(62) in which $\boldsymbol{h} = [h_1, \ldots, h_l]$. Therefore, by Lemma 12 there exists a vector $\boldsymbol{h}^0 = [h_1^0, \ldots, h_l^0]$ such that the system of equations

$$M_{h^0_i}(x_0, x_1, \ldots, x_k) = 0 \quad (\boldsymbol{i} \in I)$$

is soluble in integers. Denoting a solution by $[x_0^0, x_1^0, \ldots, x_k^0]$ we get from (60) and (61) for all $i \leqslant l$

$$wx_0^0 + \sum_{j=1}^{k} a_{h_i^0 ij0} x_j^0 - b_{h_i^0 i0} = 0,$$

$$\sum_{j=1}^{k} a_{h_i^0 ijs} x_j^0 - b_{h_i^0 is} = 0 \quad (1 \leqslant s \leqslant r)$$

hence by (59)

$$\prod_{h=1}^{g_i} \Big(\prod_{j=1}^{h} \alpha_{hij}^{x_j^0} - \beta_{hi} \Big) = 0 \quad (1 \leqslant i \leqslant l).$$

References

[1] G. Darbi, *Sulla reducibilità delle equazioni algebriche*, Annali Mat. Pura Appl. 4 (1925), pp. 185–208.

[2] H. Hasse, *Zum Existenzsatz von Grunwald in der Klassenkörpertheorie*, J. Reine Angew. Math. 188 (1950), pp. 40–64.

[3] D. Hilbert, *Die Theorie der algebraischen Zahlkörper*, Ges. Abhandlungen, Band I, Reprint New York, 1965.

[4] M. Kneser, *Lineare Abhängigkeit von Wurzeln*, Acta Arith. 26 (1975), pp. 307–308.

[5] W. H. Mills, *Characters with preassigned values*, Canad. J. Math. 15 (1963), pp. 169–171.

[6] A. Schinzel, *On power residues and exponential congruences*, Acta Arith. 27 (1975), pp. 397–420.

[7] Th. Skolem, *Anwendung exponentieller Kongruenzen zum Beweis der Unlösbarkeit gewisser diophantischer Gleichungen*, Vid. akad. Avh. Oslo I, 1937, nr 12.

[8] — *On the existence of a multiplicative basis for an arbitrary algebraic field*, Norske Vid. Selsk. Forh. (Trondheim) 20 (1947), nr 2.

[9] N. G. Tschebotaröw (Čebotarev), *Über einen Satz von Hilbert*, Vestnik Ukr. Akad. Nauk, 1923, pp. 3–7.

[10] N. Tschebotaröw, H. Schwerdtfeger, *Grundzüge der Galois'schen Theorie*, Groningen-Djakarta 1950.

[11] S. Znám, *On properties of systems of arithmetic sequences*, Acta Arith. 26 (1975), pp. 279–283.

Correction to [6]

p. 401: insert after formula (8):

provided $p_i > 2$, $a \equiv 1 \bmod p_i$ or $p_i = 2$, $a \equiv 1 \bmod 4$.

Received on 15. 10. 1975 (772)

ACTA ARITHMETICA
XXXII (1977)

Diophantine approximation in power series fields*

by

WOLFGANG M. SCHMIDT (Boulder, Colo.)

Dedicated to Professor Theodor Schneider on his 65th birthday

1. Introduction

1.1. The setting. Let K be the field of formal series

(1) $$\alpha = a_k X^k + a_{k-1} X^{k-1} + \ldots$$

with an arbitrary integer k and with coefficients in a given field F of characteristic zero. The rational functions in X with coefficients in F form a subfield $K_0 = F(X)$ of K, and the polynomials form a subring $S = F[X]$. In K we have the non-archimedean valuation with

$$|\alpha| = 2^k$$

if the leading coefficient in (1) is $a_k \neq 0$. If f is a polynomial, then

$$|f| = 2^{\deg f}.$$

Many results on "ordinary" diophantine approximation, i.e. approximation of reals by rationals, carry over to approximation of elements of K by rational functions, i.e. by elements of K_0.

For example, Dirichlet's Theorem holds: *If $\alpha \in K$ does not lie in K_0, then there are infinitely many rational functions $f/g = f(X)/g(X)$ in K_0 with*

$$|\alpha - (f/g)| \leqslant |g|^{-2}.$$

Also Liouville's Theorem holds: *If $\alpha \in K$ is algebraic over K_0 of degree $s > 1$, then for every rational function f/g, we have*

(2) $$|\alpha - (f/g)| > c_1(\alpha) |g|^{-s},$$

with a constant $c_1(\alpha) > 0$. Now just as in ordinary diophantine approxi-

* Written with partial support from NSF-MPS75-08233.

mation, where Liouville's Theorem was eventually strengthened to Roth's Theorem [5], so in the power series case, Liouville's Theorem was strengthened by Uchiyama [7] to the following result, which we shall call the Roth–Uchiyama Theorem: *If α is algebraic of degree $s > 1$, then for $\varepsilon > 0$ and for every rational function f/g, we have*

$$|\alpha-(f/g)| > c_2(\alpha, \varepsilon)|g|^{-2-\varepsilon}, \tag{3}$$

with a constant $c_2(\alpha, \varepsilon) > 0$.

As was first observed by Kolchin [1], new questions arise in the power series case in connection with algebraic differential equations. These questions have no analog in ordinary diophantine approximation. Denote the formal derivatives of a series α by $\alpha^{(1)}, \alpha^{(2)}, \ldots$ Suppose $\alpha \notin K_0$ satisfies an algebraic differential equation, i.e. an equation

$$A(\alpha, \alpha^{(1)}, \ldots, \alpha^{(l)}) = 0,$$

where $A(Y, Y_1, \ldots, Y_l)$ is a non-zero polynomial in variables $Y, Y_1, \ldots, Y^l$ with coefficients in K_0. Then Kolchin proved that for rational functions f/g,

$$|\alpha-(f/g)| > c_3(\alpha)|g|^{-d}, \tag{4}$$

where $c_3(\alpha) > 0$ and where d is the *denomination* of A, i.e. the maximum of

$$a_0+2a_1+\ldots+(l+1)a_l$$

over all monomials $Y^{a_0}Y_1^{a_1}\ldots Y_l^{a_l}$ occurring in A with non-zero coefficients. When $l = 0$, i.e. when α is algebraic, then (4) reduces to Liouville's estimate (2). It is not known whether the exponent $-d$ in (4) is best possible if $l > 0$; perhaps the exponent should always be $-(2+\varepsilon)$ for $\varepsilon > 0$, or even -2. Contributions to this question were made by Osgood [3], and special equations were discussed by Schmidt [6]. Osgood [2] used Kolchin's Theorem to prove a result on algebraic functions which is stronger than Liouville's estimate (2) and weaker* than the Roth–Uchiyama estimate (3), but which in contrast to the latter is effective in some sense.

Now in ordinary diophantine approximation one deals not only with approximation by rationals, but also with approximation by algebraic numbers. Similarly, in the power series case, one deals with approximation by algebraic functions. In the present paper we intend to discuss approximation by solutions of algebraic differential equations. In particular, we shall prove generalizations of Liouville's Theorem, dealing with approximation to a solution of an algebraic differential equation by solutions of other differential equations.

* Except that for functions of degree 3 it is stronger.

Of course, the "expansions about infinity" (1) could be replaced by other expansions, say by expansions $a_kX^k+a_{k+1}X^{k+1}+\ldots$ about zero. The expansions (1) have the advantage that now the polynomials play the role which is played by the integers in ordinary diophantine approximation.

1.2. Heights. Let Y be a "differential variable", i.e. a variable with which we associate further variables $Y_1, Y_2, \ldots$, representing "derivatives" of Y. A *differential monomial* will be an expression

$$P(\mathbf{Y}) = P(Y, Y_1, \ldots) = Y^{a_0}Y_1^{a_1}\ldots Y_l^{a_l} \tag{5}$$

with nonnegative integers $a_0, \ldots, a_l$. We write $K[\mathbf{Y}]$ (or $S[\mathbf{Y}]$) for the ring of *differential polynomials*

$$A = \sum_n \pi_n P_n, \tag{6}$$

where the sum is finite, where the P_n are differential monomials and where the coefficients π_n lie in K (or in S). Given a differential polynomial

$$A = \sum_n f_n P_n \tag{7}$$

with $f_n \in S = F[X]$ and with distinct monomials P_n, define its *height* $\mathfrak{H}(A)$ by

$$\mathfrak{H}(A) = \max_n |f_n| = \max_n (2^{\deg f_n}). \tag{8}$$

Let $A(\alpha)$ be obtained by substituting $\alpha, \alpha^{(1)}, \alpha^{(2)}, \ldots$ for $Y, Y_1, Y_2, \ldots$ into $A[\mathbf{Y}]$.

Let $\Omega(m)$ be the set of $\beta \in K$ which satisfy a linear differential equation of order $\leqslant m$:

$$B(\beta) = 0,$$

where

$$B = g_mY_m+\ldots+g_1Y_1+g_0Y+f$$

is a non-zero linear differential polynomial of order $\leqslant m$ with coefficients $g_m, \ldots, g_0, f$ in $S = F[X]$. Here $\mathfrak{H}(B) = \max(|g_m|, \ldots, |g_0|, |f|)$. Given $\beta \in \Omega(m)$, the *height* $\mathfrak{H}_m(\beta)$ is given by

$$\mathfrak{H}_m(\beta) = \min_B \mathfrak{H}(B),$$

with the minimum to be taken over all non-zero linear differential polynomials $B \in S[\mathbf{Y}]$ of order $\leqslant m$ with $B(\beta) = 0$.

If P is again given by (5), define the *order* $l(P)$ by

$$l(P) = \begin{cases} l & \text{if } a_l > 0, \\ -1 & \text{if } a_0 = \ldots = a_l = 0. \end{cases}$$

Also put

$$s(P) = a_0 + \ldots + a_l,$$

$$r(P) = a_0 + 3a_1 + \ldots + (2l-1)a_l.$$

If A is a differential polynomial, write $l(A)$, $s(A)$, $r(A)$, respectively, for the maximum of $l(P)$, $s(P)$, $r(P)$ over the monomials P occurring in A with non-zero coefficients. Then $l(A)$ is the *order* of A.

Let $\Omega(m, s)$ be the set of $\beta \in K$ which satisfy a non-trivial differential equation $B(\boldsymbol{\beta}) = 0$ with

(9) $$B \in S[Y] \quad \text{having} \quad l(B) \leqslant m, \quad s(B) \leqslant s.$$

Given $\beta \in \Omega(m, s)$, the *height* $\mathfrak{H}_{ms}(\beta)$ is given by

$$\mathfrak{H}_{ms}(\beta) = \min_{B} \mathfrak{H}(B),$$

with the minimum to be taken over all non-zero B with (9) and with $B(\boldsymbol{\beta}) = 0$.

1.3. The results. First about solutions of linear differential equations:

THEOREM 1. *Suppose $\alpha \in \Omega(l)$, but $\notin \Omega(l-1)$. Then there is a $c_4(\alpha) > 0$ such that*

(10) $$|\alpha - \beta| > c_4(\alpha)\mathfrak{H}_l(\beta)^{-l-1}$$

for every $\beta \neq \alpha$ in $\Omega(l)$.

In general we have

THEOREM 2. *Suppose $m \geqslant 0$, $s \geqslant 1$ are given. Suppose α satisfies a non-trivial differential equation $A(\boldsymbol{\alpha}) = 0$, but satisfies no non-trivial differential equation of order $< m$. Then for $\beta \neq \alpha$, $\beta \in \Omega(m, s)$, we have*

(11) $$|\alpha - \beta| > c_5 \mathfrak{H}_{ms}(\beta)^{-c_6},$$

where $c_5 = c_5(\alpha, m, s) > 0$ and where

$$c_6 = 9^{m+2}\big(r(A)s\big)^{2m+1-1}.$$

No special importance attaches to our value of the constant c_6. The restriction that α satisfies no differential equation of order $< m$ could be removed at the cost of further complication and a possible change of our value for c_6. Since $\Omega(0, 1)$ consists of rational functions, the case $m = 0$, $s = 1$ is Kolchin's Theorem, except for the just mentioned restriction and except for the value of c_6.

Both Theorems 1 and 2 are Liouville type results. We have as yet no Roth type results.

The reader will not be surprised to hear that resultants play an essential role in the proof. More surprising is the fact that existence theorems on power series solutions of differential equations are needed.

2. Differential equations [1]

2.1. Notation. Suppose $A = A(Y) = A(Y, Y_1, \ldots, Y_l)$ is a differential polynomial. The *derivative* $A^{(1)}$ of A is defined by

$$A^{(1)} = \frac{\partial A}{\partial X} + \frac{\partial A}{\partial Y} Y_1 + \frac{\partial A}{\partial Y_1} Y_2 + \ldots + \frac{\partial A}{\partial Y_l} Y_{l+1}.$$

All the usual rules on differentiation of sums and products hold. The higher derivatives $A^{(2)}, A^{(3)}, \ldots$ are defined by induction. It is easily seen that for $\alpha \in K$ we have

$$A^{(1)}(\boldsymbol{\alpha}) = \big(A(\boldsymbol{\alpha})\big)^{(1)}.$$

Given two differential polynomials A, B, put

$$A \circ B = A(B, B^{(1)}, \ldots).$$

Then $(A \circ B) \circ C = A \circ (B \circ C)$, and in particular $(A \circ B)(\boldsymbol{\alpha}) = A\big(B(\boldsymbol{\alpha})\big)$.

In the introduction we used the multiplicative valuation $|\alpha|$, in order to stress the analogy with ordinary diophantine approximation. In what follows, it will be more convenient to use the additive function v with $v(0) = -\infty$ and

$$v(\alpha) = k$$

if α is given by (1) with $a_k \neq 0$. Clearly $|\alpha| = 2^{v(\alpha)}$. We have $v(\alpha + \beta) \leqslant \max\big(v(\alpha), v(\beta)\big)$, and equality holds here if $v(\alpha) \neq v(\beta)$. Not v itself, but $-v$ is an "additive valuation".

2.2. Linear polynomials. Given a homogeneous linear differential polynomial

$$L = \lambda_0 Y + \lambda_1 Y_1 + \ldots + \lambda_l Y_l,$$

put $w(L) = -\infty$ if $L = 0$ and

$$w(L) = \max_{0 \leqslant i \leqslant l}\big(v(\lambda_i) - i\big)$$

otherwise. Then certainly $w(L_1 + L_2) \leqslant \max\big(w(L_1), w(L_2)\big)$, with equality if $w(L_1) \neq w(L_2)$. Given a non-zero homogeneous linear differential polynomial L with $w(L) = w$, we may write

(12) $$L = L_w + L_{w-1} + \ldots,$$

where

(13) $$L_j = a_{j0} X^j Y + a_{j1} X^{j+1} Y_1 + \ldots + a_{jl} X^{j+l} Y_l \quad (j = w, w-1, \ldots)$$

with constant coefficients, and where $L_w \neq 0$.

---

[1] The lemmas of this section are not new, but are collected here for convenience See e.g. [4].

Substituting $a = X^t$, we obtain

(14) $$L_j(X^t) = (a_{j0}+a_{j1}t+a_{j2}t(t-1)+\ldots+a_{jl}t(t-1)\ldots(t-l+1))X^{t+j} = p_{Lj}(t)X^{t+j},$$

say, with certain polynomials p_{Lj} of degree $\leqslant l$. The polynomial $p_L = p_{Lw}$ is not zero and is the *indicial polynomial* of L. The indicial polynomial of $L = 0$ is identically zero. It follows from (12), (14) that

(15) $$w(L) = \max_t \big(v(L(X^t)) - v(X^t)\big),$$

hence that

(16) $$w(L) = \max_a \big(v(L(a)) - v(a)\big),$$

and that in fact

(17) $$w(L) = v(L(a)) - v(a)$$

if $p_L(v(a)) \neq 0$.

Lemma 1. *Let L, M be homogeneous linear differential polynomials. Then $M\circ L$ is a homogeneous linear differential polynomial with*

(18) $$l(M\circ L) = l(M)+l(L),$$

(19) $$w(M\circ L) = w(M)+w(L),$$

(20) $$p_{M\circ L}(t) = p_L(t)p_M(t+w(L)).$$

Proof. We may suppose that L, M are non-zero. Only (19), (20) require a proof. By (12), (14),

$$L(X^t) = p_L(t)X^{t+w(L)} + p_{L,w(L)-1}(t)X^{t+w(L)-1}+\ldots$$

and

$$M(X^u) = p_M(u)X^{u+w(M)} + p_{M,w(M)-1}(u)X^{u+w(M)-1}+\ldots$$

Thus

$$(M\circ L)(X^t) = p_L(t)p_M(t+w(L))X^{t+w(L)+w(M)}+\ldots,$$

and (19), (20) become obvious.

2.3. Solutions of differential equations. Every differential polynomial A may uniquely be written as

(21) $$A = \lambda + L + \bar{A},$$

where $\lambda \in K$, where L is homogeneous and linear, and $\bar{A}$ is a sum

(22) $$\bar{A} = \sum \pi_u P_u,$$

with differential monomials P_u having $s(P_u) \geqslant 2$. Given

$$\pi_u P_u = \pi_u Y^{a_0} Y_1^{a_1} \ldots Y_l^{a_l},$$

write $w(\pi_u P_u) = v(\pi_u) - a_1 - 2a_2 - \ldots - la_l$, and set $w(\bar{A})$ for the maximum of $w(\pi_u P_u)$ over all summands in (22), with the understanding that $w(\bar{A}) = -\infty$ if $\bar{A} = 0$. Now (recall the definition of $s(P)$ in § 1.2),

$$v(\pi_u P_u(X^t)) \leqslant ts(P_u) + w(\pi_u P_u);$$

and if $t \leqslant 0$, $s(P_u) \geqslant 2$, then

$$v(\pi_u P_u(X^t)) \leqslant 2t + w(\pi_u P_u) = 2v(X^t) + w(\pi_u P_u).$$

Thus if $t \leqslant 0$, then

$$v(\bar{A}(X^t)) \leqslant 2v(X^t) + w(\bar{A}),$$

and, more generally, if $v(a) \leqslant 0$, then

(23) $$v(\bar{A}(a)) \leqslant 2v(a) + w(\bar{A}).$$

Lemma 2. *Suppose A is of the type (21) with $\lambda = 0$, $L \neq 0$. Suppose $\eta \in K$, $\eta \neq 0$, is a solution of the differential equation*

$$A(\eta) = 0.$$

Suppose that either $\bar{A} = 0$, or that $\bar{A} \neq 0$ and

(24) $$v(\eta) < \min(0, w(L) - w(\bar{A})).$$

Then $p_L(v(\eta)) = 0$.

Proof. Suppose we had $p_L(v(\eta)) \neq 0$. If $\bar{A} = 0$, then by (17),

$$-\infty = v(0) = v(L(\eta)) = v(\eta) + w(L) \neq -\infty$$

which is impossible. If $\bar{A} \neq 0$ and (24) holds, then $L(\eta) + \bar{A}(\eta) = 0$, whence by (17), (23),

$$v(\eta) + w(L) = v(L(\eta)) = v(\bar{A}(\eta)) \leqslant 2v(\eta) + w(\bar{A}) < v(\eta) + w(L),$$

which again is impossible.

2.4. Existence of solutions.

Lemma 3. *Suppose L is a non-zero homogeneous linear differential polynomial with $w(L) = w$. For a fixed t, consider series*

(25) $$\eta = a_t X^t + a_{t-1}X^{t-1} + \ldots$$

with undetermined coefficients. Then

$$L(\eta) = b_{t+w}X^{t+w} + b_{t+w-1}X^{t+w-1} + \ldots,$$

with

$$b_{t+w-j} = p_L(t-j)a_{t-j} + \hat{b}_{t+w-j},$$

where $\hat{b}_{t+w-j}$ is a linear form in $a_t, a_{t-1}, \ldots, a_{t-j+1}$. (In particular, $\hat{b}_{t+w} = 0$.)

Proof. It suffices to observe that

(a) every coefficient in the series $L(a_tX^t+\dots a_{t-j+1}X^{t-j+1})$ is a linear form in $a_t,\dots,a_{t-j+1}$,

(b) $L(a_{t-j}X^{t-j}+a_{t-j-1}X^{t-j-1}+\dots)=p_L(t-j)a_{t-j}X^{t+w-j}+\dots$

LEMMA 4. *Suppose $\bar{A}$ is a differential polynomial with summands π_uP_u having $s(P_u)\geqslant 2$. Let t, w be given with*

$$t<\min\bigl(0,w-w(\bar{A})\bigr). \tag{26}$$

Consider series (25) *with undetermined coefficients. Then*

$$\bar{A}(\eta)=c_{t+w}X^{t+w}+c_{t+w-1}X^{t+w-1}+\dots, \tag{27}$$

where c_{t+w-j} is a polynomial in $a_t,\dots,a_{t-j+1}$ with constant term zero. (In particular, $c_{t+w}=0$.)

Proof. We may suppose that $\bar{A}$ is of the special type

$$\bar{A}=X^uY_{i_1}\dots Y_{i_s}$$

with $w(\bar{A})=u-i_1-\dots-i_s$. Then the coefficients on the right hand side of (27) are forms in $a_t,a_{t-1},\dots$ of degree s. A typical summand of $\bar{A}(\eta)$ will be some constant times

$$a_{t-j_1}\dots a_{t-j_s}X^{u+(t-j_1-i_1)+\dots+(t-j_s-i_s)}.$$

The exponent here will equal $t+w-j$ if

$$t+w-j=st-j_1-\dots-j_s+w(\bar{A}).$$

Then by (26),

$$j_k\leqslant j_1+\dots+j_s=(s-1)t+w(\bar{A})-w+j\leqslant t+w(\bar{A})-w+j<j.$$

So c_{t+w-j} is a polynomial in $a_t,\dots,a_{t-j+1}$.

LEMMA 5. *Suppose A is given by* (21), *with $L\neq 0$, $\lambda\neq 0$. Suppose that $p_L(t)$ has no integer root $t\leqslant v(\lambda)-w(L)$. Suppose that either $\bar{A}=0$ or that $\bar{A}\neq 0$ and*

$$v(\lambda)<\min\bigl(w(L),2w(L)-w(\bar{A})\bigr). \tag{28}$$

Then there is an $\eta\in K$ with $v(\eta)=v(\lambda)-w(L)$ having

$$A(\eta)=0.$$

Proof. Put $w=w(L)$, $v=v(\lambda)$, $t=v-w$. Write η in the form (25). Then

$$L(\eta)+\bar{A}(\eta)=d_vX^v+d_{v-1}X^{v-1}+\dots,$$

with coefficients

$$d_{v-j}=p_L(t-j)a_{t-j}+\hat{d}_{v-j},$$

where $\hat{d}_{v-j}$ is a polynomial in $a_t,\dots,a_{t-j+1}$. This follows from Lemma 3 if $\bar{A}=0$, and it follows from Lemmas 3, 4 if $\bar{A}\neq 0$. In the latter case observe that the condition (26) is satisfied in view of (28). By our hypothesis on p_L, the coefficient $p_L(t-j)$ of a_{t-j} is non-zero for $j=0,1,\dots$ We thus can successively choose $a_t,a_{t-1},\dots$ such that $L(\eta)+\bar{A}(\eta)=-\lambda$, i.e. that $A(\eta)=0$. Since $v(\lambda)=v$, we have $d_v=p_L(t)a_t\neq 0$, whence $a_t\neq 0$, and $v(\eta)=t=v-w$.

3. Approximation by solutions of linear differential equations

3.1. Linear differential ideals. The linear differential polynomials

$$A=\lambda+L=\lambda+\lambda_0Y+\lambda_1Y_1+\dots+\lambda_lY_l$$

form a vector space V over K. Given $l\geqslant -1$, the linear differential polynomials of order $\leqslant l$ form a subspace V_l of V of dimension $l+2$ with basis $1,Y,\dots,Y_l$. A *linear differential ideal* is defined as a subspace of V which is closed under taking derivatives and which is equal to V if it contains a non-zero element of order -1. The *principal ideal* (A) generated by $A\neq 0$ is the intersection of the linear differential ideals containing A. If $l(A)\geqslant 0$, then (A) is the subspace of V spanned by $A,A^{(1)},A^{(2)},\dots$ Hence (A) consists of the polynomials $L\circ A$, where L is a homogeneous linear differential polynomial. If $l(A)=-1$, then $(A)=V$.

LEMMA 6. *Every non-zero linear differential ideal is a principal ideal.*

Proof. Given a non-zero ideal $\mathfrak{I}$, let A be a non-zero element of $\mathfrak{I}$ of least order; say $l(A)=l$. If $l=-1$ then $\mathfrak{I}=V$ and thus $\mathfrak{I}=(A)$. So suppose that $l\geqslant 0$. Let B be an arbitrary non-zero element of $\mathfrak{I}$ with $l(B)=m$. The polynomials

$$B,A,A^{(1)},\dots,A^{(m-l)} \tag{29}$$

span a subspace S of V_m. Since A with $l(A)=l$ is a non-zero element of $\mathfrak{I}$ with least possible order, the intersection $S\cap V_{l-1}=0$. So $\dim S+\dim V_{l-1}=\dim(S+V_{l-1})\leqslant\dim V_m$, and $\dim S\leqslant\dim V_m-\dim V_{l-1}=m-l+1$. We conclude that the vectors (29) are linearly dependent. Since $A,A^{(1)},\dots$ are linearly independent, B must be linear combination of $A,A^{(1)},\dots,A^{(m-l)}$. Therefore $B\in(A)$ and $\mathfrak{I}=(A)$.

If $(A)=(B)$, then clearly $l(A)=l(B)$. Thus the order $l(\mathfrak{I})$ of a linear differential ideal $\mathfrak{I}$ may be defined as the order of any of its generators.

Write (A,B) for the ideal generated by A,B.

3.2. Differential resultants. Suppose $A\neq 0$, $B\neq 0$ generate a differential ideal $\mathfrak{I}$. Put $l=l(A)$, $m=l(B)$, $r=l(\mathfrak{I})$, and suppose that

$$r<l,\quad r<m. \tag{30}$$

If

$$A = \lambda + \lambda_0 Y + \ldots + \lambda_l Y_l,$$

then

$$A^{(i)} = \lambda^{[i]} + \lambda_0^{[i]} Y + \ldots + \lambda_{l+i-1}^{[i]} Y_{l+i-1} + \lambda_l Y_{l+i} \quad (i = 1, 2, \ldots),$$

where the $\lambda^{[i]}$ and the $\lambda_j^{[i]}$ are certain linear combinations of λ and the λ_j's and their derivatives. Similarly, if

$$B = \mu + \mu_0 Y + \ldots + \mu_m Y_m,$$

then

$$B^{(i)} = \mu^{[i]} + \mu_0^{[i]} Y + \ldots + \mu_{m+i-1}^{[i]} Y_{m+i-1} + \mu_m Y_{m+i} \quad (i = 1, 2, \ldots).$$

The determinant $R(Y)$ with $l+m-2r$ rows and columns, given by

$$R(Y) = \begin{vmatrix} A(Y) & \lambda_{r+1} & \ldots & \lambda_l & & & \\ A^{(1)}(Y) & \lambda_{r+1}^{[1]} & \ldots & \lambda_l^{[1]} & \lambda_l & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ A^{(m-r-1)}(Y) & \lambda_{r+1}^{[m-r-1]} & \ldots & \lambda_l^{[m-r-1]} & \ldots & \lambda_{l+m-r-2}^{[m-r-1]} & \lambda_l \\ B(Y) & \mu_{r+1} & \ldots & \mu_m & & & \\ B^{(1)}(Y) & \mu_{r+1}^{[1]} & \ldots & \mu_m^{[1]} & \mu_m & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ B^{(l-r-1)}(Y) & \mu_{r+1}^{[l-r-1]} & \ldots & \mu_m^{[l-r-1]} & \ldots & \mu_{l+m-r-2}^{[l-r-1]} & \mu_m \end{vmatrix}$$

is a linear differential polynomial called the *resultant* of A and B.

LEMMA 7. *The resultant $R(Y)$ is a non-zero linear differential polynomial of order r. The ideal $\mathfrak{J} = (A, B)$ is generated by R: $\mathfrak{J} = (R)$.*

Proof. The resultant is a linear combination of

$$(31) \qquad A, A^{(1)}, \ldots, A^{(m-r-1)}, B, B^{(1)}, \ldots, B^{(l-r-1)},$$

hence belongs to $\mathfrak{J}$. Being a linear combination of the polynomials (31), it is a linear combination of $1, Y, \ldots, Y_{l+m-r-1}$. But from our definition of $R(Y)$ as a determinant and from elementary linear algebra it follows that the coefficients of $Y_{r+1}, \ldots, Y_{l+m-r-1}$ are in fact zero, so that $l(R) \leqslant r$. It will thus suffice to show that $R \neq 0$; since $R \in \mathfrak{J}$, this will guarantee that $l(R) = r$ and that in fact $\mathfrak{J} = (R)$.

Now for $s \geqslant l, m$, the $2s-l-m+2$ polynomials

$$(32) \qquad A, A^{(1)}, \ldots, A^{(s-l)}, B, B^{(1)}, \ldots, B^{(s-m)}$$

lie in the vector space V_s of dimension $s+2$. So for large s, the vectors (32) will be linearly dependent; let s be the smallest integer with this property. We may then suppose that $A^{(s-l)}$ is a linear combination of $A, A^{(1)}, \ldots, A^{(s-l-1)}, B, B^{(1)}, \ldots, B^{(s-m)}$. Then

$$A, A^{(1)}, \ldots, A^{(s-l-1)}, B, B^{(1)}, \ldots$$

are linearly independent and span $\mathfrak{J} = (A, B)$. In fact, if t is large, then $\mathfrak{J} \cap V_t$ is spanned by

$$A, A^{(1)}, \ldots, A^{(s-l-1)}, B, B^{(1)}, \ldots, B^{(t-m)},$$

so that $\dim(\mathfrak{J} \cap V_t) = s-l+t-m+1$. On the other hand $\mathfrak{J} = (C)$ for some C with $l(C) = r$, so that $\mathfrak{J} \cap V_t$ is spanned by $C, C^{(1)}, \ldots, C^{(t-r)}$ if $r \geqslant 0$, and by $1, Y, \ldots, Y_t$ if $r = -1$, and $\dim(\mathfrak{J} \cap V_t) = t+1-r$. Comparing our formulae we find that

$$(33) \qquad s = l+m-r.$$

The polynomials

$$(34) \qquad A, A^{(1)}, \ldots, A^{(s-l-1)}, B, B^{(1)}, \ldots, B^{(s-m-1)}$$

are linearly independent by our minimal choice of s. They span a subspace of V_{s-1} of dimension $2s-l-m = s-r$, while V_r is of dimension $r+2$. Now $(s-r)+(r+2) = s+2 > \dim V_{s-1}$. So the subspace spanned by (34) has a non-zero intersection with V_r: There exists a non-zero linear combination D of (34) which lies in V_r. Since D must be a generator of I, it is unique except for a factor $\gamma \neq 0$. Since the polynomials (34) are linearly independent, there are coefficients $\alpha, \alpha_1, \ldots, \alpha_{s-l-1}$ and $\beta, \beta_1, \ldots, \beta_{s-m-1}$, not all zero, and unique except for a common factor $\gamma \neq 0$, such that

$$\alpha A + \alpha_1 A^{(1)} + \ldots + \alpha_{s-l-1} A^{(s-l-1)} + \beta B + \beta_1 B^{(1)} + \ldots + \beta_{s-m-1} B^{(s-m-1)}$$

lies in V_r. Now since $s-l-1 = m-r-1$ and $s-m-1 = l-r-1$, this uniqueness means precisely that the submatrix obtained from the determinant for $R(Y)$ by crossing out the first column, is of rank $l+m-2r-1$. So in R, which is a linear combination of (31), some coefficient of this linear combination is non-zero. Moreover, the polynomials (31) are the same as (34); hence they are linearly independent. So $R \neq 0$.

3.3. Proof of Theorem 1. Introduce the additive height

$$h(A) = \max_n v(f_n)$$

of a differential polynomial A given by (7). Then $\mathfrak{H}(A) = 2^{h(A)}$. Define $h_m(\beta)$ and $h_{ms}(\beta)$ in the obvious way. Theorem 1 now says that *if α lies in $\Omega(l)$ but not in $\Omega(l-1)$, then*

$$(35) \qquad v(\alpha - \beta) > -(l+1) h_l(\beta) - c_7(\alpha)$$

for every $\beta \neq \alpha$ in $\Omega(l)$.

Since α lies in $\Omega(l)$, it satisfies an equation $A(\boldsymbol{a}) = 0$ with a non-zero linear differential polynomial $A \in S[Y]$ having $l(A) \leqslant l$. Since α does not lie in $\Omega(l-1)$, we have in fact $l(A) = l$, and A is *unique* except for a non-zero factor in K_0. Say $A = \lambda + L$ with $L \neq 0$.

Suppose at first that $A(\boldsymbol{\beta}) = 0$. Then $\eta = \alpha - \beta$ satisfies $L(\boldsymbol{\eta}) = 0$, and $p_L(v(\eta)) = 0$ by Lemma 2. So this case is impossible if p_L has no integer root, and $v(\alpha-\beta) = v(\eta) \geqslant t_0$ where t_0 is the least integer root of p_L, if there are such integer roots.

We may therefore suppose that $A(\boldsymbol{\beta}) \neq 0$. The element $\beta \in \Omega(l)$ satisfies a non-trivial linear differential equation $B(\boldsymbol{\beta}) = 0$ with $B \in S[Y]$, with $m = l(B) \leqslant l$ and with $h(B) = h_l(\beta)$. Since $A(\boldsymbol{\beta}) \neq 0$, the polynomials A, B are not proportional, and $B(\boldsymbol{a}) \neq 0$. Now if, say, $B = \mu + M$, then $B(\boldsymbol{a}) = B(\boldsymbol{a}) - B(\boldsymbol{\beta}) = M(\boldsymbol{a} - \boldsymbol{\beta})$ and

$$v(B(\boldsymbol{a})) \leqslant v(\alpha-\beta) + h(M) \leqslant v(\alpha-\beta) + h(B).$$

So (35) will follow if we can prove that

$$v(B(\boldsymbol{a})) > -lh(B) - c'(\alpha). \tag{36}$$

Hence it remains to prove

THEOREM 1′. *Suppose $\alpha \in \Omega(l)$, $\alpha \notin \Omega(l-1)$. Then every linear differential polynomial B with $l(B) \leqslant l$ and $B(\boldsymbol{a}) \neq 0$ has* (36).

Now let $\mathfrak{I} = (A, B)$, and put $m = l(B)$, $r = l(\mathfrak{I})$. Since A, B are not proportional, $r < l = l(A)$. If $r < m$, let R be the differential resultant of A, B. If $r = m$, set $R = B$. In either case $R \in S[Y]$ and $(R) = \mathfrak{I}$. We observe that

$$v(R(\boldsymbol{a})) \leqslant v(B(\boldsymbol{a})) + (l-r-1)h(B) + (m-r)h(A).$$

This follows from $A(\boldsymbol{a}) = 0$ and the determinant formula for R if $r < m$, and is trivial if $r = m$. Now since $r \geqslant -1$ and since $(m-r)h(A) \leqslant c''(\alpha)$, Theorem 1′ will follow from

LEMMA 8.

$$v(R(\boldsymbol{a})) \geqslant \begin{cases} 0 & \textit{if } p_L \textit{ has no integer roots,} \\ \min(0, t_0 - l) & \textit{if } t_0 \textit{ is the least integer root of } p_L \end{cases}$$

To prove Lemma 8, we observe that R is a "constant", i.e. a polynomial in $S[X]$, if $r = -1$. Then $v(R(\boldsymbol{a})) = v(R) \geqslant 0$. We may thus suppose that $r \geqslant 0$. Write

$$R = \nu + N = \nu + \nu_0 Y + \nu_1 Y_1 + \ldots + \nu_r Y_r$$

with $N \neq 0$. Since $A \in \mathfrak{I} = (R)$, there is a homogeneous linear polynomial $Q \in K[Y]$ with

$$A = Q \circ R, \quad \text{whence with} \quad L = Q \circ N. \tag{37}$$

Consider the following linear differential equation in η:

$$N(\boldsymbol{\eta}) + R(\boldsymbol{a}) = 0. \tag{38}$$

The constant term $R(\boldsymbol{a}) \neq 0$ since $l(R) < l$. Applying Q we see that every solution η of (38) has

$$L(\boldsymbol{\eta}) = L(\boldsymbol{\eta}) + A(\boldsymbol{a}) = (Q \circ N)(\boldsymbol{\eta}) + (Q \circ R)(\boldsymbol{a}) = 0. \tag{39}$$

By Lemma 2, such an η has $p_L(v(\eta)) = 0$.

Observe that by Lemma 1, the indicial polynomial p_N of N is a divisor of the indicial polynomial p_L of L. If p_L has no integer roots, then neither does p_N. Then (38) has a solution η by (the case $\overline{A} = 0$ of) Lemma 5. This η has $p_L(v(\eta)) = 0$, which is impossible. So the case $r \geqslant 0$ is impossible if p_L has no integer roots.

If p_L has integer roots and if $v(R(\boldsymbol{a})) < t_0 - l$, then p_L and hence p_N has no integer root $\leqslant v(R(\boldsymbol{a})) + l$. Since $N \in S[Y]$, we have $w(N) \geqslant -l(N) = -r > -l$, and p_L and p_N have no integer roots $\leqslant v(R(\boldsymbol{a})) - w(N)$. Again by Lemma 5, the equation (38) has a solution η with $v(\eta) = v(R(\boldsymbol{a})) - w(N)$. Since again $p_L(v(\eta)) = 0$ by (39), we have again a contradiction.

4. Approximation by solutions of general differential equations

4.1. A reduction. Let m, s, α be as in Theorem 2. So α satisfies a non-trivial equation $A(\boldsymbol{a}) = 0$ with $A \in S[Y]$ and with $l = l(A) \geqslant m$, but it satisfies no such equation of an order less than m. We may suppose that A has the minimal possible value of $r(A)$ (the functional r is defined in § 1.2), and therefore $(\partial A/\partial Y_l)(\boldsymbol{a}) \neq 0$, since $r(\partial A/\partial Y_l) < r(A)$.

The inequality (11) of Theorem 2 which is to be proved says that

$$v(\alpha - \beta) > -c_6 h_{ms}(\beta) - c \tag{40}$$

for $\beta \in \Omega(m, s)$ distinct from α. (The constants c here and in the sequel are not necessarily equal, and they will depend only on α, m, s.)

Let us first suppose that $A(\boldsymbol{\beta}) = 0$. Then $\eta = \beta - \alpha$ has $A(\boldsymbol{a} + \boldsymbol{\eta}) = 0$, or $C(\boldsymbol{\eta}) = 0$, where $C \in K[Y]$ is given by

$$C(Y) = A(\boldsymbol{a} + Y). \tag{41}$$

Writing $C = \lambda + L + \overline{C}$ as in (21), we have $\lambda = A(\boldsymbol{a}) = 0$ and

$$L = \left(\frac{\partial A}{\partial Y}(\boldsymbol{a})\right)Y + \left(\frac{\partial A}{\partial Y_1}(\boldsymbol{a})\right)Y_1 + \ldots + \left(\frac{\partial A}{\partial Y_l}(\boldsymbol{a})\right)Y_l. \tag{42}$$

Observe that L depends only on α and that $L \neq 0$ since the coefficient of Y_l is non-zero. By Lemma 2 we have either $p_L(v(\eta)) = 0$ or $\overline{C} \neq 0$ and $v(\eta) \geqslant \min(0, w(L) - w(\overline{C}))$. So certainly $v(\eta) \geqslant -c$, or $v(\alpha-\beta) \geqslant -c$.

Now α may or it may not satisfy a differential equation $A_0(\boldsymbol{a}) = 0$ of order m. If so, we may suppose A_0 to be an irreducible polynomial in $F[X, Y, \ldots, Y_m]$. Since certainly $\alpha, \ldots, \alpha^{(m-1)}$ are algebraically independent over $F(X) = K_0$, every polynomial $\hat{A}_0 \in F[X, Y, \ldots, Y_m]$ with $\hat{A}_0(\boldsymbol{a}) = 0$ would have to be a multiple of A_0. One sees as above that $v(\alpha-\beta) \geqslant -c$ if $A_0(\boldsymbol{\beta}) = 0$ and $\beta \neq \alpha$.

$\beta \in \Omega(m, s)$ satisfies a non-trivial equation $B(\boldsymbol{\beta}) = 0$ with $B \in S[\boldsymbol{Y}]$, $l(B) \leqslant m$, $s(B) \leqslant s$ and $h(B) = h_{ms}(\beta)$. Since $s(B_1B_2) = s(B_1)+s(B_2)$ and $h(B_1B_2) = h(B_1)+h(B_2)$, we may suppose B to be an irreducible polynomial in $F[X, Y, \ldots, Y_m]$. Now $B(\boldsymbol{a}) = 0$ is only possible if the polynomial A_0 above really exists and B, A_0 are proportional. Then $A_0(\boldsymbol{\beta}) = 0$, and $v(\alpha-\beta) \geqslant -c$ by what we said above. We may thus assume that $B(\boldsymbol{a}) \neq 0$.

In order to prove (40), we may clearly suppose that $v(\alpha-\beta) < v(\alpha)$, whence that

$$v(\beta) = v(\alpha).$$

If a monomial P is given by (5), then $P(\boldsymbol{a})-P(\boldsymbol{\beta})$ is a sum of terms

$$(\alpha^{(i)}-\beta^{(i)})\,\alpha^{a_0}\ldots(\alpha^{(i-1)})^{a_{i-1}}\times$$
$$\times\big((\alpha^{(i)})^{a_i-1}+\ldots+(\beta^{(i)})^{a_i-1}\big)(\beta^{(i+1)})^{a_{i+1}}\ldots(\beta^{(l)})^{a_l},$$

so that

$$v\big(P(\boldsymbol{a})-P(\boldsymbol{\beta})\big) \leqslant v(\alpha-\beta)+v(\alpha)(a_0+\ldots+a_l-1).$$

Hence

$$v\big(B(\boldsymbol{a})\big) = v\big(B(\boldsymbol{a})-B(\boldsymbol{\beta})\big) \leqslant v(\alpha-\beta)+h(B)+c.$$

Clearly (40) would follow from

$$v\big(B(\boldsymbol{a})\big) \geqslant -(c_6-1)h(B)-c.$$

Hence Theorem 2 is reduced to

THEOREM 2'. *Suppose $m \geqslant 0$, $s \geqslant 1$ are given. Suppose α satisfies a non-trivial equation $A(\boldsymbol{a}) = 0$ with $l = l(A) \geqslant m$, but satisfies no equation of order $< m$. Then every $B \in S[\boldsymbol{Y}]$ with $l(B) \leqslant m$, $s(B) \leqslant s$ and with $B(\boldsymbol{a}) \neq 0$ has*

$$v\big(B(\boldsymbol{a})\big) \geqslant -c_7h(B)-c \tag{43}$$

with

$$c_7 = (9^{m+2}-1)\big(r(A)s\big)^{2^{m+1}-1} \leqslant c_6-1. \tag{44}$$

In fact, Theorem 2 for a particular value of m follows from Theorem 2' with this particular value of m.

Theorem 2' will be proved by induction on m. Now if $B \in S[\boldsymbol{Y}]$ has $l(B) = -1$ and $B(\boldsymbol{a}) \neq 0$, then it is a non-zero constant, i.e. a non-zero polynomial in X, hence has $v\big(B(\boldsymbol{a})\big) = v(B) \geqslant 0$. Hence Theorem 2' is true for $m = -1$, with c_7 replaced by 0. Thus our inductive argument is off the ground, and we may suppose the theorem to be true for $m-1$.

4.2. An application of resultants. Let $B = B(Y, \ldots, Y_m)$ and $E = E(Y, \ldots, Y_m)$ be differential polynomials in $S[\boldsymbol{Y}]$ of respective degrees $b > 0$ and $e > 0$ in Y_m. Clearly $b \leqslant s(B)$ and $e \leqslant s(E)$. Suppose that B, E have no common factor of positive degree in $Y, \ldots, Y_m$. Then if we interpret them as polynomials in Y_m with coefficients in the field $F(X, Y, \ldots, Y_{m-1})$, they have no common factor of positive degree in Y_m. Hence their resultant $R = R(Y, \ldots, Y_{m-1})$ will not be 0. (In contrast to § 3, we are now dealing with an ordinary resultant, not a differential resultant.) The following facts follow from the theory of resultants:

$$s(R) \leqslant s(B)e+s(E)b-eb \leqslant s(B)s(E). \tag{45}$$

(Namely, each of the $(b+e)!$ summands in the determinant of order $b+e$ for R can be estimated in this way.)

$$h(R) \leqslant h(B)e+h(E)b \leqslant h(B)s(E)+h(E)s(B). \tag{46}$$

There exist polynomials $U(Y, \ldots, Y_m)$ and $V(Y, \ldots, Y_m)$ in $S[\boldsymbol{Y}]$ with

$$R = UB+VE, \tag{47}$$

having

$$s(U) \leqslant s(B)s(E)-s(B), \quad s(V) \leqslant s(B)s(E)-s(E), \tag{48}$$

$$\begin{aligned} h(U) &\leqslant h(B)s(E)+h(E)s(B)-h(B),\\ h(V) &\leqslant h(B)s(E)+h(E)s(B)-h(E). \end{aligned} \tag{49}$$

LEMMA 9. *Let $B \in S[\boldsymbol{Y}]$ be given with $l(B) = m$, $s(B) \leqslant s$, and irreducible as a polynomial in $F[X, Y, \ldots, Y_m]$. Put*

$$B_* = \partial B/\partial Y_m. \tag{50}$$

Then if α is as in Theorem 2', we have

$$\max\Big(v\big(B(\boldsymbol{a})\big),\ v\big(B_*(\boldsymbol{a})\big)\Big) > -c_8h(B)-c \tag{51}$$

with

$$c_8 = 2\cdot 9^{m+1}s\big(r(A)s^2\big)^{2^m-1} = 2sc_6\big(r(A), m-1, s^2\big). \tag{52}$$

Proof. Let R be the resultant of B and $E = B_*$. Applying the estimates on the resultant just given, we obtain

$$s(R) \leqslant s^2, \quad h(R) \leqslant 2sh(B), \tag{53}$$

and $R = UB+VB_*$ with

$$s(U), s(V) \leqslant c,$$
$$h(U), h(V) \leqslant (2s-1)h(B) < 2sh(B).$$

It follows that $v(U(a))$ and $v(V(a))$ are $< 2sh(B)+c$, whence that

$$v(R(a)) \leqslant \max\big(v(B(a)),\, v(B_*(a))\big)+2sh(B)+c. \tag{54}$$

Since $l(R) < m$, our inductive hypothesis yields

$$\begin{aligned} v(R(a)) &> -c_7(r(A),\, m-1,\, s(R))h(R)-c \\ &\geqslant -c_7(r(A),\, m-1,\, s^2)h(R)-c, \end{aligned}$$

whence by (53), (54),

$$\begin{aligned} \max\big(v(B(a)),\, v(B_*(a))\big) &\geqslant -\big(1+c_7(r(A),\, m-1,\, s^2)\big)(2s)h(B)-c \\ &\geqslant -c_6(r(A),\, m-1,\, s^2)(2s)h(B)-c \\ &\geqslant -c_8 h(B)-c. \end{aligned}$$

In our proof of Theorem 2′ we may always suppose that $l(B) = m$ and that B is irreducible. For both $h(B)$ and $v(B(a))$ are *additive* functions of B, i.e. $h(B_1B_2) = h(B_1)+h(B_2)$ and $v(B_1B_2(a)) = v(B_1(a))+ +v(B_2(a))$. Since Theorem 2′ is certainly true if $v(B(a)) > -c_8 h(B)-c$, we may suppose that

$$v(B_*(a)) > -c_8 h(B)-c. \tag{55}$$

4.3. Manipulations with polynomials. Let $m \geqslant 0$ be fixed. If P is a differential monomial given by (5), put $r_m(P) = 0$ if $l \leqslant m$, and $r_m(P) = a_{m+1}+3a_{m+2}+\ldots+(2l-2m-1)a_l$ otherwise. If A is a differential polynomial given by (6) with non-zero coefficients π_n, write $r_m(A)$ for the maximum of $r_m(P_n)$. Write $r_m(A) = 0$ if $A = 0$. Observe that $r_m(A) \leqslant r(A)$ and that

$$r_m(A^{(1)}) \leqslant r_m(A)+2.$$

If $B \in S[Y]$ with $l(B) = m$, then as in § 4.2 put $B_* = \partial B/\partial Y_m$. Then $h(B_*) \leqslant h(B)$ and $s(B_*) \leqslant s(B)-1$.

LEMMA 10. *Given $B \in S[Y]$ with $l(B) = m$, we have*

$$B^{(j)} = H_j(Y, \ldots, Y_{m+j-1})+B_* Y_{m+j} \quad (j = 1, 2, \ldots),$$

where $H_j \in S[Y]$ with

$$s(H_j) \leqslant s(B), \quad h(H_j) \leqslant h(B), \tag{56}$$

$$r_m(H_j) \leqslant 2j-2. \tag{57}$$

Proof. Observe that $B^{(1)} = H_1(Y, \ldots, Y_m)+B_* Y_{m+1}$ with

$$H_1 = \frac{\partial B}{\partial X}Y+\frac{\partial B}{\partial Y}Y_1+\ldots+\frac{\partial B}{\partial Y_{m-1}}Y_m.$$

Clearly $s(H_1) \leqslant s(B)$ and $h(H_1) \leqslant h(B)$. Since $Y_{m+1}, \ldots$ do not occur in H_1, we have $r_m(H_1) = 0$. Finally, since $H_1 \in S[Y]$, the lemma is true for $j = 1$. Assuming its truth for j, we have

$$B^{(j+1)} = H_{j+1}(Y, \ldots, Y_{m+j})+B_* Y_{m+j+1}$$

with

$$H_{j+1} = H_j^{(1)}+B_*^{(1)} Y_{m+j}.$$

So

$$s(H_{j+1}) \leqslant s(B), \quad h(H_{j+1}) \leqslant h(B)$$

and

$$r_m(H_{j+1}) \leqslant \max\big(r_m(H_j)+2,\, r_m(B_*^{(1)})+2j-1\big) \leqslant 2j.$$

Since $H_{j+1} \in S[Y]$, the lemma is established for $j+1$.

LEMMA 11. *Let m, n, l be integers with $m \geqslant 0$, $n \geqslant 1$ and $m+n = l$. Let B with $l(B) = m$ and $A = A(Y, \ldots, Y_l)$ be differential polynomials in $S[Y]$, and let $r \geqslant r_m(A)$. Then*

$$\begin{aligned} B_*^r A = D(Y, \ldots, Y_m)+C_1(Y, \ldots, Y_{m+1})B^{(1)}+\ldots+ \\ +C_n(Y, \ldots, Y_l)B^{(n)}, \end{aligned} \tag{58}$$

with polynomials $D, C_1, \ldots, C_n$ in $S[Y]$ having

$$\left.\begin{matrix} s(D) \\ s(C_iB^{(i)}) \end{matrix}\right\} \leqslant r\big(s(B)-1\big)+s(A) \quad (= s_0,\ say), \tag{59}$$

$$\left.\begin{matrix} h(D) \\ h(C_iB^{(i)}) \end{matrix}\right\} \leqslant rh(B)+h(A) \quad (= h_0,\ say). \tag{60}$$

Proof. We proceed by induction on n, and note that a degenerate version of the lemma is true for $n = 0$. Let $\mathfrak{A}(s, h)$ be the set of expressions of the type of the right hand side of (58), with $D, C_1, \ldots, C$ in $S[Y]$ having

$$\begin{aligned} \max\big(s(D),\, s(C_1B^{(1)}),\, \ldots,\, s(C_nB^{(n)})\big) \leqslant s, \\ \max\big(h(D),\, h(C_1B^{(1)}),\, \ldots,\, h(C_nB^{(n)})\big) \leqslant h. \end{aligned}$$

Then we have to show that $B_*^r A$ lies in $\mathfrak{A}(s_0, h_0)$. Since $\mathfrak{A}(s_0, h_0)$ is closed under addition, it will suffice to show that for every "monomial"

$$M = C(Y, \ldots, Y_m)Y_{m+1}^{a_{m+1}}\ldots Y_l^{a_l}$$

with $C \in S[Y]$ having $h(C) \leqslant h(A)$ and $s(C) \leqslant s(A)-a_{m+1}-\ldots-a_l$, and with $r_m(M) = a_{m+1}+\ldots+(2l-2m-1)a_l \leqslant r$, the product $B_*^r M$ lies in $\mathfrak{A}(s_0, h_0)$. Observe that

$$B_*^r M = B_*^{r-r_m(M)}C\prod_{i=1}^{n}(B_*^{2i-1}Y_{m+i})^{a_{m+i}}.$$

Clearly

(61) $B_*^{r-r_m(M)}C \in \mathfrak{A}\big((r-r_m(M))(s(B)-1)+s(C),\ (r-r_m(M))h(B)+h(C)\big)$.

We shall show that for $i = 1, \ldots, n$,

(62) $B_*^{2i-1}Y_{m+i} \in \mathfrak{A}\big((2i-1)(s(B)-1)+1,\ (2i-1)h(B)\big)$.

Now since
$$\mathfrak{A}(s_1, h_1)\mathfrak{A}(s_2, h_2) \subseteq \mathfrak{A}(s_1+s_2, h_1+h_2),$$
i.e. the product of elements in $\mathfrak{A}(s_1, h_1)$ and in $\mathfrak{A}(s_2, h_2)$ lies in $\mathfrak{A}(s_1+s_2, h_1+h_2)$, it will follow from (61), (62) that $B_*^r M \in \mathfrak{A}(s_0, h_0)$. If $n > 1$, then (62) for $i = 1, \ldots, n-1$ is true by induction. It will therefore suffice to prove (62) for $i = n$. By Lemma 10,
$$B_*^{2n-1}Y_{m+n} = B_*^{2n-2}B^{(n)} - B_*^{2n-2}H_n(Y, \ldots, Y_{m+n-1}).$$
The first summand on the right lies in
$$\mathfrak{A}\big((2n-1)(s(B)-1)+1,\ (2n-1)h(B)\big).$$
But so does the second, by virtue of $r_m(H_n) \leqslant 2n-2$ and our induction on n. The lemma is proved.

4.4. Two cases. As in § 4.1, let α satisfy $A(\alpha) = 0$ with $A \neq 0$, $A \in S[Y]$, and let it satisfy no equation of order less than m. Since $s(B_1B_2) = s(B_1)+s(B_2)$ and $h(B_1B_2) = h(B_1)+h(B_2)$, it will suffice to prove Theorem 2′ for *irreducible* polynomials B with $l(B) = m$. In what follows, B will be such a polynomial. As usual, put $l = l(A)$.

Now either

(i) $m < l$, so that $l = m+n$ with $n \geqslant 1$. Then set $r = r(A)$ and construct $D(\boldsymbol{Y}) = D(Y, \ldots, Y_m)$ as in Lemma 11. Or

(ii) $m = l$. Then put $D = A$.

In case (i), the polynomials C_i of Lemma 11 have $s(C_i) \leqslant c$ and $h(C_i) \leqslant r(A)h(B)+h(A) \leqslant r(A)h(B)+c$, so that $v(C_i(\alpha)) \leqslant r(A)h(B)+c$. Now since $A(\alpha) = 0$ and since $v(B^{(i)}(\alpha)) \leqslant v(B(\alpha))$, (58) yields

(63) $v(D(\alpha)) \leqslant v(B(\alpha))+r(A)h(B)+c$.

We also note that by (59), (60),

(64) $s(D) \leqslant r(A)s(B)$ and $h(D) \leqslant r(A)h(B)+c$.

Both (63) and (64) are trivially true in case (ii).

Again we distinguish two cases. Either $B \nmid D$, i.e. $B = B(X, Y, \ldots, Y_m)$ does not divide $D = D(X, Y, \ldots, Y_m)$. Or $B \mid D$.

4.5. The case $B \nmid D$. Since B was irreducible, the polynomials B, D have no common factor. We form the resultant of B and $E = D$, as explained in § 4.2. The equation (47) becomes

(65) $R = UB+VD$.

The polynomials U, V have

(66) $s(U) \leqslant c, \quad s(V) \leqslant c$

by (48), (64), and

(67) $h(U) \leqslant h(B)s(D)+h(D)s(B) \leqslant 2r(A)s(B)h(B)+c,$

(68) $h(V) \leqslant 2r(A)s(B)h(B)+c$

by (49), (64). We have $v(U(\alpha)) \leqslant h(U)+c$ and $v(V(\alpha)) \leqslant h(V)+c$ by (66). So by (63), (65), (67), (68), we obtain

(69)
$$v(R(\alpha)) \leqslant \max\big(v(U(\alpha))+v(B(\alpha)),\ v(V(\alpha))+v(D(\alpha))\big) \leqslant v(B(\alpha))+3r(A)s(B)h(B)+c.$$

Also note that

(70) $s(R) \leqslant s(B)s(D) \leqslant r(A)s^2(B)$

by (45), (64), and that

(71) $h(R) \leqslant 2r(A)s(B)h(B)+c$

by (46), (64).

Now $l(R) < m$. We may therefore apply our induction hypothesis to R and we obtain
$$v(R(\alpha)) > -c_7'h(R)-c,$$
with $c_7' = c_7(r(A), m-1, s(R)) \leqslant c_7(r(A), m-1, r(A)s^2)$. In conjunction with (69), (71), we obtain
$$v(B(\alpha)) \geqslant -c_9h(B)-c$$
with
$$c_9 = 2r(A)s(B)c_7'+3r(A)s(B) \leqslant c_7(r(A), m, s).$$

4.6. The case $B \mid D$. We now have $D = C_0(Y, \ldots, Y_m)B$, so that (58) becomes

(72)
$$B_*^r A = C_0(Y, \ldots, Y_m)B+C_1(Y, \ldots, Y_{m+1})B^{(1)}+ \ldots + C_n(Y, \ldots, Y_l)B^{(n)}.$$

This is certainly true in case (i), i.e. when $m < l$. In case (ii) we have $D = A$, so now B divides A, and (72) reduces to $B_*^r A = C_0 B$. In the polynomial identity (72) we substitute $\alpha+\boldsymbol{Y}$ for $\boldsymbol{Y}$, i.e. $\alpha+Y$, $\alpha^{(1)}+Y_1, \ldots$ for $Y, Y_1, \ldots$ We obtain

(73) $B_*^r(\alpha+\boldsymbol{Y})A(\alpha+\boldsymbol{Y}) = C_0(\alpha+\boldsymbol{Y})B(\alpha+\boldsymbol{Y})+ \ldots + C_n(\alpha+\boldsymbol{Y})B^{(n)}(\alpha+\boldsymbol{Y})$.

The differential polynomials on both sides of this equation are in $K[\boldsymbol{Y}]$ but not necessarily in $S[\boldsymbol{Y}]$.

Write the polynomial on either side of (73) as

$$\nu+N+\bar{E},$$

with the same conventions as in (21). Since $A(\alpha)=0$, we have $\nu=0$. Moreover, N is $B_*^r(\alpha)$ times the linear part of $A(\alpha+Y)$. If we write, as in (41), (42),

$$A(\alpha+Y)=L+\bar{C}=\left[\left(\frac{\partial A}{\partial Y}(\alpha)\right)Y+\ldots+\left(\frac{\partial A}{\partial Y_l}(\alpha)\right)Y_l\right]+\bar{C},$$

then

(74) $$N=B_*^r(\alpha)L.$$

Thus $N\neq 0$, since $L\neq 0$ and since $B_*(\alpha)\neq 0$ by (55).

In the notation (21), write

$$C_i(\alpha+Y)=C_i(\alpha)+L_i+\bar{C}_i,$$
$$B(\alpha+Y)=B(\alpha)+L_B+\bar{B},$$

where of course $C_i(\alpha)$, $B(\alpha)$ are constants (i.e. elements of K), where L_i, L_B are linear, and $\bar{C}_i$, $\bar{B}$ contain the non-linear terms. Now $B^{(i)}(\alpha+Y)=\big(B(\alpha+Y)\big)^{(i)}$, so that

$$B^{(i)}(\alpha+Y)=B^{(i)}(\alpha)+L_B^{(i)}+\bar{B}^{(i)}.$$

Further N, being the linear term of (73), may be written as

(75) $$N=N_1+N_2,$$

with

(76) $$N_1=C_0(\alpha)L_B+C_1(\alpha)L_B^{(1)}+\ldots+C_n(\alpha)L_B^{(n)},$$

(77) $$N_2=B(\alpha)L_0+B^{(1)}(\alpha)L_1+\ldots+B^{(n)}(\alpha)L_n.$$

In view of (74), $w(N)=rv\big(B_*(\alpha)\big)+w(L)$. (The functional w was defined in § 2.2). Here L and therefore $w(L)$ depends only on α, and $v\big(B_*(\alpha)\big)$ may be estimated by (55). So

(78) $$w(N)>-r(A)c_8h(B)-c.$$

On the other hand,

$$L_i=\left(\frac{\partial C_i}{\partial Y}(\alpha)\right)Y+\ldots+\left(\frac{\partial C_i}{\partial Y_l}(\alpha)\right)Y_l.$$

By (59), (60), we have $s(C_i)\leqslant c$ and $h(C_i)\leqslant r(A)h(B)+c$, whence

$$v\big((\partial C_i/\partial Y_j)(\alpha)\big)\leqslant r(A)h(B)+c, \quad \text{and} \quad w(L_i)\leqslant r(A)h(B)+c.$$

So by (77),

$$w(N_2)\leqslant v\big(B(\alpha)\big)+r(A)h(B)+c.$$

Now either $w(N_2)\geqslant w(N)$. Then by (78),

$$v\big(B(\alpha)\big)>-r(A)(c_8+1)h(B)-c\geqslant -c_7h(B)-c,$$

and (43) is true.

Or $w(N_2)<w(N)$. Then by (75),

$$w(N_1)=w(N),$$

and N, N_1 have the same indicial polynomial p_N. Looking at (76) we see that

$$N_1=Q\circ L_B$$

with $Q=C_0(\alpha)Y+\ldots+C_n(\alpha)Y_n$. So by Lemma 1, the indicial polynomial p_{L_B} of L_B is a divisor of the indicial polynomial p_N of N_1. Moreover, $v\big(C_i(\alpha)\big)\leqslant r(A)h(B)+c$, whence $w(Q)\leqslant r(A)h(B)+c$, and

(79) $$w(L_B)=w(N_1)-w(Q)\geqslant w(N)-r(A)h(B)-c.$$

4.7. Solving a differential equation. Consider the following differential equation for η:

(80) $$B(\alpha+\eta)=0,$$

or

$$B(\alpha)+L_B(\eta)+\bar{B}(\eta)=0.$$

Now $w(\bar{B})\leqslant h(B)+c$, so that

$$\min=\min\big(w(L_B),\,2w(L_B)-w(\bar{B})\big)$$

has

$$\min>-(2c_8+3)r(A)h(B)-c$$

by (78), (79). Now if $\min\leqslant v\big(B(\alpha)\big)$, then (43) of Theorem 2′ holds. If not, then condition (28) of Lemma 5 holds. So there is a solution η of (80) with $v(\eta)=v\big(B(\alpha)\big)-w(L_B)$. Putting $\beta=\alpha+\eta$ we have

$$B(\beta)=0$$

and, by (78), (79),

(81) $$v(\alpha-\beta)=v(\eta)\leqslant v\big(B(\alpha)\big)+(c_8+1)r(A)h(B)+c.$$

Now obviously the right hand side of (72) vanishes if we substitute β, and hence either $A(\beta)=0$ or $B_*(\beta)=0$. If $A(\beta)=0$, then $v(\alpha-\beta)\geqslant -c$ by what we said in § 4.1, and (81) implies the desired (43) of Theorem 2′.

There remains the case $B_*(\beta)=0$. Then $R(\beta)=0$, where R is the resultant of B, B_* constructed in the proof of Lemma 9. Since $l(R)<m$, it follows from the case $m-1$ of Theorem 2 (which follows from the case

$m-1$ of Theorem 2′), that

$$\begin{aligned} v(\alpha-\beta) &\geqslant -c_6\big(r(A), m-1, s(R)\big)h(R)-c \\ &\geqslant -c_6\big(r(A), m-1, s^2\big)2sh(B)-c \end{aligned}$$

by (53). Together with (81), and observing (52), we obtain

$$v\big(B(\boldsymbol{a})\big) \geqslant -\big((1+c_8)r(A)+c_3\big)h(B)-c \geqslant -c_7h(B)-c.$$

We finally remark that we threw away the solution β of $B(\boldsymbol{\beta})=0$ in going from Theorem 2 to Theorem 2′. Then at the end we had to construct a solution β of $B(\boldsymbol{\beta})=0$. This may seem a wasteful argument. But in our inductive argument, we may have to replace B by a new B with a smaller value of $m=l(B)$. In other words, the solution β with $B(\boldsymbol{\beta})=0$ gets lost in the inductive argument.

References

[1] E. R. Kolchin, *Rational approximation to solutions of algebraic differential equations*, Proc. Amer. Math. Soc. 10 (1959), pp. 238–244.
[2] C. F. Osgood, *An effective lower bound on the "Diophantine Approximation" of Algebraic Functions by Rational Functions*, Mathematika 20 (1973), pp. 4–15.
[3] — *Concerning a possible "Thue-Siegel-Roth Theorem" for Algebraic differential equations* (to appear).
[4] E. G. C. Poole, *Introduction to the theory of linear differential equations*, New York 1960.
[5] K. F. Roth, *Rational approximations to algebraic numbers*, Mathematika 2 (1955), pp. 1–20.
[6] W. M. Schmidt, *Rational approximation to solutions of linear differential equations with algebraic coefficients*, Proc. Amer. Math. Soc. 53 (1975), pp. 285–289.
[7] S. Uchiyama, *Rational approximations to algebraic functions*, J. Fac. Sci. Hokkaido Univ. Ser. 15 (1961), pp. 173–192.

UNIVERSITY OF COLORADO, Boulder
UNIVERSITÄT WIEN, Österreich

Received on 12. 10. 1975

ACTA ARITHMETICA
XXXII (1977)

Large values of Dirichlet polynomials, IV

by

M. N. Huxley (Cardiff) and M. Jutila (Turku)

1. Introduction. This paper continues [4]–[6], [8], [9]. Our object is to estimate the size of a set of pairs (s, χ) at which a Dirichlet polynomial $F(s, \chi)$ can be large. A precise statement is given in the next section where the notation is introduced. Our main tool is the reflection argument of [4], which we use in a simplified form due to Jutila [8], [9] as Lemma 8 below. It relates Dirichlet polynomials of length N to those of length about D/N, where D measures the range in which the pairs (s, χ) can lie. It is useful to have a peak function which is itself a Dirichlet polynomial: we use the H series discussed in Sections 3 and 4, which are modified Dirichlet L-functions. It is sometimes possible to use $F(s, \chi)$ itself as a peak function, as in Lemma 10 below. The L-functions can be approximated by H series of length $D^{1/2}$ (the so-called approximate functional equation), as in Lemma 14 below. Lemma 14 is implicit in the literature; we sketch the proof out of duty. Jutila [9] has a new lemma (our Lemma 7) in which $F(s, \chi)$ is raised to an even integral power, and obtains sharper results than those of [6] when $F(s, \chi)$ is very large, for instance when the exponent a of (2.23) is $4/5$.

In this paper we explore the consequences of Jutila's new lemma. Our arguments are purely combinatoric (except Lemma 14). To make the work accessible, we have summarised the main ideas of previous papers as a sequence of lemmas, stressing the combinatory rather than the analytic aspects. Our result is Theorem 2 of Section 5. It enables us to improve the zero-density theorems for Dirichlet L-functions. For instance we extend the range of the density hypotheses. Let $N(\alpha, T, \chi)$ be the number of zeros $\beta+i\gamma$ of $L(s, \chi)$ in $\beta \geqslant \alpha$, $|\gamma| \leqslant T$. Then

$$\sum_{\chi \bmod q} N(\alpha, T, \chi) \ll (qT)^{2-2\alpha+\varepsilon} \tag{1.1}$$

holds for $\alpha > 4/5$. Let an asterisk denote a sum over proper characters. Then

$$\sum_{q \leqslant Q} \sideset{}{^*}\sum_{\chi \bmod q} N(\alpha, T, \chi) \ll (Q^2T)^{2-2\alpha+\varepsilon} \tag{1.2}$$