Can. J. Math., Vol. XLI, No. 6, 1989, pp. 1101–1116

GENERALIZED NEWTON-PUISEUX THEORY AND HENSEL'S LEMMA IN C[[x, y]]

TZEE-CHAR KUO

The Newton polygon and the Newton-Puiseux algorithm ([**3**], p. 370, [**8**], p. 98), and their generalizations, serve as a powerful tool for analysing the singularities of a given function. Yet experts know how difficult it is to keep track of them when one, or several, blowing-ups are applied. Thus many interesting theorems are stated under the strong, rather undesirable, assumption that the Newton faces are non-degenerate.

In this paper, we introduce a method which is parallel to the classical Newton-Puiseux theory, yet avoids blowing-ups and fractional power series, except in the proofs.

Given an irreducible curve germ, Γ, at $O \in \mathbf{C}^2$, and given $f(x,y)$, we define, in Section 2, the notion of Taylor's expansion of f at Γ. When Γ is smooth, this reduces to the usual Taylor expansion at O. When Γ is singular, there is a succession of blowing-ups, β, which desingularizes Γ to a curve Γ^*, having a point O^* corresponding to O. Then, morally speaking, the Taylor expansion at Γ serves as a "remote control" on the behavior of $f \circ \beta$ near O^*.

The notion of the Newton polygon, and that of the associated polynomial equation of an edge ([**8**], p. 100), can likewise be generalised. We then have the Generalised Hensel's Lemma which gives a necessary and sufficient condition for reducibility. (Compare [**6**].)

Then, in Section 5, we present an algorithm for factoring f into its irreducible components, of which the classical Newton-Puiseux algorithm can be considered as a special case.

A corner stone of this work is a complete list of irreducible curve germs and their defining equations, given in Section 1, which is indexed on the characteristic sequences: one equation (involving some parameters) for each isotopy class. (A different listing is given in [**2**].)

The defining equation of an irreducible curve germ, Γ, also gives rise, in a natural manner, to what we call the Γ-adic expansion base in Section 1. This is a special case of the G-adic expansion base defined by Abhyankar and Moh ([**1**], p. 29). The fact that the Γ-adic base is tied up with an irreducible curve germ (rather than being a general base) has strong implications which are vital for the results.

The author would like to thank S. Abhyankar, Pierre Milman and Mutsuo Oka for stimulating conversations during their visits in Sydney; without Pierre's encouragement this paper would not have been written up.

Received November 2, 1988.

1. General equation of an irreducible curve germ. Consider a finite, or infinte, sequence of pairs of positive integers

$$\mathcal{P} = \{(d_0, n_0), (d_1, n_1), \ldots, (d_s, n_s), \ldots\}$$

where $d_0 = n_0 = 1 < d_i, d_i, n_i$ are relatively prime, and

$$1 < \frac{n_1}{d_1} < \frac{n_2}{d_1 d_2} < \cdots < \frac{n_s}{d_1 \cdots d_s} < \cdots. \tag{1}$$

The following shorthand will be used throughout this paper:

$$D_i = d_0 \cdots d_i ; p_i = \frac{n_i}{D_i} ; \nu_{i+1} = p_{i+1} - p_i ; \quad i \geqq 0.$$

We may call p_i the Puiseux exponents, and ν_i the Newton exponents.

Gives $s \geqq 1$, let us write $\mathcal{P}_s$ for the truncated sequence

$$\mathcal{P}_s = \{(d_1, n_1), \ldots, (d_s, n_s)\}.$$

We shall determine the general equation of an irreducible curve germ, Γ_s, having $\mathcal{P}_s$ as its characteristic sequence. Such a curve germ will be called a $\mathcal{P}_s$-curve (germ). By a $\mathcal{P}_0$-curve we shall mean the germ of a smooth curve.

Now, let $\mathcal{P}$ be given, satisfying (1). Consider an open subset of $\mathbf{C}^2$ with a coordinate system $\{x, y\}$. A sequence of monic polynomials in x, with coefficients in $\mathbf{C}[[y]]$,

$$G_{-1} = y, G_0, \ldots, G_s, \ldots,$$

is defined recursively as follows. First, take any complex number c_0 and define

$$g_0 = x - c_0 G_{-1}, G_0 = g_0 + a_1(G_{-1})$$

where a_1 is any formal power series with $O(a_1) > 1$. Clearly, $G_0 = 0$ is the general equation of a $\mathcal{P}_0$-curve, Γ_0, transverse to the x-axis.

Asssume, by induction, that Γ_i and its defining equation $G_i = 0$, for $0 \leqq i \leqq s-1$, have been defined, and that a rational number $w(\Gamma_i)$, called the weight of G_i, has been defined for each $i \leqq s-2$, where $w(G_{-1}) = 1$. We then define $w(G_{s-1})$ by the formula

$$w(G_{s-1}) = \sum_{i=0}^{s-1} (d_i - 1) w(G_{i-1}) + p_s. \tag{2}$$

As an easy consequence, we have

$$w(G_{s-1}) = d_{s-1} w(G_{s-2}) + \nu_s > d_{s-1} w(G_{s-2}). \tag{3}$$

Definition. A Γ_{s-1}*-adic monomial*, or simply a Γ_{s-1}*-monomial*, is an expression of the form $cG_{-1}^{e_{-1}}G_0^{e_0}\dots G_{s-1}^{e_{s-1}}$, where

$$c \in \mathbf{C}, e_{-1} \geqq 0, e_{s-1} \geqq 0, d_{i+1} - 1 \geqq e_i \geqq 0, i = 0, \dots, s-2.$$

A *Weierstrass* Γ_{s-1}*-polynomial* is a monic polynomial in G_{s-1} of the form

$$G_{s-1}^k + a_1(G_{-1}, \dots, G_{s-2})G_{s-1}^{k-1} + \dots + a_k(G_{-1}, \dots, G_{s-2})$$

which is also a series (i.e., a finite, or infinite, sum) of Γ_{s-1}-monomials.

LEMMA 1. *Consider N/D_s, where $N \in \mathbf{Z}^+$ is given.*

(*A*) *There exists a unique $(s+1)$-tuple $(e_{-1}, e_0, \dots, e_{s-1})$ of integers such that*

$$N/D_s = e_{-1} + e_0 w(G_0) + \dots + e_{s-1} w(G_{s-1})$$

where $d_{i+1} - 1 \geqq e_i \geqq 0$ for $0 \leqq i \leqq s-1$. (We merely have $e_{-1} \in \mathbf{Z}$.)

(*B*) *In case $N/D_s = d_s w(G_{s-1})$, we then have $e_{-1} > p_s > 0$ and $e_{s-1} = 0$. (In fact, $e_{s-1} = 0$ if and only if N is divisible by d_s.)*

(*C*) *In case $N/D_s > d_s w(G_{s-1})$, we still have $e_{-1} > p_s > 0$.*

The proof, of arithmetic nature, is given at the end of the section.

Now we define Γ_s and G_s. Choose $(e_{-1}, \dots, e_{s-1})$, with $e_{s-1} = 0$, according to (B). Take a complex number $c_s \neq 0$, and set

$$g_s = G_{s-1}^{d_s} - c_s G_{-1}^{e_{-1}} G_0^{e_0} \dots G_{s-2}^{e_{s-2}}.$$

Then take G_s to be a Weierstrass Γ_{s-1}-polynomial of the form

$$G_s = g_s + a_1(G_{-1}, \dots, G_{s-2})G_{s-1}^{d_s - 1} + \dots + a_{d_s}(G_{-1}, \dots, G_{s-2}) \tag{4}$$

with $O(a_j) > jw(G_{s-1}), 1 \leqq j \leqq d_s$.

Here $O(a_j)$ is the order of a_j when weights are assigned to G_i according to (2).

Attention. When $\mathcal{P}$ is a finite sequence terminating at (d_s, n_s), the above construction finishes at G_s; and then $w(G_s)$ is *not* defined. We call $\{G_{-1}, \dots, G_s\}$ a Γ_s*-adic expansion base* in $\mathbf{C}[[x, y]]$.

THEOREM 1. *The general equation of a $\mathcal{P}_s$-curve, Γ_s, is $G_s = 0$.*

That is to say, for any choice of c_i and a_j in the above construction, the resulting equation $G_s = 0$ defines a $\mathcal{P}_s$-curve; and conversely, the defining equation of any $\mathcal{P}_s$-curve can be obtained in this way, up to a unit factor and a rotation of the coordinate axis.

Note. When we take all $a_j = 0$, the resulting g_s may be called the "simplest" polynomial defining a $\mathcal{P}_s$-curve. However, in general, its degree is not the smallest. For example, both $g_1 = x^2 - y^7$ and $(x - y^2)^2 - yx^3$ define a (2,7)-curve.

https://doi.org/10.4153/CJM-1989-048-7 Published online by Cambridge University Press

Proof of Lemma 1. The integers

$$n_s, d_s n_{s-1}, d_s d_{s-1} n_{s-2}, \ldots, d_s \cdots d_2 n_1, d_s \cdots d_1$$

have no common factor > 1. Hence we can find integers E_i,

$$N/D_s = E_{s-1} p_s + \cdots + E_0 p_1 + E_{-1},$$

where we can assume $0 \leqq E_i \leqq d_{i+1} - 1$ for $i = 0, \ldots, s-1$.

By a repeated application of (2), we shall have

$$(5) \qquad N/D_s = e_{s-1} w(G_{s-1}) + \cdots + e_0 w(G_0) + e_{-1},$$

where $e_{-1} \in \mathbf{Z}, d_{i+1} - 1 \geqq e_i \geqq 0$ for $i = 0, \ldots, s-1$.

Since $w(G_i)$ is of the form N/D_{i+1}, but not of the form N/D_i, uniqueness follows, completing the proof of (A).

For (B), note that N must be divisible by d_s in this case. Hence $E_{s-1} = e_{s-1} = 0$. Using (2), (5) and the fact that $d_s > 1$, we then have

$$e_{-1} > w(G_{s-1}) - \sum_{i=0}^{s-1} (d_i - 1) w(G_{i-1}) = p_s > 0.$$

The proof of (C) is the same.

Examples. $(x^2 - y^3)^2 - y^7$ is not of the form g_2, hence reducible; $(x^2 - y^3)^2 - y^5 x$ is of the form g_2, having characteristic sequence $\{(2, 3), (2, 7)\}$, which is shared by the Eisenbud-Neumann example ([**4**] p. 58) $x^4 - 2y^3x^2 - 4a^2y^5x + y^6 - a^4y^7 = (x^2 - y^3)^2 - 4a^2y^5x - a^4y^7$.

2. Generalized Taylor expansion, Newton polygon and Hensel's lemma. Let Γ be a given $\mathcal{P}_s$-curve, $\{G_{-1} = y, G_0, \ldots, G_s\}$ a Γ-adic expansion base as constructed in Section 1, where $G_s = 0$ defines Γ. This base will be fixed in the rest of this paper. Note that the degree of G_i (in x) divides that of G_{i+1}. Hence it is easy to see that a given $f(x, y) \in \mathbf{C}[[x, y]]$ can be expressed, uniquely, as a series of Γ-monomials.

$$(6) \qquad f(x, y) = \sum c_{(e_{-1}, \ldots, e_s)} G_{-1}^{e_{-1}} G_0^{e_0} \ldots G_s^{e_s}.$$

We call (6) *the Taylor expansion* of f at Γ. (This notion readily generalizes to the n-variable case.)

In a coordinate plane, let us plot a dot, called a Newton dot, at the point (u, v) where

$$u = e_s, v = \sum_{i=-1}^{s-1} e_i w(G_i),$$

for every non-zero term in (6).

Definition. The *Newton Polygon of f* with respect to Γ is the boundary of the convex hull generated by the quadrants

$$(u, v) + \{(s, t) \in \mathbf{R}^2 : s \geqq 0, t \geqq 0\},$$

for all Newton dots (u, v).

Let us now choose an arbitrary angle θ for which

$$\tan\theta \geqq d_s w(G_{s-1}).$$

We are merely interested in the case when $\tan\theta$ is rational. Let it be written as

$$(7) \qquad \tan\theta = \frac{n}{D_s d}, \quad n, d \text{ relatively prime, } d \geqq 1.$$

We like to collect the Newton dots along a given line with slope $-\tan\theta$. The equation of such a line is

$$(8) \qquad \mathcal{L}_w : u \tan\theta + v = w, \quad w \text{ a constant.}$$

Let m denote the smallest value of w for which $\mathcal{L}_m$ contains at least one Newton dot. Amongst all the Newton dots on $\mathcal{L}_m$, let $(\mu_s, \sum_{i=-1}^{s-1} \mu_i w(G_i))$ be the one with maximal u-coordinate μ_s.

Dividing μ_s by d yields.

$$(9) \qquad \mu_s = qd + r \quad 0 \leqq r < d.$$

It is then quite clear that any Newton dot on $\mathcal{L}_m$ can only be one of the points $(u_j, v_j), 0 \leqq j \leqq q$, where

$$(10) \qquad u_j = \mu_s - jd, v_j = m - u_j \tan\theta.$$

Using Lemma 1, (B), (C), we can choose a unique $(s + 1)$-tuple $(h_{-1}, \ldots, h_{s-1})$,

$$(11) \qquad d \tan\theta = \sum_{i=-1}^{s-1} h_i w(G_i)$$

where $h_{-1} > 0, d_{i+1} - 1 \geqq h_i \geqq 0, i = 0, \ldots, s - 1$. The above (10) can be rewritten as

$$u_j = \mu_s - jd, v_j = \sum(\mu_i + jh_i)w(G_i).$$

Notation. $\Delta \equiv G_{-1}^{h_{-1}} \cdots G_{s-1}^{h_{s-1}}, \tau_j \equiv G_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}} \Delta^j, 0 \leqq j \leqq q$.

The total exponent of G_i in τ_j is $\mu_i + jh_i$, which may be $\geqq d_{i+1}$ for some i. When this happens, we ought to expand τ_j into its own Taylor expansion at Γ. The following lemma gives information on the Newton dots.

More generally, let $\tau \equiv G_{-1}^{v_{-1}} \ldots G_{s-1}^{v_{s-1}} G_s^{u^*}$ be given. Let us write

$$v^* = \sum_{i=-1}^{s-1} v_i w(G_i), w^* = d_s w(G_{s-1}) u^* + v^*.$$

LEMMA 2. *The Taylor expansion of τ has its Newton dots lying in the region*

$$R(u^*, v^*) = \{(u, v) : d_s w(G_{s-1}) u + v \geqq w^*, u \geqq u^*, v \geqq 0\}.$$

There is definitely a Newton dot at the corner point (u^*, v^*); if $v_{s-1} \leqq d_s - 1$ then there is no other dot on the line $\mathcal{L}_{w^*}^* : d_s w(G_{s-1}) u + v = w^*$.

The proof is by induction, the hypothesis being

(I_k). The above assertion is true for all τ such that

$$0 \leqq v_i \leqq d_{i+1} - 1, \quad \text{for} \quad k + 1 \leqq i \leqq s - 1.$$

(No restriction on $v_i, -1 \leqq i \leqq k$.)

Of course, I_{-1} is true.

Assuming $I_k, k < s - 1$, to prove I_{k+1}, we use induction again:

(A_N). The assertion holds for all τ such that

$$0 \leqq v_{k+1} \leqq N, 0 \leqq v_i \leqq d_{i+1} - 1, k + 2 \leqq i \leqq s - 1.$$

When $N \leqq d_{k+2} - 1, A_N$ is already true.

Assuming $A_N, N + 1 \geqq d_{k+2}$, to prove A_{N+1}, we take a τ with $v_{k+1} = N + 1$ and use the formula

$$G_{k+1}^{d_{k+2}} = G_{k+2} + c_{k+2} G_{-1}^{e_{-1}} \cdots G_k^{e_k} - \sum_{j=1}^{d_{k+2}} a_j(G_{-1}, \ldots, G_k) G_{k+1}^{d_{k+2}-j}$$

which is just the definition of G_{k+2}, (4), to reduce τ, yielding

$$\tau = \tau^{(1)} + \sigma^{(1)} + \sum_j \sigma_j^{(1)}$$

where $\tau^{(1)}$ is τ with $G_{k+1}^{v_{k+1}}$ and $G_{k+2}^{v_{k+2}}$ replaced by $G_{k+1}^{v_{k+1}-d_{k+2}}$ and $G_{k+2}^{v_{k+2}+1}$ respectively; the meaning of the $\sigma's$ is obvious.

By the induction hypothesis, the Newton dots of $\sigma^{(1)}$ lie in $R(u^*, v^*)$, and (u^*, v^*) is one of the dots.

https://doi.org/10.4153/CJM-1989-048-7 Published online by Cambridge University Press

Similarly, the Newton dots of $\sigma_j^{(1)}$ lie in the region $R(u^*, v_j^*)$ where

$$v_j^* = v^* - d_{k+2}w(G_{k+1}) + O(a_j) + (d_{k+2} - j)w(G_{k+1}).$$

From the definition of G_{k+2} we know $v_j^* > v^*$. Hence the Newton dots of $\sigma_j^{(1)}$ lie above the line $\mathcal{L}_{w^*}^*$.

It remains to consider $\tau^{(1)}$. The argument is divided into three cases. (A): $k+1 < s-1, v_{k+2}+1 \leqq d_{k+3}-1$, (B): $k+1 < s-1, v_{k+2}+1 \geqq d_{k+3}$, and (C): $k+1 = s-1$.

For (A), the induction hypothesis applies to $\tau^{(1)}$. By (3), with $s-1 = k+2$, the Newton dots of $\tau^{(1)}$ lie above the line $\mathcal{L}_{w^*}^*$.

For (C), the induction hypothesis still applies. The Newton dots of $\tau^{(1)}$ are contained in $R(u^*+1, v^* - d_s w(G_{s-1}))$. The proof is also finished.

When (B) happens, the reduction process can be iterated on G_{k+2}, yielding

$$\tau^{(1)} = \tau^{(2)} + \sigma^{(2)} + \sum \sigma_j^{(2)}.$$

The Newton dots of $\sigma^{(2)}$ and $\sigma_j^{(2)}$ lie above $\mathcal{L}_{w^*}^*$, causing no trouble. As for $\tau^{(2)}$, again the argument is divided into cases (A), (B) and (C). If (B) happens, the reduction continues.

But (B) can not happen more than $s-k$ times. Hence Lemma 2 is proved.

Now, take a term γ_j in (6) which is represented by (u_j, v_j) on $\mathcal{L}_m$. Using Lemma 2, there is a *unique* constant $a_j \neq 0$ such that γ_j appears as a term in the Taylor expansion of $a_j \tau_j G_s^{u_j}$.

In case (u_j, v_j) does not represent a non-zero term in (6), define $a_j = 0$.

Definition. Given the Taylor expansion (6). The polynomial associated to the given angle θ is

$$\varphi_\theta(z) = z^r[a_0 z^q + \cdots + a_q] \tag{12}$$

where q, r, are defined in (9).

We know $a_0 \neq 0$. There is another $a_j \neq 0$ if and only if the Newton Polygon has an edge E with $\theta_E = \theta$. Here θ_E denotes the angle between E and the negative u-direction.

Given E, the polynomial $\varphi_{E_\theta}(z)$ will be written simply as $\varphi_E(z)$. We also write

$$\Phi_E(z) \equiv a_0 z^q + \cdots + a^q.$$

Illustrative Examples. (A) Consider $f(x, y) = (x^2 - 2y^3)^2 + y^7$, we have $G_0 = x, G_1 = x^2 - 2y^3, w(G_0) = 3/2; \tan\theta = 7/2, d = 1, q = 2, r = 0, h_{-1} = 2, h_0 = 1, \Delta = y^2 x, \Delta^2 = 2y^7 + y^4(x^2 - 2y^3)$. Hence $a_0 = 1, a_1 = 0, a_2 = 1/2, \varphi_E(z) =$

$z^2 + 1/2$ which has two distinct roots. By the Generalized Hensel's Lemma (see below), f is reducible.

(B). Consider $f(x,y) = (x^2 - 2y^3)^2 - xy^5$. Again $G_0 = x, G_1 = x^2 - 2y^3$, while $\tan\theta = 13/4, d = 2, q = 1, r = 0, d\tan\theta = 13/2, h_{-1} = 5, h_0 = 1, \Delta = y^5x$ and $\varphi_E(z) = z - 1$, which has only one root. We know, from Section 1, that f is irreducible.

Definition. An edge, E, of the Newton Polygon of (6) is *relevant* if

$$\tan\theta_E \geqq d_s w(G_{s-1}).$$

Call E *strictly relevant* if this is a strict inequality.

When Γ is defined by $x = 0$, E is relevant ($\tan\theta_E \geqq 1$) if and only if the Puiseux roots arising from E have order $\geqq 1$.

THEOREM 2 (Generalized Hensel's Lemma). *A formal power series $f(x,y)$, having no multiple factors, is reducible if and only if there exists an irreducible curve germ, Γ, with respect to which the Newton Polygon of f has a relevant edge, E, whose associated polynomial equation $\varphi_E(z) = 0$ has two, or more, distinct roots. In this case, given a factorization in* $\mathbf{C}[z]$:

$$(13) \qquad \varphi_E(z) = \eta(z)\zeta(z), \eta, \zeta \textit{ being relatively prime},$$

there is a corresponding factorization in $\mathbf{C}[\![x,y]\!]$:

$$(14) \qquad f(x,y) = h(x,y)k(x,y)$$

such that η, ζ are polynomials associated to θ_E for h, k respectively.

As a corollary, we derive the following interesting result of M. Oka, which is contained implictly in his paper [**7**].

First, observe that if $\Phi_E(z) = 0$ has no multiple non-zero roots, then each non-zero root gives rise to an irreducible factor of f; and different roots give rise to different irreducible factors. Call E non-degenerate in this case.

Now, consider the Newton Polygon of f in the usual sense. The number of integral (lattice) points on a given edge E equals the number of non-zero roots minus 1.

COROLLARY (M. Oka). *Suppose the Newton Polygon of f has a vertex on each coordinate axis, and every edge is non-degenerate, then the number of irreducible factors of f equals $N(f) - 1$, where $N(f)$ denotes the number of integral points on the Newton Polygon. Moreover, these factors are all different.*

3. Proof of theorem 1. Let us consider the following two induction hypothesis:

(I_s). $G_i = 0$ is the general equation of a $\mathcal{P}_i$-curve $0 \leqq i \leqq s$.

(II_s). Fix any $i, 1 \leqq i \leqq s$. Take a Puiseux root, $x = \lambda(y)$, of $G_i = 0$. Let $\hat{\lambda}(y)$ denote the series which is $\lambda(y)$ omitting all terms of degree $\geqq p_i$. Then $\hat{\lambda}(y)$ is a Puiseux root of $G_{i-1} = 0$.

We have seen that (I_0) is true; (II_0) says nothing, hence is true. Note that all Puiseux roots of an irreducible curve are conjugate, hence (II_s) is independent of the choice of λ.

A number of important consequences can be derived from the above hypothesis.

Consider a given $i \geqq 1$. There are D_i Puiseux roots of $G_i = 0$; let them be denoted by $\lambda_1, \ldots, \lambda_m, m = D_i$. We define

$$l(G_i) = \sum_{j=2}^{m} O(\lambda_j(y) - \lambda_1(y)).$$

(This number is closely related to the self-linking number of the knot, see [**5**], p. 301.)

We also define $l(G_0) = 0$.

Since the λ's are conjugate, $l(G_i)$ is well defined.

An examination on the tree-models of G_i and G_{i-1} ([**5**], p. 308), superposed according to (II_s), leads to the following identity:

$$l(G_i) + p_i = d_i p_i + d_i l(G_{i-1}), \quad 0 \leqq i \leqq s,$$

which can be rewritten as

$$(15) \qquad l(G_i) = d_i l(G_{i-1}) + p_i(d_i - 1).$$

The details of the proof is omitted.

On the other hand, combining (3), (15) and an easy induction yields

$$w(G_i) - l(G_i) = p_{i+1}, \quad 0 \leqq i \leqq s.$$

Let us take a Puiseux root, $\sigma(y)$, of $G_s = 0$. Then consider $\sigma(y) + \tau y^{p_s}$, where τ is an indeterminant. An examination of the superposed tree-models of G_s and G_i leads to

$$(16) \qquad O(G_s(\sigma(y) + \tau y^{p_s}, y)) = l(G_s) + p_s$$

$$O(G_i(\sigma(y) + \tau y^{p_s}, y)) = l(G_i) + p_{i+1} = w(G_i), \quad 0 \leqq i \leqq s - 1.$$

In fact, the following stronger formulae hold:

$$G_s(\sigma(y) + \tau y^{p_s}, y) = \tau y^{p_s}[a y^{l(G_s)} + R_s]$$

$$G_i(\sigma(y)+\tau y^{p_s}, y) = b_i y^{w(G_i)} + R_i, \quad 0 \leqq i \leqq s-1$$

where the constants $a \neq 0, b_i \neq 0$ are independent of τ, and where $R_i, 0 \leqq i \leqq s$, are power series in y^{1/D_s}, with coefficients in $\mathbf{C}[\tau]$, such that

$$O_y(R_s) \geqq l(G_s), O_y(R_i) \geqq w(G_i), 0 \leqq i \leqq s-1.$$

Moreover, when setting $\tau = 0$, these become strict inequalities.

We are now ready to find the Puiseux roots $\tau = \tau(y)$ of the equation

$$G_{s+1}(\sigma(y)+\tau y^{p_s}, y) = 0,$$

which can be rewritten in the form

$$\left\{ a^{d_{s+1}} \tau^{d_{s+1}} - c_{s+1} \left[\prod_{i=-1}^{s-1} b_i^{e_i} \right] y^{d_{s+1}\nu_{s+1}} \right\} + \cdots = 0.$$

There are d_{s+1} Puiseux roots of the form

$$\tau = \tau_0 y^{\nu_{s+1}} + \cdots \quad \text{(power series in } y^{1/D_{s+1}}\text{)}$$

where τ_0 are the roots of

$$a^{d_{s+1}} z^{d_{s+1}} - c_{s+1} \prod b_i^{e_i} = 0.$$

These Puiseux roots are conjugate under the group of d_{s+1}-th roots of unity. Hence, as we run through all D_s conjugate choices of $\sigma(Y)$, we will find altogether D_{s+1} Puiseux roots of $G_{s+1} = 0$, which are all conjugate. Since G_{s+1} is regular in x of order D_{s+1}, there is no other Puiseux root of order > 0. Hence $G_{s+1} = 0$ is a $\mathcal{P}_{s+1}$-curve.

Now we prove the converse. Let a $\mathcal{P}_{s+1}$-curve be given. We can assume it is not tangent to the x-axis. Take one of its Puiseux roots

$$\lambda(y) = \cdots + b_1 y^{p_1} + \cdots + b_s y^{p_s} + \cdots + b_{s+1} y^{p_{s+1}} + \cdots, \qquad b_i \neq 0.$$

Let $\sigma(y)$ denote $\lambda(y)$ with all terms of degree $\geqq p_{s+1}$ omitted.

Using (I_s), we can find an expansion base $\{G_{-1} = y, G_0, \ldots, G_s\}$ such that $G_s = 0$ defines a $\mathcal{P}_s$-curve having $\sigma(y)$ as a root.

Let us consider

$$H_1(x,y) = G_s^{d_{s+1}} - a G_{-1}^{e_{-1}} G_0^{e_0} \cdots G_{s-1}^{e_{s-1}}$$

where a, e_i are determined as follows.

We can (uniquely) choose values for e_i according to Lemma 1(B), so that

$$e_{-1} > 0, e_{-1} + e_0 w(G_0) + \cdots + e_{s-1} w(G_{s-1}) = d_{s+1} w(G_s).$$

An examination of the superposed tree-model of $G_i, 0 \leqq i \leqq s$, yields

$$O(G_i(\lambda(y), y)) = w(G_i), \quad 0 \leqq i \leqq s.$$

Hence we can uniquely choose a value $a \neq 0$ so that

$$O(H_1(\lambda(y), y) > d_{s+1} w(G_s).$$

Note that the left hand side is a number of the form N/D_{s+1}.

Having defined H_1, we then consider

$$H_2(x, y) = H_1(x, y) - aG_{-1}^{e_{-1}} \cdots G_{s-1}^{e_{s-1}} G_s^{e_s}$$

where a, e_i are determined as follows.

Using a similar argument, we can find values $e_i (d_{s+1} > e_s)$ and $a \neq 0$ such that

$$O(H_2(\lambda(y), y)) > O(H_1(\lambda(y), y)),$$

where the left hand side is of the form N/D_{s+1}.

The construction can be repeated recursively to yield a sequence $\{H_n\}$ such that

$$O(H_n(\lambda(y), y)) > O(H_{n-1}) = O(H_n - H_{n-1}),$$

and these numbers are of the form N/D_{s+1}.

Now define

$$g_{s+1}(x, y) = H_1(x, y), G_{s+1} = g_{s+1} + \sum_{n=1}^{\infty} (H_n - H_{n-1}).$$

Then, clearly, $\lambda(y)$ is a root of $G_{s+1} = 0$, and so are its D_{s+1} conjugates.

It follows that the given $\mathcal{P}_{s+1}$-curve must be defined by $G_{s+1} = 0$, up to a unit factor, thus completing the proof of (I_{s+1}), (II_{s+1}).

4. Proof of theorem 2, the "if" part. Consider a relevant edge, E, and assume a factorization (13) exists. Take E_θ to be the angle θ in Section 2.

Given a polynomial $p(z)$, let $p(z_1, z_2)$ denote its homogenization, i.e., the homogeneous form having the same degree as $p(z)$ with $p(z, 1) = p(z)$.

Consider the line $\mathcal{L}_m$ defined by (8) for E_θ.

LEMMA 3. *Let L_m denote the sum of terms in* (6) *whose Newton dots lie on $\mathcal{L}_m$. Then*

$$L_m = G_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}} G_s^r \Phi_E(G_s^d, \Delta) \tag{17}$$

modulo Γ-*monomials lying above* $\mathcal{L}_m$.

Proof. In case $\tan\theta_E > d_s w(G_{s-1})$, (17) is an immediate consequence of Lemma 2, applied to each $\tau_j, 0 \leqq j \leqq q$.

Now suppose $\tan\theta_E = d_s w(G_{s-1})$. This condition has strong implications. Since $d_s w(G_{s-1})$ is a number of the form N/D_{s-1}, we must have: $d = 1, n$ is divisible by d_s, and also $h_{s-1} = 0$ in (11), where n, d were defined in (7).

Thus $\mu_i + jh_i \geqq d_{i+1}$ can happen only when $i \leqq s-2$. Hence, by the last part of Lemma 2, in the Taylor expansion of $\tau_j G_s^{u_j}, (u_j, v_j)$ is the only Newton dot on $\mathcal{L}_m$, all other dots lie above it.

Now, $\sum_j a_j \tau_j G_s^{u_j}$ is just the right hand side of (17), proving Lemma 3.

Given a number w, of the form $N/D_s d, N \in \mathbf{Z}^+$, by an E-form of weight w we mean a sum of Γ-monomials lying on the line $\mathcal{L}_w$.

LEMMA 4. *Let L be a given E-form, say of weight w^*. Suppose $w^* > m$. Then there exist two E-forms Q and R,*

$$L = QG_s^r \Phi_E(G_s^d, \Delta) + RG_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}}$$

modulo Γ-*monomials lying above* $\mathcal{L}_{w^*}$.

Proof. Lemma 3 can be applied to L also, giving

$$L = G_{-1}^{\mu_{-1}^*} \cdots G_{s-1}^{\mu_{s-1}^*} G_s^{r^*} \Phi_E^*(G_s^d, \Delta)$$

modulo Γ-monomials lying above $\mathcal{L}_{w^*}$. (When L has only one term, $\Phi_E^* = 1$.)

Of course, we don't necessarily have $\mu_i^* \geqq \mu_i$.

Now $G_s^{r^*}\Phi_E^*, G_s^r \Phi_E$ can be considered as weighted homogeneous forms in G_s and Δ, when G_s and Δ are given weights 1 and d respectively; both are monic in G_s. Let the former be divided by the latter, yielding

$$G_s^{r^*}\Phi_E^*(G_s^d, \Delta) = \tilde{Q} G_s^r \Phi_E(G_s^r, \Delta) + \Delta^j \tilde{R} \tag{18}$$

where $\tilde{Q}, \tilde{R}$ are weighted homogeneous forms in G_s, Δ,

$$jd + (\deg \tilde{R} \text{ in } G_s) = \mu_s^*, \deg \tilde{R} \text{ in } G_s < \mu_s.$$

Let us first consider the case $\tan\theta_E > d_s w(G_{s-1})$.

Then we show that $G_{-1}^{\mu_{-1}^*} \cdots G_{s-1}^{\mu_{s-1}^*} \Delta^j$ is "divisible" by $G_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}}$.

Using the assumption $w^* - m > 0$, we find

$$(\deg \tilde{R}) \tan\theta_E + \sum (\mu_i^* + jh_i) w(G_i) > \mu_s \tan\theta_E + \sum \mu_i w(G_i)$$

and hence

$$\sum (\mu_i^* + jh_i) w(G_i) - \sum \mu_i w(G_i) > \tan\theta_E \geqq d_s w(G_{s-1}).$$

By Lemma 1(C), there exists $(e_{-1}, \ldots, e_{s-1})$,

$$(19) \qquad \sum(\mu_i^* + jh_i)w(G_i) - \sum \mu_i w(G_i) = \sum e_i w(G_i),$$

whence, by Lemma 2,

$$(20) \qquad aG_{-1}^{\mu_{-1}^*} \cdots G_{s-1}^{\mu_{s-1}^*} \Delta^j \tilde{R} = G_{-1}^{e_{-1}} \cdots G_{s-1}^{e_{s-1}} G_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}} \tilde{R}$$

modulo Γ-monomials lying above $\mathcal{L}_{w^*}$. Here $a \neq 0$ is a constant.

Lemma 4 follows from (18) and (20) in this case.

It remains to consider the case $\tan\theta_E = d_s w(G_{s-1})$. The proof is by induction on $\deg \Phi_E$.

Note that, as before, we must have $d = 1$, and $r = 0$. Hence Φ_E decomposes into a product of linear factors, possibly repeated. Let us take one of the factors, say $G_s - c\Delta, c \in \mathbf{C}$, and make the substitution $G_s = \tilde{G}_s + c\Delta$ in L and L_m.

We shall write $\tilde{G}_s$ simply as G_s, abusing notations.

Then Φ_E is divisable by G_s and can be written in the form

$$\Phi_E = G_s \psi_E(G_s, \Delta).$$

Again we choose e_i satisfying (19), and then, by Lemma 2, (20) holds modulo monomials on $\mathcal{L}_{w*}$ which are divisible by G_s, and monomials lying above $\mathcal{L}_{w^*}$. It follows that L can be written in the form

$$L = G_s L^* + G_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}} M^*$$

modulo monomials above $\mathcal{L}_{w^*}$.

An application of the induction hypothesis to $G_s^{-1} L_m$ and L^* completes the proof of Lemma 4.

Now, a recursive application of Lemma 4 generates two sequences of E-forms $\{Q_n\}, \{R_n\}$, such that

$$f(x,y) = [G_{-1}^{\mu_{-1}} \cdots G_{s-1}^{\mu_{s-1}} + Q_1 + Q_2 + \cdots][G_s^r \Phi_E(G_s^d, \Delta) + R_1 + R_2 + \cdots],$$

the weights being increasing in each factor.

It remains to decompose the second factor, using (13).

The factor z^r in (12), if $r \geqq 1$, is contained entirely in $\eta(z)$ or in $\zeta(z)$, which are relatively prime. Let us assume it is in $\eta(z)$, and write $\eta(z) = z^r \tilde{\eta}(z)$.

We then choose polynomials $a(z), b(z)$,

$$a(z) z^r \tilde{\eta}(z) + b(z)\zeta(z) = 1.$$

And, using a similar recursive argument as above, we can find two sequences of E-forms $\{A_n\}, \{B_n\}$ such that

$$G_s^r \Phi_E + R_1 + R_2 + \cdots = [G_s^r \tilde{\eta}(G_s^d, \Delta) + B_1 + \cdots][\zeta(G_s^d, \Delta) + A_1 + \cdots],$$

the weight being increasing in each factor.

Finally, writing Δ and its powers in terms of Γ-monomials will yield the desired factorization (14).

5. A factorization algorithm (generalized Newton-Puiseux algorithm). The proof of the "only if" part of Theorem 2 is also incorporated into the description of the algorithm.

Let $f(x, y)$ be given, having no multiple factors. We shall construct recursively a finite sequence

$$\Gamma_i \equiv \Gamma_i^{(0)}, \Gamma_i^{(1)}, \ldots, \Gamma_i^{(k_i)},$$

of $\mathcal{P}_i$-curves, for $i = 0, \ldots, N$, where N is to be determined. Then, at the end, we shall either arrive at a $\mathcal{P}_N$-curve, which coincides with $f = 0$, or else find that f is reducible, the algorithm is then continued on each factor.

Let $D = O(f)$ be the order of f. In case $D = 1$ f is equivalent to x; this case is trivial.

Suppose $D \geqq 2$. We can apply a linear transformation so that f becomes regular in x:

$$f(x, 0) = a_0 x^D + \text{ higher order terms}, \quad a_0 \neq 0.$$

In case the initial form of f has two, or more, distinct factors, f is reducible, by the well-known Hensel's Lemma. Otherwise, we can perform a linear transformation, if necessary, so that x^D is the initial form.

Let us take $\Gamma_0 = \Gamma_0^{(0)}$ to be the curve defined by $x = 0$. Let $G_{-1} = y, G_0 = x$. Then consider the Newton Polygon of f (with respect to Γ_0). There must be a compact edge, which is also strictly relevant, having $(D, 0)$ as a vertex. Since otherwise f would be divisible by x^D, a contradiction.

Now, assume $\Gamma_s^{(j)}, s \geqq 0, j \geqq 0$, have been defined, for which the following holds: The Newton Polygon of f with respect to $\Gamma_s^{(j)}$ has a compact, strictly relevant, edge, E, having $(D/D_s, 0)$ as a vertex.

Let $\{G_{-1} = y, \ldots, G_s\}$ denote the $\Gamma_s^{(j)}$-adic expansion base. (Strictly speaking, since this base depends also on j, we ought to write $G_i^{(j)}$ instead of G_i here.) There are four cases to consider.

Terminating Case: $D/D_s = 1$.

Reducible Case: $D/D_s > 1$, but the other vertex of E is not on the v-axis.

Stable Case: The other vertex lies on the v-axis, $d = 1$.

Unstable Case: The other vertex lies on the v-axis, $d > 1$.

When $D/D_s = 1$, the curve $f = 0$ coincides with a $\mathcal{P}_s$-curve defined by an equation of the form $G_s + a_1(G_{-1}, \ldots, G_{s-1}) = 0$. There is nothing more to do.

In the second case, the associated polynomial equation, $\varphi_E = 0$, has 0 as a root, and at least one non-zero root, regardless of whether $d = 1$ or $d > 1$ in (7). By the "if" part, proved in the last section, f is reducible; the algorithm is then continued on each factor of (14).

Now consider the Stable Case. When $d = 1$, φ_E has the form

$$\varphi_E(z) = a_0 z^D + a_1 z^{D-1} + \cdots + a_D.$$

An application of the Shreedharacharya-Tschirnhausen transformation $\tilde{z} = z - a_1/Da_0$ turns φ_E into a polynomial of the form

$$a_0\tilde{z}^D + \tilde{a}_2\tilde{z}^{D-2} + \cdots + \tilde{a}_D.$$

Notice that $\varphi_E(z) = 0$ has no distinct roots if and only if $\tilde{a}_j = 0$ for $2 \leqq j \leqq D$. (The author is indebted to S. Abhyankar for pointing out to him this simple, but important, fact.)

In case $\varphi_E = 0$ has distinct roots, f is reducible, the algorithm is continued on each factor. Otherwise, define

$$\tilde{G}_s = G_s - (a_1/Da_0)\Delta$$

and let $\Gamma_s^{(j+1)}$ be the $\mathcal{P}_s$-curve defined by $\tilde{G}_s = 0$. The Newton Polygon of f with respect to $\Gamma_s^{(j+1)}$ must have a compact edge, E', having $(D/D_s, 0)$ as a vertex, and

$$\tan\theta_{E'} > \tan\theta_E.$$

It follows that E' is strictly relevant.

Finally, consider the Unstable Case. Again, if φ_E has distinct roots, f will be reducible. Otherwise, we must have $r = 0$ in (12). Again, the transformation $\tilde{z} = z - a_1/Da_0$ reduces φ_E to $a_0\tilde{z}^q$. We then set

$$G_{s+1} = G_s^d - (a_1/Da_0)\Delta, d_{s+1} = d, n_{s+1} = n,$$

and let Γ_{s+1} be the $\mathcal{P}_{s+1}$-curve defined by $G_{s+1} = 0$. The Newton Polygon of f with respect to Γ_{s+1} has a vertex at $(D/D_{s+1}, 0)$.

When $D/D_{s+1} = 1$, this reduces to the Terminating Case. When $D/D_{s+1} > 1$, we have one of the three remaining cases.

The Unstable Case can not occur infinitely many times, since D/D_s keeps dropping. The Stable Case can not either, for if it did, f would have a factor of the form $G_k^{D/D_k}, D/D_k > 1$, which is a contradiction.

After a finite number of such applications, we shall arrive at a complete decomposition of f into irreducible factors, without resorting to fractional power series.

Example ([**4**], p. 58). For $f = x^4 - 2y^3x^2 + y^6 - 4a^2y^5x - a^4y^7, a \neq 0$, we find, following the algorithm, that $G_{-1} = y, G_0 = x, G_1 = x^2 - y^3$, and, finally, $G_2 = (x^2 - y^3)^2 - 4a^2y^5x$. Hence f is irreducible.

Following the algorithm, one can decide effectively whether $f(x,y)$ is reducible or not, i.e., the algorithm is a computer programming, at least when f is a polynomial.

If case f is known to be reducible, to find its irreducible factors, one has to solve the associated polynomial equations $\varphi_E = 0$. Apart from this, the process is also effective.

The classical Newton-Puiseux algorithm for finding the fractional power series roots can be considered as a special case of the above, when the fractional powers of y are introduced one by one. We omit the details.

REFERENCES

1. S. Abhyankar and T. T. Moh, *Newton-Puiseux expansion and generalized Tshianhausen transformation II*, J. reine und angew Math *261* (1973), 29–54.

2. S. Abhyankar, *Irreducibility Criteria for germs of analytic functions of two complex variables*, Advances in Math, to appear.

3. E. Brieskorn and H. Knörrer, *Plane algebraic curves* (Birkhäuser, 1986).

4. D. Eisenbud and W. Neumann, *Three-dimensional link theory and invariants of plane curve singularities*, Annals of Maths Studies *110* (Princeton University Press, 1985).

5. T. C. Kuo and Y. C. Lu, *On analytic function germs of two complex variables*, Topology *16* (1977), 299–310.

6. T. T. Moh, *On approximate roots of a polynomial*, J. reine und angew. Math. *278/279* (1975), 301–306.

7. M. Oka, *On the stability of the Newton boundary*, Proceeding of Symposia in Pure Math *40*, part 2 (1983), 259–268.

8. R. Walker, *Algebraic curves* (Princeton University Press, 1950).

University of Sydney,
Sydney, Australia