On the Thue-Siegel-Dyson theorem

by

ENRICO BOMBIERI

The Institute for Advanced Study Princeton, N.J., U.S.A. and Institut Mittag-Leffler, Djursholm, Sweden

I. Introduction

I.1. The well-known Thue-Siegel theorem, in the refined form obtained by Dyson and Gelfond, asserts that if α is an algebraic number of degree $r \geq 2$ and if $\varepsilon > 0$ then

$$\left|\alpha - \frac{p}{q}\right| > q^{-\sqrt{2r}-\varepsilon}$$

for $q \geq q_0(\alpha, \varepsilon)$. The constant $q_0(\alpha, \varepsilon)$ in this result turns out to be not effectively computable.

In fact, Thue proved a result of this type with the exponent $r/2+1$, Siegel improved this to the exponent $\min(r/(s+1)+s)$ for $s=0, 1, \ldots, r-1$ and finally, by using full freedom in the construction of the auxiliary polynomial, Dyson and Gelfond independently arrived at the exponent $\sqrt{2r}$.

The common feature in the approach of Thue, Siegel, Dyson and Gelfond is the consideration of two approximations $p_1/q_1, p_2/q_2$ to α and the construction of an auxiliary polynomial $p(x_1, x_2)$ with integral coefficients vanishing to a high order at (α, α) and vanishing only to a low order at $(p_1/q_1, p_2/q_2)$. Although Siegel and Schneider soon realized that further improvements could be obtained by the consideration of several distinct approximations $p_1/q_1, \ldots, p_m/q_m$ to α and by the construction of an auxiliary polynomial $P(x_1, \ldots, x_m)$ in many variables, it took about thirty years before Roth showed how to prove that P would vanish only to a low order at the point $(p_1/q_1, \ldots, p_m/q_m)$. In this way Roth was able to prove his celebrated theorem

$$\left|\alpha - \frac{p}{q}\right| > q^{-2-\varepsilon}$$

for $q \geq q_0(\alpha, \varepsilon)$.

I.2. All results mentioned above are ineffective in the sense that the method does not allow the calculation of $q_0(\alpha, \varepsilon)$. On the other hand, Thue himself noted that one could obtain a statement of the following kind ([T], Theorem III, p. 249).

THUE'S THEOREM. *Let α be algebraic of degree r and let h, k be given positive numbers. There is an effectively computable constant $G_0 = G_0(\alpha, h, k)$ such that if there exist p_0, q_0 with*

$$\left|\alpha - \frac{p_0}{q_0}\right| < q_0^{-\frac{r}{2}-1-k}, \quad q_0 > G_0$$

then we can determine effectively $G = G(\alpha, h, k, q_0)$ such that

$$\left|a - \frac{p}{q}\right| > q^{-\frac{r}{2(k+1)}-1-h}$$

for all $q > G$.

Gelfond ([G], Theorem 1, p. 22) obtained a similar result, but with the exponents ϑ, ϑ_1 in place of $r/2+1+k$ and $\frac{1}{2}r/(k+1)+1+h$, provided $2 \leqslant \vartheta \leqslant \vartheta_1 \leqslant r$ and $\vartheta\vartheta_1 = 2r(1+\varepsilon)$. He also related approximations to two different algebraic numbers in the same field, with essentially the same conclusions. It should be noted that because of a basic difference between the auxiliary constructions of Thue and Gelfond the result of Thue and the result of Gelfond are somewhat different. Further refinements are in Hyyrö ([H]).

The meaning of these results is simply that if there is an exceptionally good approximation to α then all other approximations cannot be too good from some point onward. Now the question is whether exceptionally good approximations exist in nature or not. As far as we can see, the constant G_0 in Thue's theorem, or the analogous constant in Gelfond's result, turns out to be far too large and as of today no pair $(\alpha, p_0/q_0)$ has been found which verifies the hypothesis $q_0 > G_0$ of Thue's theorem. This means that no effective result for the approximation of α by rationals has been found using Thue's or Siegel's or Roth's approach.

I.3. In order to illustrate one of the difficulties which appears let us consider the following problem.

Let $P(x)$ be a polynomial of degree d, with integral coefficients bounded by H. Let α be an algebraic number of degree r and let p/q be a rational number. Now suppose that P vanishes α to order $e(\alpha)$ and suppose also that P vanishes at p/q to order $e(p/q)$.

We would like to find conditions which make sure that $e(p/q)$ is small, even if $e(\alpha)$ is relatively large.

Let $e=e(p/q)$. Then $(qx-p)^e$ divides P and hence we must have $H\geqslant q^e$. This yields

$$e\left(\frac{p}{q}\right)\leqslant\frac{\log H}{\log q},$$

which shows that $e(p/q)$ is small if q is relatively large compared with H. On the other hand, let $e=e(\alpha)$ and let $f(x)$ be an irreducible polynomial with integral coefficients bounded by $H(\alpha)$ and defining α. Then we can write $P=f^eQ$ where Q has integral coefficients and now $H\geqslant C^{-d}H(\alpha)^e$ with an absolute constant C, by a basic result on heights. Thus we see that this argument cannot produce a bound better than

$$e(p/q)\leqslant e(\alpha)\frac{\log H(\alpha)}{\log q}+O\left(\frac{d}{\log q}\right).$$

If we disregard the $O(\ldots)$ term, we see that in order to have $e(p/q)<e(\alpha)$ we need something like $\log q>\log H(\alpha)$.

It is the appropriate generalization of this argument to polynomials in two or more variables which is needed in Siegel's and Roth's approach and this leads to results which are useful only in the case of approximations p/q in which q is fairly large compared with $H(\alpha)$. We have been unable so far to construct pairs $(\alpha, p/q)$ in which p/q is an excellent approximation to α and also q is large enough for this method to work.

I.4. If we look again at the preceding problem there is an obvious inequality which has to be satisfied if $\alpha\neq p/q$, namelv

$$e(p/q)+e(\alpha)r\leqslant d$$

because P vanishes at p/q to order $e(p/q)$ and also vanishes at α and at all conjugates of α to order $e(\alpha)$.

This purely algebraic approach is the one followed by Dyson in his work and, as far as we know, it has not been the object of further study since the appearance of Dyson's paper [D]. On the other hand, it has the distinct advantage of being entirely free from considerations of heights and hence it allows any approximation p_0/q_0 to α to be a candidate for a starting point to obtain effective results. This is exactly what we do in this paper. We shall obtain a very explicit formulation of Gelfond's result and as a

17-812904 Acta mathematica 148. Imprimé le 31 août 1982

special case we shall exhibit an infinite class of number fields of large degree ($\geqslant 40$) in which an effective form of Thue's theorem holds.

The question of effective measures of irrationality for an algebraic number is of considerable importance and it has attracted much attention; it has also proved to be a difficult one. The first non-trivial approximation theorems for a class of algebraic numbers were obtained by Baker [Ba1] using techniques related to the construction of explicit Padé approximants to algebraic functions of one variable. These techniques have been further developed to deal with simultaneous approximations [Ba2], so that it is possible to obtain uniform approximation results for certain fields of the type $Q(\sqrt[n]{a/b})$ with a/b sufficiently close to 1. A notable success of this method has been Baker's result

$$\left|\sqrt[3]{2}-\frac{p}{q}\right| > 10^{-6}q^{-2.955};$$

further improvements have been announced in the work of G. V. Choodnovsky.

An entirely different approach to the problem of effective bounds has been made possible by Baker's work on linear forms of logarithms; an exposition of the current state of the theory can be found in [Ba3]. It suffices here to mention that the work of Baker, together with some fundamental improvements by Feldman [F], has led to the first general non-trivial effective improvement in the exponent of approximation, for every algebraic number. On the other hand, the gain in the exponent appears to be extremely small and it depends badly on the height of the number of be approximated. However, it has been pointed out by Baker [Ba4] that for numbers of the type $\alpha^{1/m}$, m large, his technique yields an exponent of the kind $c(\alpha)\log m$, which for *fixed* α and large m is even better than the Thue-Siegel exponent $2\sqrt{m}$.

To sum up, our knowledge about effective approximations is of the following kind:

(A) good effective exponents (approaching Roth's exponent 2) for numbers of the type $\sqrt[n]{a/b}$, obtained through the use of Padé approximations to algebraic hypergeometric functions;

(B) a small effective improvement on the Liouville exponent, which however applies to every algebraic number, obtained through the use of the theory of linear forms in logarithms. For special numbers, the exponents obtained are also good.

To these results, we can now add:

(C) good effective exponents for all generators of certain number fields, the exponent being the same for all generators, obtained by refining the original Thue-Siegel method.

I.5. The refinements of the Thue-Siegel method which are relevant to us require a careful use of Dyson's idea for proving the non-vanishing of the auxiliary polynomial and also require that losses in estimating should be reduced to a minimum. As a consequence, besides using Dyson's approach, it is absolutely necessary that all estimates be carried out with the utmost precision. Also the p-adic generalizations of the Thue-Siegel theorem must be considered. For this reason, we shall use in a systematic way the absolute height (Mahler's measure) rather than more familiar notions of height, as well as Lang's local to global technique (see [L], Chapter VI).

Our notations and definitions are as follows.

Let K be a number field. We write $d=[K: Q]$ and for every place v of K we write $d_v=[K_v: Q_v]$. If the finite place v of K lies over the prime number p, we write $v|p$. We normalize the absolute value $|\ |_v$ so that

(i) if $v|p$ then

$$|p|_v = p^{-d_v/d},$$

(ii) if $v|\infty$ and v is real then

$$|x|_v = |x|^{1/d},$$

(iii) if $v|\infty$ and v is complex then

$$|x|_v = |x|^{2/d};$$

here $|\ |$ denotes the Euclidean absolute value in $\mathbf{R}$ or $\mathbf{C}$. In view of our normalizations, we have the product formula

$$\prod_v |x|_v = 1$$

if $x \in K^*$.

We define the absolute height of $x \in K$ by the formula

$$h(x) = \prod_v \max(1, |x|_v);$$

one of its main properties is that it does not change if we replace K by a finite extension (see Weil [W]); we also define

$$\log^+ a = \log \max(1, a) \quad \text{for} \quad a \geqslant 0,$$

so that

$$\log h(x) = \sum_v \log^+ |x|_v .$$

Let S be a set of places of K and let $x \in K^*$. From the product formula we have

$$\sum_S \log |x|_v = -\sum_{v \notin S} \log |x|_v \geqslant -\sum_{v \notin S} \log^+ |x|_v \geqslant -\log h(x).$$

Moreover, if $x \neq 0$ the product formula yields

$$\log h(1/x) = \log h(x),$$

which combined with the previous inequality gives

$$\sum_S \log |x|_v \leqslant \log h(1/x) = \log h(x).$$

From the last two inequalities we deduce the

Fundamental inequality. Let $x \in K^*$ and let S be any set of places of K. Then we have

$$-\sum_{v \notin S} \log^+ |x|_v \leqslant \sum_S \log |x|_v \leqslant \sum_{v \notin S} \log^+ |1/x|_v$$

and in particular

$$-\log h(x) \leqslant \sum_S \log |x|_v \leqslant \log h(x),$$

where $h(x)$ is the absolute height.

This notion of height is easily extended to vectors in the following way. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with $x_j \in K$ we define

$$h(\mathbf{x}) = \prod_v \max(1, |x_1|_v, \ldots, |x_n|_v).$$

We may further extend this notion to polynomials in any number of variables and to vectors of polynomials, by taking the height of the vector whose components are the

coefficients of the polynomial, and similarly we proceed with matrices; we also define $|\mathbf{x}|_v = \max |x_j|_v$.

The following properties of height are easily established. First of all, the inequality $\max(1, ab) \leqslant \max(1, a) \max(1, b)$ yields

$$h(xy) \leqslant h(x)\, h(y)$$

for $x, y \in K$; also

$$h(x^a) = h(x)^{|a|}$$

if x is algebraic and $a \in Q$. Next, we may note that if v is a finite place then

$$|x_1 + \ldots + x_n|_v \leqslant \max |x_j|_v$$

while if v is infinite then

$$|x_1 + \ldots + x_n|_v \leqslant n^{d_v/d} \max |x_j|_v;$$

sine $\sum_{v|\infty} d_v = d$, we obtain:

If $x_1, \ldots, x_n \in K$ then

$$h(x_1 \ldots x_n) \leqslant h(x_1) \ldots h(x_n),\ h(x_1 + \ldots + x_n) \leqslant n h(\mathbf{x}),$$

where $\mathbf{x}$ *is the vector* $\mathbf{x} = (x_1, \ldots, x_n)$.

Moreover of $\mathbf{x}$ and $\mathbf{y}$ are any vectors with components in K, then

$$h(\mathbf{x} \otimes \mathbf{y}) \leqslant h(\mathbf{x}) h(\mathbf{y}).$$

If α is an algebraic number of degree r and if $H(\alpha)$ is the maximum of the coefficients of an irreducible equation for α over Z, we have

$$C_1(r) H(\alpha)^{1/r} \leqslant h(\alpha) \leqslant C_2(r)\, H(\alpha)^{1/r}$$

where $C_1(r)$, $C_2(r)$ depend only on r; this may be useful in visualizing the size of $h(\alpha)$.

Finally, we shall abbreviate

$$\lambda(\alpha) = \log h(\alpha)$$

and call $\lambda(\alpha)$ the logarithmic height.

I.6. The content of this paper is as follows.

In section II, we prove Dyson's lemma, essentially using his arguments.

In section III, we construct the auxiliary polynomial $P(x_1, x_2)$ and its vanishing at (α_1, α_2) and (β_1, β_2) is controlled by means of Lemma 5 and Lemma 6; Lemma 5 represents the application of Dyson's lemma.

In section IV, we prove a general effective result relating an approximation β_1 to α_1 with another approximation β_2 to α_2. This result is stated in Theorem 2. Theorems 3 and 4 are special cases of Theorem 2, formulated in a more familiar way.

In section V, we show with examples that our results are sufficiently explicit and precise to yield some cases of effectiveness for the Thue-Siegel theorem; in particular our Example 2 is completely explicit in all numerical constants.([1])

Finally, I would like to thank here the Mittag-Leffler Institut for providing computer time and assisting in the numerical calculations of section V.

Note. I wish to take the occasion of the publication of this paper to clarify the relative status of two preceding papers, namely "Algebraic values of meromorphic maps", Inv. Math., 10, 267–287 (1970) by Enrico Bombieri and "Analytic subgroups of Group varieties", Inv. Math., 11, 1–14 (1970) by Enrico Bombieri and Serge Lang.

In fact, although submitted and received at the same time (June 29, 1970) and although my paper appeared before the joint paper, the joint paper actually precedes the other. In particular the Schwarz lemma of the joint paper, which appears also as Proposition 4 in the first paper, originates entirely in the joint paper.

Finally I must point out that the references to the first paper which appear in our joint paper were added a posteriori at my request, for the sake of completeness. The fact that my paper does not contain a similar cross reference to the joint paper is simply due to my oversight. I wish to apologize to Professor Lang if this fact has caused misunderstandings in attributing to me alone ideas and results obtained in collaboration with him.

II. Dyson's lemma

II.1. Let K be an algebraically closed field of characteristic 0 and let $\xi_\mu = (\xi_{\mu 1}, \ldots, \xi_{\mu n})$, $\mu = 1, \ldots, m$ be m points in K^n. We shall assume that *for* $i = 1, 2, \ldots, n$

([1]) We have made no effort here in trying to single out various classes of number fields to which our method applies. Indeed, the reader will perceive that strong effective results can be obtained in many cases of ample generality. Any improvement of Lemma 6 would lead to better results and it is easily seen that if we could replace the constant $r\varphi_2(t)/(1-r\varphi_2(t))$ by 1 as a multiplier of $d_1\lambda(\alpha_1)$ (and we allow worse multipliers for $d_2\lambda(\alpha_2)$) then the optimal exponent 2 would follow effectively for every algebraic number. We hope to return to these questions in future papers.

the m numbers $\xi_{1i}, \xi_{2i}, \ldots, \xi_{mi}$ *are distinct;* in this case the set of points ξ_μ will be called admissible. Let $\vartheta_i > 0$, $i=1, \ldots, n$ be real numbers.

Let $d_1, \ldots, d_n$ be positive numbers (not necessarily integers) and let t_μ, $\mu=1, \ldots, m$ be real numbers. We define

$$\mathcal{P}(\mathbf{d}, \vartheta; t_1, \ldots, t_m | \xi_1, \ldots, \xi_m),$$

and abbreviate $\mathcal{P}(\mathbf{d}; t_\mu)$, to be the vector space over K consisting of all polynomials $P=P(x_1, \ldots, x_n)$ in n variables with coefficients in K and satisfying

$$\deg_{x_i} P \leqslant d_i \tag{A}$$

for $i=1, \ldots, n$;

$$\Delta^I P(\xi_\mu) = 0 \tag{B}$$

for

$$\Delta^I = \frac{\partial^{i_1 + \ldots + i_n}}{\partial x_1^{i_1} \ldots \partial x_n^{i_n}}$$

and all indices $I = (i_1, \ldots, i_n)$ *with*

$$\vartheta_1 \frac{i_1}{d_1} + \ldots + \vartheta_n \frac{i_n}{d_n} < t_\mu,$$

for $\mu=1, \ldots, m$. It is convenient to allow t_μ to be negative or 0, in which case no condition on P at the point ξ_μ is imposed.

Let us define

$$\varphi_n(t) = \int_0^1 \ldots \int_0^1 dx_1 \ldots dx_n,$$
$$\vartheta_1 x_1 + \ldots + \vartheta_n x_n \leqslant t$$

so that the number of solutions of

$$\vartheta_1 \frac{i_1}{d_1} + \vartheta_2 \frac{i_2}{d_2} + \ldots + \vartheta_n \frac{i_n}{d_n} < t, \quad i_\nu \leqslant d_\nu$$

for large d_ν is asymptotic to $\varphi_n(t) d_1 d_2 \ldots d_n$. Since the number of indices $(i_1, i_2, \ldots, i_n)$ at out disposal is asymptotic to $d_1 d_2 \ldots d_n$, we see that $\varphi_n(t)$ measures the number of

indices satisfying the given conditions, as the degrees $d_1, \dots, d_n$ go to infinity. We have $\varphi_n(t)=0$ if $t\leqslant 0$, $\varphi_n(t)=1$ if $t \geqslant \Sigma\vartheta_i$.

If we fix $\xi_1, \dots, \xi_m$ and $t_1, \dots, t_m$ with

$$\sum_\mu \varphi_n(t_\mu) < 1,$$

the vector space $\mathscr{P}(\mathbf{d}; t_\mu)$ is not (0) as soon as the degrees $d_1, \dots, d_n$ are sufficiently large. Indeed each equation in (B) is a homogeneous linear equation in the coefficients of P and we have $(d_1+1)\dots(d_n+1) \sim d_1 \dots d_n$ coefficients. Now $\Sigma_\mu \varphi_n(t_\mu)<1$ implies that for large $d_1, \dots d_n$ we have more unknowns than equations in our linear system, which makes the result obvious. Conversely, the question arises whether the condition $\Sigma_\mu \varphi_n(t_\mu)\leqslant 1$ is also necessary for having $\mathscr{P}(\mathbf{d}; t_\mu)\neq(0)$. This is so if $\xi_1, \dots, \xi_m$ are generic in an appropriate sense (which we do not need to make precise here) and this gives some support to the view that it may always be so. It is the content of Dyson's lemma that this is essentially the case if $\xi_1, \dots, \xi_m$ is admissible, if $n=2$ and if the ratio d_2/d_1 is small. We have

THEOREM 1. *Let $\xi_1, \dots, \xi_m$ be admissible and let $n=2$. Then if $\mathscr{P}(\mathbf{d}; t_\mu)\neq(0)$ we have*

$$\sum_\mu \varphi_2(t_\mu) \leqslant 1+\max\left(\frac{m-2}{2}, 0\right)\frac{d_2}{d_1}.$$

In order to appreciate the meaning of the admissibility condition, let us consider the case $n=1$. Then (B) implies that $P(x)$ has a zero at ξ_μ with multiplicity not less than $\max(t_\mu, 0)d=\varphi_1(t_\mu)d$. If the ξ_μ are admissible, that is distinct, then

$$d\sum \varphi_1(t_\mu) \leqslant \sum (\text{multiplicity at } \xi_\mu) \leqslant \deg P \leqslant d$$

and

$$\sum \varphi_1(t_\mu) \leqslant 1.$$

This simple argument makes clear why admissibility is needed, for otherwise the example of $P(x_i)$ taken as polynomial in several variables $x_1, \dots, x_n$ shows that no result like Theorem 1 can possibly hold.

We shall deduce Theorem 1 from the apparently weaker

DYSON'S LEMMA. *Let* $\xi_1, \ldots, \xi_m$ *be admissible and let* $n=2$. *Then if* $\mathscr{P}(\mathbf{d};\ t_\mu)\neq(0)$ *and* $0<t_\mu\leqslant\min(\vartheta_1, \vartheta_2)$ *we have*

$$\sum \frac{1}{2\vartheta_1\vartheta_2} t_\mu^2 \leqslant 1+\max\left(\frac{m-2}{2},\ 0\right)\frac{d_2}{d_1}.$$

II.2. In this section we prove some simple facts on polynomials $P\in\mathscr{P}(\mathbf{d};\ t_\mu)$; our results are stated for polynomials in several variables.

LEMMA 1. *We have*:

(a) *if* $t_\mu^*\leqslant t_\mu$ *then*

$$\mathscr{P}(\mathbf{d};\ t_\mu)\subset\mathscr{P}(\mathbf{d};\ t_\mu^*);$$

(b) *if* $\alpha\geqslant 1$ *then*

$$\mathscr{P}(\mathbf{d};\ t_\mu)\subset\mathscr{P}(\alpha\mathbf{d};\ \alpha^{-1}t_\mu);$$

(c) *if* $P_\varrho\in\mathscr{P}(\mathbf{d}_\varrho;\ t_{\varrho\mu})$, $\varrho=1, \ldots, r$ *then*

$$P_1\ldots P_r\in\mathscr{P}\left(\sum_\varrho \mathbf{d}_\varrho;\ \sum_\varrho \min_i\left(\frac{d_{\varrho i}}{d_{1i}+\ldots+d_{ri}}\right)t_{\varrho\mu}\right);$$

(d) *if* $P\in\mathscr{P}(\mathbf{d};\ t_\mu)$ *then*

$$P^N\in\mathscr{P}(N\mathbf{d};\ t_\mu);$$

(e)

$$\Delta^I\mathscr{P}(\mathbf{d};\ t_\mu)\in\mathscr{P}\left(\alpha\mathbf{d};\ \alpha^{-1}\left(t_\mu-\sum\frac{i_\nu}{d_\nu}\right)\right)$$

for every $\alpha\geqslant 1-\min i_\nu/d_\nu$;

(f) $\mathscr{P}(\mathbf{d};\ t_\mu)=(0)$ *if* $t_\mu>\Sigma\vartheta_i$ *for some* μ.

Proof. (a), (b), (e) follow readily from the definition of $\mathscr{P}$. (d) is a special case of (c). For (c), we note that

$$\Delta^I(P_1\ldots P_r)=\sum(\Delta^{I_1}P)\ldots(\Delta^{I_r}P_r)$$

where the sum is over all $I_1, \dots, I_r$ with $I_1 + \dots + I_r = I$. Now

$$\sum_{\varrho} \min_{i} \left(\frac{d_{\varrho i}}{d_{1i} + \dots + d_{ri}} \right) \left(\vartheta_1 \frac{i_{\varrho 1}}{d_{\varrho 1}} + \dots + \vartheta_n \frac{i_{\varrho n}}{d_{\varrho n}} \right)$$

$$\leqslant \sum_{\nu} \vartheta_\nu \frac{\sum_{\varrho} i_{\varrho \nu}}{\left(\sum_{\varrho} d_{\varrho \nu} \right)} = \sum_{\nu} \vartheta_\nu \frac{i_\nu}{\left(\sum_{\varrho} d_{\varrho \nu} \right)},$$

hence if

$$\sum_{\nu} \vartheta_\nu \frac{i_\nu}{\left(\sum_{\varrho} d_{\varrho \nu} \right)} < \sum_{\varrho} \min_{i} \left(\frac{d_{\varrho i}}{\sum_{\varrho} d_{\varrho i}} t_{\varrho \mu} \right)$$

then

$$\vartheta_1 \frac{i_{\varrho 1}}{d_{\varrho 1}} + \dots + \vartheta_n \frac{i_{\varrho n}}{d_{\varrho n}} < t_{\varrho \mu}$$

for at least one ϱ and the corresponding derivative $\Delta^{I_\varrho} P_\varrho$ vanishes at ξ_μ. Since this holds for every decomposition, we see that $\Delta^I(P_1 \dots P_r)(\xi) = 0$, which proves (c). Finally (f) follows from Taylor's theorem.

Let $P(x_1, \dots, x_n)$ be a polynomial and let us consider decompositions

$$P = \sum_{j=0}^{s} f_j(\mathbf{x}') g_j(x_n)$$

where $f_j(\mathbf{x}') = f_j(x_1, \dots, x_{n-1})$ and $g_j(x_n)$ are polynomials. We certainly have decompositions in which $f_0, f_1, \dots, f_s$ are linearly independent over K and also $g_0, g_1, \dots, g_s$ are linearly independent over K; indeed, any decomposition in which s is minimal has the required property. We define

$$s_n(P) = \max s$$

where the max is over all decompositions with $f_0, \dots, f_s$ linearly independent over K and $g_0, \dots, g_s$ linearly independent over K. It is clear that

$$0 \leqslant s_n(P) \leqslant d_n$$

because $g_0, \dots, g_s$ are polynomials in one variable of degree $\leqslant d_n$ and thus not more than $d_n + 1$ of them can be linearly independent.

LEMMA 2. *For every P we can find a polynomial $L=a+b_1x_1+\ldots+b_nx_n$ of degree* 1 *such that*

$$s_n(LP)\geqslant s_n(P)+1.$$

Proof. We use matrix notation and write $\mathbf{f}$, $\mathbf{g}$ for the corresponding column vectors; we have

$$P={}^t\mathbf{g}\cdot\mathbf{f}$$

where ${}^t(\ )$ denotes the transpose. It is clear that at least one of $x_ng_0,\ldots,x_ng_s$ is linearly independent from $g_0,\ldots,g_s$ (just check degrees); let $\mathbf{g}_\mathrm{I}$ be a maximal subset of $g_0,\ldots,g_s$ such that $g_0,\ldots,g_s,x_ng_i$ with $i\in I$ are linearly independent and let $\mathbf{g}_\mathrm{II}$ be the complementary set. Then we can write

$$x_n{}^t\mathbf{g}_\mathrm{II}={}^t\mathbf{g}\cdot A+x_n{}^t\mathbf{g}_\mathrm{I}\cdot B$$

for suitable constant matrices A, B. Let us write $L(\mathbf{x})=L_0(\mathbf{x}')+b_nx_n$, where $\mathbf{x}'=(x_1,\ldots,x_{n-1})$. We have, with an obvious notation for $A_\mathrm{I},A_\mathrm{II}$:

$$\begin{aligned}(L_0(\mathbf{x}')+b_nx_n)P&=(L_0(\mathbf{x}')+b_nx_n)\,{}^t\mathbf{g}\cdot\mathbf{f}\\&={}^t\mathbf{g}_\mathrm{I}(L_0(\mathbf{x}')\mathbf{f}_\mathrm{I}+b_nA_\mathrm{I}\mathbf{f}_\mathrm{II})+{}^t\mathbf{g}_\mathrm{II}(L_0(\mathbf{x}')\mathbf{f}_\mathrm{II}+b_nA_\mathrm{II}\mathbf{f}_\mathrm{II})+x_n{}^t\mathbf{g}_\mathrm{I}\,b_n\cdot\end{aligned}$$

$(\mathbf{f}_\mathrm{I}+B\mathbf{f}_\mathrm{II})$,

and ${}^t\mathbf{g}_\mathrm{I}$, ${}^t\mathbf{g}_\mathrm{II}$, $x_n{}^t\mathbf{g}_\mathrm{I}$ are linearly independent by construction. If we show that the column rank of

$$\begin{pmatrix}L_0(\mathbf{x}')\mathbf{f}_\mathrm{I}+b_nA_\mathrm{I}\mathbf{f}_\mathrm{II}\\L_0(\mathbf{x}')\mathbf{f}_\mathrm{II}+b_nA_\mathrm{II}\mathbf{f}_\mathrm{II}\\\mathbf{f}_\mathrm{I}+B\mathbf{f}_\mathrm{II}\end{pmatrix}$$

is at least $s+1$ for a *generic* choice of $L_0(\mathbf{x}')$ and b_n, the result follows. Since the rank does not increase if we specialize $L(\mathbf{x})$ we may set $b_n=0$ and prove that for generic $L_0(\mathbf{x}')$ the rank of

$$\begin{pmatrix}L_0(\mathbf{x}')\mathbf{f}_\mathrm{I}\\L_0(\mathbf{x}')\mathbf{f}_\mathrm{II}\\\mathbf{f}_\mathrm{I}+B\mathbf{f}_\mathrm{II}\end{pmatrix}$$

is at least $s+1$. If not, then since the rows $L_0(\mathbf{x}')\mathbf{f}_{\mathrm{I}}$ and $L_0(\mathbf{x}')\mathbf{f}_{\mathrm{II}}$ are linearly independent we must have

$$\mathbf{f}_{\mathrm{I}}+B\mathbf{f}_{\mathrm{II}} = N_L \cdot L_0(\mathbf{x}') \begin{pmatrix} \mathbf{f}_{\mathrm{I}} \\ \mathbf{f}_{\mathrm{II}} \end{pmatrix}$$

for a suitable constant matrix N_L. This means that $\mathbf{f}_{\mathrm{I}}+B\mathbf{f}_{\mathrm{II}}$ is divisible by $L_0(\mathbf{x}')$ and since $L_0(\mathbf{x}')$ is arbitrary we must have

$$\mathbf{f}_{\mathrm{I}}+B\mathbf{f}_{\mathrm{II}} = 0.$$

This contradicts the fact that $f_0, \dots, f_s$ are linearly independent and proves our result.

COROLLARY. *Let* $\mathscr{P}(\mathbf{d}; t_\mu) \neq (0)$ *and let* $\alpha > 1$. *Then there is* $P \in \mathscr{P}(\alpha\mathbf{d}; \alpha^{-1} t_\mu)$ *with* $P \neq 0$ *and*

$$s_n(P) \geqslant [(\alpha-1) \min_\nu d_\nu].$$

II.3. In what follows, $P \in \mathscr{P}(\mathbf{d}; t_\mu)$ and ξ_μ is admissible. We have a decomposition

$$P = \mathrm{f}_0(\mathbf{x}')\, g_0(x_n) + \dots + f_s(\mathbf{x}')\, g_s(x_n)$$

where f_0, $\mathrm{f}_1, \dots, f_s$ are linearly independent and similary g_0, $g_1, \dots, g_s$ are linearly independent. In view of this property of linear independence we know that some generalized Wronskian of $f_0, \dots, f_s$, and also the Wronskian of $g_0, \dots, g_s$, is not identically 0. To be more explicit, there are differential operators δ_0, $\delta_1, \dots, \delta_s$ such that δ_i has degree $\leqslant i$ and such that

$$F(\mathbf{x}') = \det (\delta_i f_j)_{i,j=0,\dots,s}$$

is not identically 0. Similarly,

$$G(x_n) = \det \left(\left(\frac{\partial}{\partial x_n} \right)^k g_j \right)_{j,k=0,\dots,s}$$

is not identically 0.

LEMMA 3. *We have*

$$F(\mathbf{x}')\, G(x_n) = \det \left(\delta_i \left(\frac{\partial}{\partial x_n} \right)^k P \right)_{j,k=0,\dots,s}.$$

Moreover,

$$\deg_{x_n} G \leqslant (s+1)(d_n-s).$$

Proof. Since the variables $\mathbf{x}'$ and x_n in $f_j(\mathbf{x}')$ and $g_j(x_n)$ are separated, we have

$$\delta_i\left(\frac{\partial}{\partial x_n}\right)^k P = \sum_{j=0}^{s} (\delta_i f_j)\left(\frac{\partial}{\partial x_n}\right)^k g_j;$$

now the identity of Lemma 3 follows by multiplication of the matrices associated to $F(\mathbf{x}')$ and $G(x_n)$. In order to prove the last statement in Lemma 3, we note that there is a basis g_j^* of the vector space over K spanned by $g_0, \ldots, g_s$, such that

$$d_n \geqslant \deg g_0^* > \deg g_1^* > \ldots > \deg g_s^* \geqslant 0.$$

The Wronskian of the g_j^* is proportional to $G(x_n)$ and hence we may assume that the degrees $\deg g_j$ form a strictly decreasing sequence of integers and in particular

$$\deg g_j \leqslant d_n - j.$$

A typical term in the expansion of the determinant for $G(x_n)$ is

$$\prod_{j=0}^{s}\left(\frac{\partial}{\partial x_n}\right)^{k_j} g_j$$

where the k_j form a permutation of $0, \ldots, s$. The degree of this product is

$$\sum_{j=0}^{s} (\deg g_j - k_j) \leqslant \sum_{j=0}^{s} (d_n - j - k_j) = (s+1)(d_n - s),$$

because the k_j are a permutation of $0, \ldots, s$. This proves Lemma 3.

Remark. In the special case in which $n=2$ we must have $\delta_i = (\partial/\partial x_1)^i$ and hence we can repeat the previous argument and find

$$\deg_{x_1} F(x_1) \leqslant (s+1)(d_1 - s).$$

II.4. In this section we prove Dyson's lemma. Let $P \neq 0$, $P \in \mathscr{P}(\mathbf{d}; t_\mu)$. If we replace P by P^N with $N \to \infty$ and use Lemma 2 we see that we may suppose in proving Dyson's lemma that s is arbitrarily large.

Let $P = \sum_{j=0}^{s} f_j(x_1)\, g_j(x_2)$ be our decomposition of P, and let us consider μ as being fixed. Let A_μ be a constant $(s+1)\times(s+1)$ invertible matrix. We have

$$P = {}^{t}\mathbf{f}\cdot\mathbf{g} = {}^{t}\mathbf{f}A_\mu^{-1}\cdot A_\mu\mathbf{g} = {}^{t}({}^{t}A_\mu^{-1}\mathbf{f})\,(A_\mu\mathbf{g})$$

and we choose A_μ so that if we write $\mathbf{g}^{(\mu)}=A_\mu\mathbf{g}$ then

$$0\leqslant \operatorname{ord}_{\xi_{\mu 2}} g_0^{(\mu)} < \operatorname{ord}_{\xi_{\mu 2}} g_1^{(\mu)} < \ldots < \operatorname{ord}_{\xi_{\mu 2}} g_s^{(\mu)} \leqslant d_2.$$

We may do this in several ways; we select one for each μ and define

$$u_{\mu j} = \operatorname{ord}_{\xi_{\mu 2}} g_j^{(\mu)}.$$

Let

$$G(x_2) = \det\left(\frac{\partial^k}{\partial x_2^k} g_j\right)_{j,k=0,\ldots,s};$$

by Lemma 3, we have

$$\deg T(x_2) \leqslant (s+1)\,(d_2-s).$$

In order to compute the order of zero of $G(x_2)$ at $\xi_{\mu 2}$, we note that

$$\det\left(\frac{\partial^k}{\partial x_2^k} g_j^{(\mu)}\right)_{j,k=0,\ldots,s} = \det(A_\mu)\, G(x_2)$$

and hence it has the same order of zero at $\xi_{\mu 2}$ as $G(x_2)$ because A_μ is invertible. Now a typical term in the expansion of the determinant is

$$\pm\prod\left(\frac{\partial}{\partial x_2}\right)^{k_j} g_j^{(\mu)}$$

where the k_j's are a permutation of $0, 1, \ldots, s$, and hence it vanishes at $\xi_{\mu 2}$ at least to the order

$$\sum_{j=0}^{s}(u_{\mu j}-k_j) = \sum_{j=0}^{s}(u_{\mu j}-j).$$

Thus we have shown that

$$\operatorname{ord}_{\xi_{\mu 2}} G(x_2) \geqslant \sum_{j=0}^{s}(u_{\mu j}-j).$$

Now the points $\xi_{\mu 2}$ are distinct, hence

$$\sum_{\mu} \operatorname{ord}_{\xi_{\mu 2}} G(x_2) \leqslant \deg G(x_2)$$

which yields

$$\sum_{j=0}^{s} \sum_{\mu} (u_{\mu j} - j) \leqslant (s+1)(d_2 - s)$$

It is here that we have Dyson's important remark that, since the $u_{\mu j}$ are integers strictly increasing in j, the quantity $\Sigma_{\mu}\,(u_{\mu j} - j)$ is increasing as a function of j. It follows that

$$\sum_{j=0}^{b} \sum_{\mu} (u_{\mu j} - j) \leqslant (b+1)(d_2 - s)$$

for $b = 0, 1, \ldots, s$.

Our next task consists in obtaining a lower bound for $\operatorname{ord}_{\xi_{\mu 1}} F(x_1)$. This is done as follows. Let us consider

$$\det\left(\left(\frac{\partial}{\partial x_1}\right)^{i}\left(\frac{\partial}{\partial x_2}\right)^{u_{\mu k}} P(x_1, \xi_{\mu 2})\right)_{i,k=0,\ldots,s}.$$

If we write $\mathbf{f}^{(\mu)} = {}^{t}A_{\mu}^{-1}\,\mathbf{f}$ and recall that $P = {}^{t}\mathbf{f}^{(\mu)} \cdot \mathbf{g}^{(\mu)}$ then we see that this determinant is

$$\det({}^{t}A_{\mu}^{-1})\, F(x_1) \det\left(\left(\frac{\partial}{\partial x_2}\right)^{u_{\mu k}} g_j^{(\mu)}(\xi_{\mu 2})\right)_{j,k=0,\ldots,s}.$$

Now

$$\left(\frac{\partial}{\partial x_2}\right)^{u_{\mu k}} g_j^{(\mu)}(\xi_{\mu 2}) \neq 0$$

if $k = j$, while

$$\left(\frac{\partial}{\partial x_2}\right)^{u_{\mu k}} g_j^{(\mu)}(\xi_{\mu 2}) = 0$$

if $k < j$, and it follows that

$$\det\left(\left(\frac{\partial}{\partial x_2}\right)^{u_{\mu k}} g_j^{(\mu)}(\xi_{\mu 2})\right)_{j,k=0,\ldots,s} \neq 0$$

since the matrix is non-singular triangular. This shows that $F(x_1)$ vanishes at $\xi_{\mu 1}$ exactly as

$$\det\left(\left(\frac{\partial}{\partial x_1}\right)^i\left(\frac{\partial}{\partial x_2}\right)^{u_{\mu k}}P(x_1,\xi_{\mu 2})\right)_{i,k=0,\dots,s}$$

By our hypotheses, $(\partial/\partial x_2)^{u_{\mu k}}P(x_1,\xi_{\mu 2})$ vanishes at $x_1=\xi_{\mu 1}$ at least to the order

$$\frac{1}{\vartheta_1}d_1\max\left(t_\mu-\vartheta_2\frac{u_{\mu k}}{d_2},\,0\right),$$

and hence a typical term in the expansion of the determinant vanishes at $x_1=\xi_{\mu 1}$ at least to the order

$$\frac{1}{\vartheta_1}d_1\sum_{k=0}^{s}\max\left(t_\mu-\vartheta_1\frac{i_k}{d_1}-\vartheta_2\frac{u_{\mu k}}{d_2},\,0\right)$$

where the i_k's are a permutation of $0,1,\dots,s$. *A fortiori*, the determinant itself and $F(x_1)$ vanish at $\xi_{\mu 1}$ at least to the order

$$\frac{1}{\vartheta_1}d_1\sum_{k=0}^{s}\max\left(t_\mu-\vartheta_2\frac{k}{d_2}-\vartheta_2\frac{(u_{\mu k}-k)}{d_2},\,0\right)-\frac{s(s+1)}{2},$$

and since the $\xi_{\mu 1}$'s are distinct, we obtain

$$\frac{1}{\vartheta_1}d_1\sum_{\mu}\sum_{k=0}^{s}\max\left(t_\mu-\vartheta_2\frac{k}{d_2}-\vartheta_2\frac{(u_{\mu k}-k)}{d_2},\,0\right)-m\frac{s(s+1)}{2}\leqslant(s+1)(d_1-s)$$

by our bound on $\deg F$ obtained in the remark to Lemma 3.

Let

$$t=\max t_\mu$$

and let

$$b=\min\left(s,\frac{1}{\vartheta_2}d_2t\right).$$

Then we can replace $\Sigma_{k=0}^{s}$ by $\Sigma_{k=0}^{b}$ in the last inequality and deduce the basic inequality

$$\begin{aligned}\frac{1}{\vartheta_1}d_1\sum_{\mu}\sum_{k=0}^{b}\max\left(t_\mu-\vartheta_2\frac{k}{d_2},\,0\right)&\leqslant(s+1)(d_1-s)+\frac{1}{\vartheta_1}d_1\vartheta_2\frac{1}{d_2}\sum_{k=0}^{b}(u_{\mu k}-k)\\&+\frac{m}{2}s(s+1)\leqslant(s+1)d_1+\frac{\vartheta_2}{\vartheta_1}\frac{d_1}{d_2}(b+1)(d_2-s)+\left(\frac{m}{2}-1\right)s(s+1),\end{aligned}$$

the last step coming from the inequality obtained by Dyson's remark.

Let us write

$$s = \sigma d_2, \quad b = \beta d_2,$$

so that $0<\sigma\leqslant 1$ and

$$\beta = \min\left(\sigma, \frac{1}{\vartheta_2} t\right),$$

where

$$t = \max t_\mu.$$

We divide both sides of our basic inequality by $d_1 d_2$ and we approximate Σ_k by an integral. This yields

$$\frac{1}{\vartheta_1 \vartheta_2} \sum_\mu \int_0^{\beta\vartheta_2} \max(t_\mu - x, 0)\, dx \leqslant \sigma + \frac{\vartheta_2}{\vartheta_1}\beta(1-\sigma) + \left(\frac{m}{2}-1\right)\sigma^2 \frac{d_2}{d_1} + O\left(\frac{1}{d_2}\right).$$

Now we obtain a lower bound for the integral as follows.

Case I. $\beta\vartheta_2 \geqslant t_\mu$.

In this case,

$$\int_0^{\beta\vartheta_2} \max(t_\mu - x, 0)\, dx = \frac{1}{2} t_\mu^2$$

and also

$$\sigma + \frac{\vartheta_2}{\vartheta_1}\beta(1-\sigma) \leqslant 1$$

because

$$\vartheta_2 \beta \leqslant t \leqslant \min(\vartheta_1, \vartheta_2) \leqslant \vartheta_1$$

by our hypotheses on t_μ. Thus

$$\int_0^{\beta\vartheta_2} \max(t_\mu - x, 0)\, dx \geqslant \frac{1}{2} t_\mu^2 \left\{\sigma + \frac{\vartheta_2}{\vartheta_1}\beta(1-\sigma)\right\}$$

in this case.

Case II. $\beta\vartheta_2 \leqslant t_\mu$.

In this case, we have $\beta \leqslant t/\vartheta_2$ hence $\beta=\sigma\leqslant t/\vartheta_2$. We have

$$\frac{\int_0^{\beta\vartheta_2} \max(t_\mu - x, 0)\,dx}{\sigma+\frac{\vartheta_2}{\vartheta_1}\beta(1-\sigma)} = \frac{t_\mu\vartheta_2\sigma - \frac{1}{2}(\vartheta_2\sigma)^2}{\sigma+\frac{\vartheta_2}{\vartheta_1}\sigma(1-\sigma)} = \frac{t_\mu\vartheta_2 - \frac{1}{2}\vartheta_2^2\sigma}{1+\frac{\vartheta_2}{\vartheta_1}-\frac{\vartheta_2}{\vartheta_1}\sigma}.$$

The last expression is decreasing in σ because

$$t_\mu \leqslant \min(\vartheta_1, \vartheta_2) \leqslant \frac{1}{2}(\vartheta_1+\vartheta_2)$$

and hence a lower bound is obtained by choosing σ as large as possible, which yields

$$\beta = \sigma = \frac{1}{\vartheta_2} t_\mu.$$

This is in the situation of Case I, and we get the lower bound $\frac{1}{2}t_\mu^2$ once again.

We have shown

$$\sum_\mu \frac{1}{2\vartheta_1\vartheta_2} t_\mu^2 \leqslant 1+\max\left(\frac{m}{2}-1, 0\right)\frac{d_2}{d_1}+O\left(\frac{1}{\sigma d_2}\right);$$

and since we may suppose that $\sigma d_2 = s$ is arbitrarily large we obtain Dyson's lemma.

II.5. We show here how to deduce Theorem 1 from Dyson's lemma. If we replace $P(x_1, x_2)$ by $P^N(x_1, x_2)$ and let $N\to\infty$ we see that the statement of Theorem 1 depends only on the values of $\vartheta_1, \vartheta_2, t_\mu$ and the ratio d_2/d_1. Without loss of generality, we may assume $\vartheta_1=1$ and write ϑ for ϑ_2; we also define δ by

$$\delta = d_2/d_1.$$

The polynomial P gives us a set of parameters $(t_\mu; \vartheta, \delta)$; conversely, we say that $(t_\mu; \vartheta, \delta)$ occurs if there is a sequence of polynomials P_i with parameters $(t_\mu^{(i)}; \vartheta^{(i)}, \delta^{(i)}) \to (t_\mu; \vartheta, \delta)$.

Let us consider the case $\vartheta \leqslant 1$. We divide the set of μ's into subsets M_1, M_2, M_3 as follows:

$$M_1 = \{\mu \mid t_\mu \leqslant \vartheta\}, \quad M_2 = \{\mu \mid \vartheta < t_\mu \leqslant 1\}, \quad M_3 = \{\mu \mid 1 < t_\mu\}.$$

LEMMA 4. *$|M_3|=0$ or 1. If $|M_3|=1$ then M_2 is empty. If $\vartheta<1$ then $M_3 \cup \{\mu | t_\mu=1\}$ has at most one element.*

Proof. We assume that $(t_1, t_2; \vartheta, \delta)$ occurs with $t_1>1$, $t_2=1$ and we will reach a contradiction.

Since we have $\vartheta \leqslant 1$, we may increase ϑ to 1 and $(t_1, 1;, 1, \delta)$ will occur *a fortiori*. Thus we may suppose $\vartheta=1$. Let P be a polynomial with parameters $(t_1-\varepsilon, 1-\varepsilon; 1, \delta)$, where $\varepsilon>0$ is arbitrarily small. Since $t_1>1$, we see that P factorizes as

$$P(x_1, x_2) = (x_1-\xi_{11})^{a_1} (x_2-\xi_{12})^{a_2} Q(x_1, x_2)$$

where $a_i=[(t_1-1-\varepsilon)d_i]$.

Now $\xi_{11} \neq \xi_{21}$ and $\xi_{12} \neq \xi_{22}$ because ξ_1 and ξ_2 are admissible and thus Q vanishes exactly as P at ξ_2, while we must lose a_1 derivatives with respect to x_1 and a_2 derivatives with respect to x_2 in the vanishing at the point ξ_1. It is an easy matter to obtain parameters for Q and if we consider $d_1, d_2 \to \infty$ and then let $\varepsilon \to 0$, as we may, we deduce that the new set of parameters

$$\left(1, \frac{1}{2-t_1}; 1, \delta\right)$$

also occurs. We note that $1/(2-t_1)>1$ and start our transformation once again but considering of course the vanishing at ξ_2. Then we see that

$$\left(\frac{2-t_1}{3-2t_1}, 1; 1, \delta\right)$$

also occurs. By iteration, we arrive at new parameters $t_\mu^{(k)}, k=1, 2, \ldots$ for which

$$\max t_\mu^{(k)} \geqslant \frac{k-(k-1)t_1}{(k+1)-kt_1}$$

and thus $\max t_\mu^{(k)}>2$ if k is sufficiently large. This contradicts the last clause of Lemma 1.

Now suppose that $(t_1, t_2, \ldots; \vartheta, \delta)$ occurs and either $t_1>1$ or $t_1=1$ and $\vartheta<1$. Then the same argument as before shows that

$$\left(1, \frac{t_2}{1+\vartheta-t_1}, \frac{t_3}{1+\vartheta-t_1}, \ldots; 1, \vartheta^{-1}\delta\right)$$

also occurs. This proves that we must have $t_\mu/(1+\vartheta-t_1)\leqslant 1$ for every $\mu\geqslant 2$, hence $t_\mu<\vartheta$ because $t_1>1$, and the proof of Lemma 4 is complete.

If we look at the last set of parameters we see that we can apply Dyson's lemma to it. We obtain

$$\frac{1}{2}+\sum_{\mu\geqslant 2}\frac{1}{2}\left(\frac{t_\mu}{1+\vartheta-t_1}\right)^2\leqslant 1+\max\left(\frac{m-2}{2},0\right)\vartheta^{-1}\delta$$

which is easily transformed into

$$1-\frac{1}{2\vartheta}(1+\vartheta-t_1)^2+\sum_{\mu\geqslant 2}\frac{1}{2\vartheta}t_\mu^2\leqslant 1+\max\left(\frac{m-2}{2},0\right)\left(\frac{1+\vartheta-t_1}{\vartheta}\right)^2\delta$$

$$<1+\max\left(\frac{m-2}{2},0\right)\delta;$$

since $\varphi_2(t)=1-\frac{1}{2}(1+\vartheta-t)^2/\vartheta$ if $t>1$, we have obtained Theorem 1 if some t_μ is $t_\mu>1$.

Now suppose that $(t_1,t_2,\ldots;\vartheta,\delta)$ occurs and $M_3=\varnothing$ but $|M_2|>1$. If P gives parameters

$$(t_1-\varepsilon,t_2-\varepsilon,\ldots;\vartheta,\delta)$$

and $\varepsilon>0$ is sufficiently small then again P factorizes as

$$P(x_1,x_2)=\prod_{\mu\in M_2}(x_1-\xi_{\mu 1})^{[(t_\mu-\vartheta-\varepsilon)d_1]}Q(x_1,x_2)$$

and now, by letting $d_1,d_2\to\infty$ and $\varepsilon\to 0$, we see that Q gives rise to parameters

$$(t_1^*,t_2^*,\ldots;\vartheta^*,\delta^*)$$

where

$$t_\mu^*=\frac{\vartheta}{1-S}=\vartheta^*\quad\text{if }\mu\in M_2,$$

$$t_\mu^*\leqslant\frac{t_\mu}{1-S}\leqslant\vartheta^*\quad\text{if }\mu\in M_1,$$

$$S=\sum_{\mu\in M_2}(t_\mu-\vartheta);$$

of course, we must have $S<1$. We also have $\vartheta^*\leqslant 1$. Otherwise, if we divide the condition $i_1/d_1+\vartheta^* i_2/d_2<t_\mu^*$ by ϑ^* we see that

$$(t_1^*/\vartheta^*, t_2^*/\vartheta^*, \ldots; (\vartheta^*)^{-1}, (\delta^*)^{-1})$$

also occurs as a set of parameters, and $(\vartheta^*)^{-1}<1$. By the last clause of Lemma 4, M_2 consists of a single element, say μ_0, because $t_\mu^*/\vartheta^*=1$ for $\mu\in M_2$. In this case however

$$\vartheta^* = \frac{\vartheta}{1-S} = \frac{\vartheta}{1+\vartheta-t_{\mu_0}} \leqslant 1$$

because $t_{\mu_0}\leqslant 1$, and we reach a contradiction.

Since $\vartheta^*\leqslant 1$ we can apply Dyson's lemma once more and we find

$$\sum_{\mu\in M_2} \frac{\vartheta}{2(1-S)} + \sum_{\mu\in M_1} \frac{1-S}{2\vartheta}\left(\frac{t_\mu}{1-S}\right)^2 \leqslant 1+\max\left(\frac{m-2}{2}, 0\right)\frac{1}{1-S}\delta,$$

which is easily transformed into

$$\sum_{\mu\in M_2}\left(t_\mu-\frac{\vartheta}{2}\right)+\sum_{\mu\in M_1}\frac{1}{2\vartheta}t_\mu^2 \leqslant 1+\max\left(\frac{m-2}{2}, 0\right)\delta;$$

since $\varphi_2(t)=t-\vartheta/2$ if $\vartheta<t\leqslant 1$, we have completed the proof of Theorem 1 when $\vartheta\leqslant 1$. Finally if $\vartheta>1$ the proof of Theorem 1 is reduced to the preceding case, by noting that if $(t_\mu;\vartheta,\delta)$ occurs then $(t_\mu/\vartheta;\vartheta^{-1},\delta^{-1})$ also occurs and noting that we can state Dyson's lemma in the form:

$$\sum\frac{1}{2\vartheta}t_\mu^2 \leqslant 1+\max\left(\frac{m-2}{2}, 0\right)\min(\delta, \delta^{-1})$$

if $(t_\mu;\vartheta,\delta)$ occurs and $0<\vartheta\leqslant 1, 0<t_\mu\leqslant\vartheta$.

III. The auxiliary polynomial

III.1. Let k be an algebraic number field, let α_1,α_2 be algebraic with $k(\alpha_1)=k(\alpha_2)=K$; let $r=[K:k]$ be the degree of α_1 and α_2 over k.

We are interested in the approximation properties of α_1,α_2 by elements of k, say β_1,β_2. We follow Thue's ideas (for the case $k=Q$) and construct a polynomial $P\in k[x_1,x_2]$ with the following properties:

(i) P vanishes at the point (α_1,α_2) to high order;

(ii) P vanishes at the point (β_1,β_2) only to low order;

(iii) the height of P is not too large.

To be precise we want P such that

$$\deg_{x_i} P \leqslant d_i, \quad i=1,2$$

where $d_1 \geqslant d_2$, d_2 is large, and such that for some t, $t>0$, we have

$$\Delta^I P(\alpha_1,\alpha_2)=0,$$

where

$$\Delta^I = \frac{\partial^{i_1+i_2}}{\partial x_1^{i_1}\,\partial x_2^{i_2}},$$

for all $I=(i_1,i_2)$ with

$$\vartheta^{-1}\frac{i_1}{d_1}+\vartheta\frac{i_2}{d_2}<t.$$

In carrying out our estimates we shall suppose that d_2 and hence d_1 are large and eventually go to infinity. The algebraic numbers are kept fixed; $(\alpha_1^{(\mu)},\alpha_2^{(\mu)}),\mu=1,2,\ldots,r$ will denote a set of conjugates of (α_1,α_2). We always assume $(\beta_1,\beta_2)\neq(\alpha_1,\alpha_2)$. Let

$$\varphi_2(t)=\int_0^1\int_0^1 dx\,dy,$$
$$\vartheta^{-1}x+\vartheta y<t.$$

LEMMA 5. *Let P be as before and let us suppose that $r\varphi_2(t)<1$. Let τ be defined by*

$$\varphi_1(\tau)=1-r\varphi_2(t)+\frac{r}{2}\frac{d_2}{d_1}.$$

Then there is $I^*=(i_1^*,i_2^*)$ such that

$$\Delta^{I^*}P(\beta_1,\beta_2)\neq 0$$

and

$$\vartheta^{-1}\frac{i_1^*}{d_1}+\vartheta\frac{i_2^*}{d_2}\leqslant\tau.$$

Proof. Let $(\alpha_1^{(\mu)},\alpha_2^{(\mu)}),\mu=1,\ldots,r$ be the conjugates of (α_1,α_2) over k. Since we assume $K=k(\alpha_1)=k(\alpha_1,\alpha_2)$, the numbers $\alpha_1=\alpha_1^{(1)},\ldots,\alpha_1^{(r)}$, are all distinct; similarly,

$\alpha_2=\alpha_2^{(1)},\dots,\alpha_2^{(r)}$, are all distinct. It is clear that $\beta_i\neq\alpha_i^{(\mu)}$; hence the set of $r+1$ points $\xi_\mu=(\alpha_1^{(\mu)},\alpha_2^{(\mu)})$ if $\mu=1,\dots,r$ and $\xi_{r+1}=(\beta_1,\beta_2)$ is admissible for the application of Theorem 1. We apply Theorem 1 with $t_\mu=t,\mu=1,\dots,r$ and with $t_{r+1}=\tau^*>\tau$ close to τ, to the effect that $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2<\tau^*$ implies $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2<\tau$. Then if we had $\Delta^I P(\beta_1,\beta_2)=0$ for all I with $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2<\tau^*$, we obtain by Theorem 1

$$\sum_{\mu=1}^{r+1}\varphi_2(t_\mu)\leqslant 1+\frac{r}{2}\frac{d_2}{d_1}$$

which implies

$$\varphi_2(\tau)<\varphi_2(\tau^*)\leqslant 1-r\varphi_2(t)+\frac{r}{2}\frac{d_2}{d_1},$$

which contradicts our choice of τ.

It remains to construct P with all the desired properties. This is achieved using the fundamental construction of Thue and Siegel. We prove the following invariant form of the familiar Siegel's lemma.

Let k be a number field, K a finite extension of k, of degree $[K:k]=r$, and let $L_i(\mathbf{x})=\sum_{j=1}^N a_{ij}x_j$, $i=1,2,\dots,M$ be M linear forms with coefficients $a_{ij}\in K$, in the N variables $x_1,\dots,x_N$. We have:

SIEGEL'S LEMMA. ([1]) *There is a constant c_1 depending only on the fields k, K but not otherwise on M, N or the forms L_i, with the following property.*

If $N>rM$ then there is a solution $\mathbf{x}\in k^N$, $\mathbf{x}\neq 0$ to the linear system

$$L_i(\mathbf{x})=0,\quad i=1,\dots,M$$

satisfying

$$h(\mathbf{x})\leqslant c_1(c_1N)^{\frac{rM}{N-rM}}\left(\prod_{i=1}^M h(L_i)\right)^{\frac{r}{N-rM}}$$

where $h(\mathbf{x})=\Pi_v\max_i|x_i|_v$ is the absolute homogeneous height.

Proof. If $k=K$ this is Siegel's lemma as formulated in our paper "On G-functions" ([Bo]). Now suppose $k\in K$, $k\neq K$. There is a number field F such that $k\subset K\subset F$ and F is

([1]) Perhaps, and more appropriately, results of this type should be called Thue's lemma.

a Galois extension of k, with Galois group G, and also is a Galois extension of K, with Galois group H. Clearly H is a subgroup of G with index

$$[G:H] = [K:k] = r.$$

We consider the coefficients a_{ij} of L_i as elements of F and look at the linear system

$$\sum_{j=1}^{N} \gamma(a_{ij}) X_j = 0$$

for $i=1,2,\dots,M$ and $\gamma \in G$, to be solved with $X_j \in F$. Since $\eta(a_{ij})=a_{ij}$ for every $\eta \in H$, the number of independent equations is at most rM (we may restrict our attention to γ running over a set of representatives of cosets of H in G). Now we apply Siegel's lemma for the field F and obtain a solution $\mathbf{X} \in F^N$ to the above system, with $\mathbf{X} \neq \mathbf{0}$ and

$$h(\mathbf{X}) \leqslant c_2(Nc_2)^{\frac{rM}{N-rM}} \prod_{i=1}^{M} \left(\prod_{\gamma \in G/H} h(\gamma(L_i)) \right)^{\frac{1}{N-rM}}.$$

The constant c_2 depends only on the field F. Since $h(\gamma(L_i))=h(L_i)$ for every $\gamma \in G$, our bound for $h(\mathbf{X})$ simplifies to

$$h(\mathbf{X}) \leqslant c_2(Nc_2)^{\frac{rM}{N-rM}} \left(\prod_{i=1}^{M} h(L_i) \right)^{\frac{1}{N-rM}}.$$

The solution we have found is in F^N while we want a solution in k^N. Let $\lambda \in F$, $\lambda \neq 0$. We have

$$\sum_{j=1}^{N} \gamma(a_{ij}) (\lambda X_j) = 0$$

for every $\gamma \in G$, hence

$$\sum_{j=1}^{N} a_{ij} \gamma (\lambda X_j) = 0$$

for every $\gamma \in G$. Now we take traces in k and we find

$$\sum_{j=1}^{N} a_{ij} x_j = 0$$

with

$$x_j = \mathrm{Tr}_{F/k}(\lambda X_j) = \sum_{\gamma \in G} \gamma(\lambda X_j) \in k.$$

If we let λ run through a basis of F over K we cannot always have $\mathrm{Tr}_{F/k}(\lambda \mathbf{X})=\mathbf{0}$ unless $\mathbf{X}=\mathbf{0}$ to start with. This shows that we may choose λ such that $\mathrm{Tr}_{F/k}(\lambda \mathbf{X}) \neq \mathbf{0}$, from a finite set which depends only on F and k, but which is independent of $\mathbf{X}$; only the choice of λ will depend on $\mathbf{X}$.

In order to complete the proof of Siegel's lemma, we have to show that we can choose the solution $\mathbf{X} \in F^N$ so that $h(\mathrm{Tr}_{F/k}(\lambda \mathbf{X}))$ is comparable with $h(\mathbf{X})$. This need not be always true and an additional argument is needed to overcome this difficulty.

Let $\mathbf{X} \in F^N$, $\mathbf{X} \neq \mathbf{0}$. We claim that there is $\mu \in F$ such that

$$\max_j \|\mu X_j\|_v \leqslant 1 \quad (v \text{ finite}),$$

$$\max_j \|\mu X_j\|_v \leqslant A h(\mathbf{X}) \quad (v \text{ infinite}),$$

where we have written for simplicity $\| \ \|_v = | \ |^{d/d_v}$ and where $A=A(F)$ depends only on the field F. Now we complete the proof of Siegel's lemma as follows. We have

$$\max_j \|\lambda \mu X_j\|_v \leqslant \|\lambda\|_v \quad (v \text{ finite}),$$

$$\max_j \|\lambda \mu X_j\|_v \leqslant \|\lambda\|_v A h(\mathbf{X}) \quad (v \text{ infinite}),$$

and we have $\|\lambda\|_v = \|\alpha\|_{\gamma^{-1}v}$ for every $\gamma \in G$ and $\alpha \in F$, $\alpha \neq 0$; this implies

$$\max_{\gamma, j} \|\gamma(\lambda \mu X_j)\|_v \leqslant \max_{\gamma} \|\gamma \lambda\|_v \quad (v \text{ finite}),$$

$$\max_{\gamma, j} \|\gamma(\lambda \mu X_j)\|_v \leqslant \max_{\gamma} \|\gamma \lambda\|_v A h(\mathbf{X}) \quad (v \text{ infinite}),$$

and finally

$$h(\mathrm{Tr}_{F/K}(\lambda \mu \mathbf{X})) \leqslant |G| \prod_v \max_{\gamma, j} \|\gamma(\lambda \mu X_j)\|_v^{d_v/d}$$

19–812904 Acta mathematica 148. Imprimé le 31 août 1982

$$\leqslant |G|\left(\prod_v \max_\gamma |\gamma\lambda|_v\right)\prod_{v|\infty}(Ah(\mathbf{X}))^{d_v/d}$$

$$= |G|h(G\lambda)Ah(\mathbf{X}).$$

Since λ belongs to a fixed finite set, we have $|G|h(G\lambda)A\leqslant c_3$ where c_3 depends only on the fields F, k; this proves the conclusion of Siegel's lemma with $\mathbf{x}=\lambda\mu\mathbf{X}$.

It remains to show that we can find $\mu\neq 0$ with the required properties, namely

$$\|\mu\|_v \leqslant 1/\max_j \|X_j\|_v \quad (v \text{ finite}),$$

$$\|\mu\|_v \leqslant Ah(\mathbf{X})/\max_j \|X_j\|_v \quad (v \text{ infinite}).$$

Let

$$\lambda(v) = \begin{cases} 1/\max_j \|X_j\|_v & (v \text{ finite}), \\ h(X)/\max_j \|X_j\|_v & (v \text{ infinite}). \end{cases}$$

It is clear that $\lambda(v)=1$ for almost all v, $\lambda(v)=\|\pi_v\|_v^{l_v}$ with $l_v\in Z$ and π_v a uniformizing parameter of F_v if v is finite, and also $\Pi_v\lambda(v)^{d_v}=1$ by definition of height. Thus the collection $\{\lambda(v)\}$ forms a ceiling of the field F, as defined by Mahler ([M]). By Mahler's theorem ([M], Section 12, p. 440) if we choose

$$A = d(F)^{2d(F)}|D_F|^{\frac{1}{2}}$$

where $d(F)=[F:Q]$ and D_F is the absolute discriminant of F, there is at least one $\mu\neq 0$ with the required property

$$\|\mu\|_v \leqslant \lambda(v) \text{ if } v \text{ is finite},$$

$$\|\mu\|_v \leqslant A\lambda(v) \text{ if } v \text{ is infinite};$$

this completes the proof of Siegel's lemma.

In what follows, we abbreviate

$$\lambda(x) = \log h(x)$$

and call $\lambda(x)$ the *logarithmic height*.

LEMMA 6. *Let t be given with $r\varphi_2(t)<1$ and let d_1, d_2 be large integers. Then we can find $P \in k[x_1, x_2]$ not identically 0, of degree $\deg_{x_i} P \leqslant d_i$, such that*

$$\Delta^I P(\alpha_1, \alpha_2) = 0$$

for all $I=(i_1, i_2)$ with

$$\vartheta^{-1}\frac{i_1}{d_1}+\vartheta\frac{i_2}{d_2}<t,$$

and such that

$$\lambda(P)=\log h(P)\leqslant\frac{r\varphi_2(t)}{1-r\varphi_2(t)}(d_1\lambda(\alpha_1)+d_2\lambda(\alpha_2)+(d_1+d_2)\log 2)+o(d_1+d_2).$$

Remark. The proof of Lemma 6 will yield a slightly stronger but more complicated bound. This may be useful in dealing with specific cases.

Proof. Let us write

$$P(z_1, z_2) = \sum x_{j_1, j_2} z_1^{j_1} z_2^{j_2},$$

regarding the coefficients of P as unknowns in the field k. We have

$$\frac{1}{I!}\Delta^I P(\alpha_1, \alpha_2) = \sum_{j_1, j_2} \binom{j_1}{i_1}\binom{j_2}{i_2}\alpha_1^{j_1-i_1}\alpha_2^{j_2-i_2}x_{j_1, j_2},$$

where $I!=i_1!\,i_2!$. The associated linear from L_I has height

$$\begin{aligned}
h(L_I) &= \prod_v \max_J \left|\binom{j_1}{i_1}\binom{j_2}{i_2}\alpha_1^{j_1-i_1}\alpha_2^{j_2-i_2}\right|_v \\
&\leqslant \prod_v \max_j \left|\binom{j}{i_1}\alpha_1^{j-i_1}\right|_v \cdot \prod_v \max_j \left|\binom{j}{i_2}\alpha_2^{j-i_2}\right|_v \\
&\leqslant \binom{d_1}{i_1}\prod_v \max(1, |\alpha_1|_v)^{d_1-i_1}\cdot\binom{d_2}{i_2}\prod_v \max(1, |\alpha_2|_v)^{d_2-i_2} \\
&= \binom{d_1}{i_1}\binom{d_2}{i_2} h(\alpha_1)^{d_1-i_1}h(\alpha_2)^{d_2-i_2}.
\end{aligned}$$

The number N of unknowns is $(d_1+1)(d_2+1)$ while the number M of equations is the number of solutions of $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2<t$, which is asymptotic to $\varphi_2(t)d_1d_2$ as d_1, d_2 go independently to ∞. Thus

$$N\sim d_1d_2, \quad N-rM\sim(1-r\varphi_2(t))d_1d_2$$

and Siegel's lemma shows that we can find $x_J\in k$, not all 0, such that $\Delta^I P(\alpha_1,\alpha_2)=0$ and

$$\lambda(P)\leqslant\frac{N}{N-rM}\log(c_1N)+\frac{r}{[1+o(1)]\,d_1d_2(1-r\varphi_2(t))}\cdot\left(\sum_I\log\binom{d_1}{i_1}+\log\binom{d_2}{i_2}+(d_1-i_1)\lambda(\alpha_1)+(d_2-i_2)\lambda(\alpha_2)\right),$$

with $c_1=c_1(k, K)$. The first term is $O(\log(d_1d_2))=o(d_1+d_2)$.
Also

$$\log\binom{d}{i}\leqslant d\log 2,$$

$$\sum_I 1\leqslant\varphi_2(t)d_1d_2,$$

and Lemma 6 follows.

Remark. If $t<\min(\vartheta^{-1},\vartheta)$ a more careful estimate yields the following result. Let

$$g(u)=\frac{1}{6}u^3\log\frac{1}{u}+\frac{1}{6}(1-u)^3\log\frac{1}{1-u}-\frac{1}{6}u+\frac{5}{12}u^2.$$

Then we have

$$\lambda(P)\leqslant\frac{2r}{2-rt^2}\left(\frac{1}{\vartheta^2}g(\vartheta t)+\left(\frac{t^2}{2}-\vartheta\frac{t^3}{6}\right)\lambda(\alpha_1)\right)d_1+\frac{2r}{2-rt^2}\left(\vartheta^2 g\left(\frac{1}{\vartheta}t\right)+\left(\frac{t^2}{2}-\vartheta^{-1}\frac{t^3}{6}\right)\lambda(\alpha_2)\right)d_2.$$

We also have $g(u)\leqslant\frac{1}{4}u^2$ for $0<u<1$, and

$$\lambda(P)\leqslant\frac{rt^2}{2-rt^2}(d_1\lambda(\alpha_1)+d_2\lambda(\alpha_2)+\tfrac{1}{2}(d_1+d_2)).$$

IV. The Thue-Siegel theorem

IV.1. Let $k, \alpha_1, \alpha_2, K=k(\alpha_1)=k(\alpha_2)$ be as in the preceding section, let $r=[K:k]\geqslant 2$ be the degree of α_1, α_2 over k and let β_1, β_2 be two approximations to α_1, α_2 relative to a same set S of places v of k. This is to be understood in the following sense. Let S be a finite set of places of k together with an extension to the field k and for $v\in S$ let $|\ |_v$ be the absolute value associated to v normalized relative to the field k; then we say that $\beta\in k$ approximates α relative to the set S if

$$|\alpha-\beta|_v<1 \text{ for } v\in S.$$

THEOREM 2. *Let k, K, α_1, α_2, β_1, β_2, S be as before and let $\vartheta, t, \tau, \delta_1, \delta_2$ be positive numbers such that*

$$r\varphi_2(t)<1, \quad 0<\tau<t$$

and

$$\frac{r}{2}\frac{\delta_2}{\delta_1}\leqslant r\varphi_2(t)+\varphi_2(\tau)-1.$$

Then we have

$$\begin{aligned}(t-\tau)\sum_{v\in S}\min\left(\vartheta\delta_1\log\frac{1}{|\alpha_1-\beta_2|_v}, \vartheta^{-1}\delta_2\log\frac{1}{|\alpha_2-\beta_2|_v}\right)&\\ \leqslant\delta_1\left(\lambda(\beta_1)+\frac{1}{1-r\varphi_2(t)}\lambda(\alpha_1)+\frac{\log 3}{1-r\varphi_2(t)}\right)&\\ +\delta_2\left(\lambda(\beta_2)+\frac{1}{1-r\varphi_2(t)}\lambda(\alpha_2)+\frac{\log 3}{1-r\varphi_2(t)}\right)&\end{aligned}$$

COROLLARY (Thue-Siegel theorem). *Let v be a place of k extended to K, let $\alpha\in K$, $k(\alpha)=K$, and let $\varepsilon>0$. There are only finitely many $\beta\in k$ such that*

$$|\alpha-\beta|_v<h(\beta)^{-\sqrt{2r}-\varepsilon}$$

This corollary is the refinement of the Thue-Siegel theorem obtained by Dyson and Gelfond. As usual, this is proven by contradiction. If we had infinitely many solutions, we could find β_1, β_2 with $\lambda(\beta_1), \lambda(\beta_2)$ arbitrarily large, $\lambda(\beta_1)/\lambda(\beta_2)$ arbitrarily small and now we could apply Theorem 2 with $\delta_1=1/\lambda(\beta_1)$, $\delta_2=1/\lambda(\beta_2)$, $\vartheta=1$, t arbitrarily close to $\sqrt{2/r}$, τ arbitrarily small, to obtain a contradiction. In fact this argument is so well-

known that we can safely omit the details. On the other hand, our Theorem 2 is very explicit and it is possible to draw some interesting consequences out of it. Let us consider in detail the case in which S consists of a single place v.

We assume

$$0<\tau<t<\min(\vartheta^{-1},\vartheta)$$

so that $\varphi_2(t)=\frac{1}{2}t^2$, $\varphi_2(\tau)=\frac{1}{2}\tau^2$ and we choose

$$\delta_1=\left(\log h(\beta_1)+\frac{2}{2-rt^2}\log 3h(\alpha_1)\right)^{-1}$$

$$\delta_2=\left(\log h(\beta_2)+\frac{2}{2-rt^2}\log 3h(\alpha_2)\right)^{-1}$$

Then our inequality becomes

$$(t-\tau)\min\left(\vartheta\delta_1\log\frac{1}{|\alpha_1-\beta_1|_v},\vartheta^{-1}\delta_2\log\frac{1}{|\alpha_2-\beta_2|_v}\right)\leqslant 2$$

that is:
either

$$|\alpha_2-\beta_2|_v\geqslant(3h(\alpha_2))^{-\frac{4\vartheta}{(2-rt^2)(t-\tau)}}h(\beta_2)^{-\frac{2\vartheta}{t-\tau}}.$$

or

$$|\alpha_1-\beta_1|_v\geqslant(3h(\alpha_1))^{-\frac{4/\vartheta}{(2-rt^2)(t-\tau)}}h(\beta_1)^{-\frac{2/\vartheta}{t-\tau}},$$

On the other hand, it could be that our choice of δ_1,δ_2 does not satisfy the condition in the hypotheses of Theorem 2, so that in this case we cannot say that the preceding alternative holds. In this case however we must have $r(\delta_2/\delta_1)>rt^2+\tau^2-2$ whence we obtain

$$\log h(\beta_2)+\frac{2}{2-rt^2}\log 3h(\alpha_2)<\frac{r}{rt^2+\tau^2-2}\left(\log h(\beta_1)+\frac{2}{2-rt^2}\log 3h(\alpha_1)\right).$$

We have shown:

THEOREM 3. *Let $k\subset K$ be number fields, let $r=[K:k]\geqslant 2$ and let v be a place of k extended to K, with absolute value $|\ |_v$ normalized relative to k. Let λ,t,τ be positive real numbers with*

$$0<\sqrt{2-rt^2}<\tau<t<\sqrt{\frac{2}{r}}\leqslant\min(\vartheta^{-1},\vartheta).$$

Let α_i,β_i, $i=1,2$ be such that $k(\alpha_i)=K$, $\beta_i\in k$, $|\alpha_i-\beta_i|_v<1$. Then we have: either

$$|\alpha_1-\beta_1|_v\geqslant(3h(\alpha_1))^{-\frac{4/\vartheta}{(2-rt^2)(t-\tau)}}h(\beta_1)^{-\frac{2/\vartheta}{t-\tau}},$$

or

$$|\alpha_2-\beta_2|_v\geqslant(3h(\alpha_2))^{-\frac{4\vartheta}{(2-rt^2)(t-\tau)}}h(\beta_2)^{-\frac{2\vartheta}{t-\tau}},$$

or

$$\log h(\beta_2)+\frac{2}{2-rt^2}\log 3h(\alpha_2)<\frac{r}{rt^2+\tau^2-2}\left(\log h(\beta_1)+\frac{2}{2-rt^2}\log 3h(\alpha_1)\right).$$

The special case $\vartheta=1$ can be given a simpler formulation.

THEOREM 4. *Let $k\subset K$ be number fields, let $r=[K:k]\geqslant 2$ and let v be a place of k extended to K, with absolute value $|\ |_v$, normalized relative to k. Let also t, τ be positive real numbers with*

$$0<\sqrt{2-rt^2}<\tau<t<\sqrt{2/r}.$$

Then the following two statements hold.

(A) *For all but finitely many pairs (α,β) with $k(\alpha)=K$, $\beta\in k$ we have*

$$|\alpha-\beta|_v\geqslant(3h(\alpha))^{-\frac{4}{(2-rt^2)(t-\tau)}}h(\beta)^{-\frac{2}{t-\tau}}.$$

(B) *The quantity*

$$\log h(\beta)+\frac{2}{2-rt^2}\log 3h(\alpha)$$

lies in some fixed interval

$$\left(X,\frac{r}{rt^2+\tau^2-2}X\right)$$

as (α,β) runs over all pairs which do not satisfy (A).

It is now clear that if we can determine one pair which does not satisfy (A), then we can determine an interval containing

$$\log h(\beta)+\frac{2}{2-rt^2}\log 3h(\alpha)$$

and then determine effectively all exceptions to (A). We shall give later on some examples of fields K admitting pairs (α,β) which do not satisfy (A) and for which the exponent $2/(t-\tau)$ is $O(\sqrt{r})$; for these fields, we then obtain an effective Thue-Siegel theorem.

IV.2. Let P be the polynomial constructed in the preceding section. We have P, t, τ, I^*, d_i with $\deg_{x_i} P \leqq d_i$, and we have the following facts:

(i) $r\varphi_2(t)<1$

(ii) $\varphi_2(\tau)=1-r\varphi_2(t)+\frac{r}{2}\frac{d_2}{d_1}\leqq 1$

(iii) $\Delta^I P(\alpha_1, \alpha_2)=0 \quad \text{for} \quad \vartheta^{-1}\frac{i_1}{d_1}+\vartheta\frac{i_2}{d_2}<t$

(iv) there is $I^*=(i_1^*, i_2^*,)$ with $\Delta^{I^*}P(\beta_1,\beta_2)\neq 0$ and

$$\vartheta^{-1}\frac{i_1^*}{d_1}+\vartheta\frac{i_2^*}{d_2}\leqq\tau$$

(v) $\lambda(P)\lesssim\frac{r\varphi_2(t)}{1-r\varphi_2(t)}(d_1\lambda(\alpha_1)+d_2\lambda(\alpha_2)+(d_1+d_2)\log 2)$

(vi) if $v\in S$ then $|\alpha_i-\beta_i|_v<1 \quad \text{for } i=1, 2;$

here (iv) and (v) follow from Lemma 4 and Lemma 5.

We abbreviate $I!=i_1!\,i_2!$ and

$$\binom{J}{I}=\binom{j_1}{i_1}\binom{j_2}{i_2}.$$

Let $\gamma=(1/I^*!)\Delta^{I^*}P(\beta_1,\beta_2)$ so that $\gamma\neq 0$; note also that $\gamma\in k$ because $P\in k\,[x_1,x_2]$ and $\beta_1,\beta_2\in k$. Thus the product formula yields $\Sigma_v\log|\gamma|_v=0$ where the summation is over all places v of k and where $|\ |_v$ is normalized relative to k. Now we proceed to estimate $\log|\gamma|_v$ separately for each v, as follows.

Case I. $v\in S$, *v finite.*
We have

$$\gamma=\frac{1}{I^*!}\Delta^{I^*}P(\beta_1,\beta_2)=\sum_I\binom{I^*+I}{I}\frac{1}{(I^*+I)!}\Delta^{I^*+I}P(\alpha_1,\alpha_2)(\beta_1-\alpha_1)^{i_1}(\beta_2-\alpha_2)^{i_2}$$

and, by property (iii), we also have $\Delta^{I^*+I}P(\alpha_1,\alpha_2)=0$ for all I with $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2<t-\vartheta^{-1}i_1^*/d_1-\vartheta i_2^*/d_2$ and *a fortiori* for all I with $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2<t-\tau$.

Since $(1/I!)\Delta^I$ does not introduce denominators we see, using the fact that v is finite, that

$$\log\left|\frac{1}{I!}\Delta^I P(\alpha_1,\alpha_2)\right|_v \leqslant \log|P|_v + d_1\log^+|\alpha_1|_v + d_2\log^+|\alpha_2|_v.$$

Also $\log|\beta_i-\alpha_i|_v<0$, hence if $t-\tau>0$ we have

$$\max^*\log\left|(\beta_1-\alpha_1)^{i_1}(\beta_2-\alpha_2)^{i_2}\right|_v$$
$$\leqslant -(t-\tau)\min\left(\vartheta d_1\log\frac{1}{|\alpha_1-\beta_1|_v}, \vartheta^{-1}d_2\log\frac{1}{|\alpha_2-\beta_2|_v}\right)$$

where max* is taken over all I with $\vartheta^{-1}i_1/d_1+\vartheta i_2/d_2\geqslant t-\tau$. Thus we have shown that in Case I if $t>\tau$ we have

$$\log|\gamma|_v \leqslant \log|P|_v + d_1\log^+|\alpha_1|_v + d_2\log^+|\alpha_2|_v$$
$$-(t-\tau)\min\left(\vartheta d_1\log\frac{1}{|\alpha_1-\beta_1|_v}, \vartheta^{-1}d_2\log\frac{1}{|\alpha_2-\beta_2|_v}\right).$$

Case II. $v\in S$, v *infinite*.

In this case we proceed as in Case I, but taking into account the contribution arising from differentiating P. Let $\varepsilon_v=1$ if v is real, $\varepsilon_v=2$ if v is complex. Then we obtain

$$\log|\gamma|_v \leqslant \log|P|_v + d_1\log^+|\alpha_1|_v + d_2\log^+|\alpha_2|_v + \frac{\varepsilon_v}{[k:Q]}(d_1+d_2)\log 3$$
$$-(t-\tau)\min\left(\vartheta d_1\log\frac{1}{|\alpha_1-\beta_1|_v}, \vartheta^{-1}d_2\log\frac{1}{|\alpha_2-\beta_2|_v}\right)+o(d_1+d_2).$$

Case III. $v\notin S$, v *finite*.

In this case we do not use that Taylor series expansion for $(1/I^*!)\,\Delta^{I^*}P$ but rather estimate directly this quantity. Thus we obtain

$$\log|\gamma|_v \leqslant \log|P|_v + d_1\log^+|\beta_1|_v + d_2\log^+|\beta_2|_v.$$

Case IV. $v\notin S$, v *infinite*.

In this case we proceed as in Case III, but taking into account the contribution arising from differentiation in $(1/I^*!)\,\Delta^{I^*}P$. Then we obtain

$$\log|\gamma|_v \leqslant \log|P|_v + d_1 \log^+|\beta_1|_v + d_2 \log^+|\beta_2|_v + \frac{\varepsilon_v}{[k:Q]}(d_1+d_2)\log 2 + o(d_1+d_2).$$

With respect to the estimates in Cases I and II, we note that we have $|\alpha_i - \beta_i|_v < 1$ in these cases because $v \in S$. Now it is clear that $\log^+|\alpha_i|_v = \log^+|\beta_i|_v$ if $v \in S$, v finite, while $\log^+|\alpha_i|_v \leqslant \log^+|\beta_i|_v + (\varepsilon_v/[k:Q])\log 2$ if $v \in S$ and v is infinite, and thus we can replace $\log^+|\alpha_i|_v$ by $\log^+|\beta_i|_v + (\varepsilon_v/[k:Q])\log 2$ in our estimates.

We combine our local estimates of $\log|\gamma|_v$ with the global result $\Sigma \log|\gamma|_v = 0$ and we find

$$0 = \Sigma \log|\gamma|_v \leqslant \Sigma \log|P|_v + d_1 \Sigma \log^+|\beta_1|_v + d_2 \Sigma \log^+|\beta_2|_v + \left(\sum_{v|\infty} \frac{\varepsilon_v}{[k:Q]}\right)(d_1+d_2)\log 6 + o(d_1+d_2)$$
$$-(t-\tau)\sum_{v \in S} \min\left(\vartheta d_1 \log\frac{1}{|\alpha_1-\beta_1|_v}, \vartheta^{-1} d_2 \log\frac{1}{|\alpha_2-\beta_2|_v}\right)$$

which simplifies to

$$(t-\tau)\sum_{v \in S} \min\left(\vartheta d_1 \log\frac{1}{|\alpha_1-\beta_1|_v}, \vartheta^{-1} d_2 \log\frac{1}{|\alpha_2-\beta_2|_v}\right)$$
$$\leqslant d_1\lambda(\beta_1) + d_2\lambda(\beta_2) + (d_1+d_2)\log 6 + \lambda(P) + o(d_1+d_2).$$

By property (v) we have

$$\lambda(P) \leqslant \frac{r\varphi_2(t)}{1-r\varphi_2(t)}(d_1\lambda(\alpha_1) + d_2\lambda(\alpha_2) + (d_1+d_2)\log 2)$$

and if we combine this bound with the last inequality we obtain, noting that $r \geqslant 2$ and $(r+1)\varphi_2(t) > 1$ because $\tau < t$:

$$(t-\tau)\sum_{v \in S} \min\left(\vartheta d_1 \log\frac{1}{|\alpha_1-\beta_1|_v}, \vartheta^{-1} d_2 \log\frac{1}{|\alpha_2-\beta_2|_v}\right)$$
$$\leqslant d_1\left(\lambda(\beta_1) + \frac{1}{1-r\varphi_2(t)}\lambda(\alpha_1) + \frac{\log 3}{1-r\varphi_2(t)}\right)$$
$$+ d_2\left(\lambda(\beta_2) + \frac{1}{1-r\varphi_2(t)}\lambda(\alpha_2) + \frac{\log 3}{1-r\varphi_2(t)}\right).$$

This inequality holds with $\varphi_2(\tau)=1=r\varphi_2(t)+\frac{1}{2}r\,d_2/d_1$ whenever $r\varphi_2(t)<1$ and $0<\tau<t$.

If we divide both sides of the asymptotic inequality above by d_1 and let d_1, d_2 go to ∞ keeping the ratio d_2/d_1 fixed, we see that the asymptotic inequality above can be replaced by an exact inequality. This completes the proof of Theorem 2.

Remark. If we deal with specific cases it appears that interesting results can be obtained only if $t<\min(\vartheta^{-1}, \vartheta)$ and hence $\varphi_2(t)=\frac{1}{2}t^2$, $\varphi_2(\tau)=\frac{1}{2}\tau^2$. In this case however we can use the remark at the end of the proof of Lemma 6 to obtain slightly better results.

V. Some cases of effectiveness

V.1. Let $m\geq 3$ be a positive integer and let α be the real root $\alpha>1$ of the irreducible equation

$$x^r-mx^{r-1}+1=0.$$

We take $k=Q$, $K=Q(\alpha)$ and for v we take the real place of Q, extended to K so that $\alpha>1$, i.e. $|\ |_v$ is the usual euclidean absolute value $|\ |$ in $\mathbf{R}$.

The equation $\alpha^r-m\alpha^{r-1}+1=0$ yields

$$m^{-r+1}<|\alpha-m|=|\alpha|^{-r+1}<(m-1)^{-r+1}$$

which shows that $\beta_1=m^{-1}$ is an excellent approximation to $\alpha_1=\alpha^{-1}$. We have

$$|\alpha_1-\beta_1|<(m-1)^{-r-1}.$$

We want to show that if m and r are sufficiently large then this approximation is so good that it does not verify (A) of Theorem 4.

First of all, we compute $h(\alpha)$ as follows. Since α is an algebraic integer, we have

$$\log|\alpha|_v\leq 0 \quad \text{if} \quad v \text{ is finite.}$$

If instead v is infinite, we have

$$\log|\alpha|_v<0 \quad \text{for all } v\neq v^*$$

where v^* is the real place of K for which $\alpha>1$. Finally we have

$$\log|\alpha|_{v^*}=\frac{1}{r}\log|\alpha|<\frac{1}{r}\log m.$$

If we put together these local estimates we find

$$\lambda(\alpha_1)=\lambda(\alpha)=\log h(\alpha)<\frac{1}{r}\log m.$$

It is obvious that for $\beta_1=m^{-1}$ we have

$$\lambda(\beta_1)=\log h(\beta_1)=\log m.$$

In view of these estimates we see that the pair (α_1,β_1) does not satisfy the bound (A) in Theorem 4 as soon as

$$(m-1)^{-r-1}<\left(3m^{\frac{1}{r}}\right)^{-\frac{4}{(2-rt^2)(t-\tau)}}m^{-\frac{2}{t-\tau}};$$

if m is sufficiently large, we need only

$$r+1>\frac{2}{t-\tau}+\frac{4}{r(2-rt^2)(t-\tau)}$$

with $0<\sqrt{2-rt^2}<\tau<t<\sqrt{2/r}$. If we set $t=\sqrt{2/(r+a)}$ where $0<a<1$ we have $2-rt^2=2a/(r+a)=at^2$ and the condition on τ becomes $\sqrt{a}\,t<\tau<t$; of course, we can choose τ arbitrarily close to $\sqrt{a}\,t$, at the expense of making $r/(rt^2+\tau^2-2)$ very large.

It is easily seen that we can fulfill the condition

$$r+1>\frac{2}{t-\tau}+\frac{4}{r(2-rt^2)(t-\tau)}$$

by choosing $r\geq 200$ and $t=\sqrt{2/(r+a)}$ with $a=0.35$. For example a rough numerical calculation, of which we omit the details, leads to the following explicit result.

Example 1. *Let K be the field generated by the root $x>1$ of the equation*

$$x^{200}-mx^{199}+1=0,$$

where m is an integer $m\geq 10^{1731}$. Let $\alpha\in K$ generate K over Q. Then

$$\left|\alpha-\frac{p}{q}\right|\geq 10^{-13656}h(\alpha)^{-28622}(\max(|p|,|q|))^{-50}$$

for all p, q with $(p,q)=1$ and with

$$\max(|p|,|q|)\geq m^{8076815}h(\alpha)^{-572}.$$

We may combine this result with the trivial Liouville estimate and obtain a lower bound which holds for all α which generate K over Q and for all p/q. If we formulate our inequality in terms of $H(\alpha)$, the maximum of the coefficients of an irreducible equation for α over Z, we obtain

Example 2. *Let K be the field of Example* 1 *and let $Q(\alpha)=K$. Then we have for all p/q:*

$$\left|\alpha-\frac{p}{q}\right| \geqslant (10m)^{-10^{10}} H(\alpha)^{-144} H\left(\frac{p}{q}\right)^{-50}.$$

V.2. It is clear from what preceeds that effective results for the equation $x^r-mx^{r-1}+1=0$ usually need r or m to be very large, so that no result of practical or intrinsic interest can be obtained. On the other hand, since it appears that the above examples are the first fields not of type $Q(\sqrt[n]{a/b})$ in which a uniform result of Thue-Siegel type holds effectively for every generating element of the field, we believe that it is of some theoretical interest to investigate fields of lowest possible degree for which a non-trivial result can be obtained and also to investigate the best exponents we can get for fields of large degree.

In order to obtain results on these lines we use Theorem 3 with various values for ϑ. We take the same equation, $x^r-mx^{r-1}+1=0$, and choose τ arbitrarily close to $\sqrt{2-rt^2}$. We take (α_1,β_1) as our anchor pair and allow $h(\beta_2)$ to be as large as needed to obtain our conclusions, of course in an effective way. This means that the third alternative in Theorem 3 may be disregarded and hence, if the first alternative does not hold, the second alternative must be true. If m is large enough the first alternative of Theorem 3 cannot hold if $\sqrt{2/r} \leqslant \min(\vartheta^{-1}, \vartheta)$ and

$$r+1>\frac{2/\vartheta}{(t-\tau)}\left(1+\frac{2}{r(2-rt^2)}\right)$$

and we can take here $\tau=\sqrt{2-rt^2}$ for the purpose of checking this inequality. If this inequality holds then the Thue-Siegel exponent for $h(\beta_2)$ will be

$$\frac{2\vartheta}{t-\tau}.$$

If we write $t=\sqrt{2/(r+a)}$, $t=\sqrt{a}\,t$ and if we choose ϑ in an optimal way we obtain

$$\frac{2\vartheta}{t-\tau}=\frac{2}{(1-\sqrt{a})^2}\left(1+\frac{1}{a}+\frac{1}{r}\right)\frac{r+a}{r+1}<\frac{2}{(1-\sqrt{a})^2}\left(1+\frac{1}{a}\right)$$

provided $a \leq \frac{1}{2}$, which we may suppose. Since

$$\min_a \frac{2}{(1-\sqrt{a})^2}\left(1+\frac{1}{a}\right) = 39.2573250\ldots$$

at $a=0.20556943\ldots$, we have obtained

Example 3. Let $r \geq 40$ *and let* K *be the field generated by the root* $x>1$ *of the equation*

$$x^r - mx^{r-1} + 1 = 0,$$

where $m \geq m_0(r)$ *and* $m_0(r)$ *is effectively computable. Let also* α *be such that* $Q(\alpha)=K$. *There is an effectively computable* $q_0(\alpha)$ *such that for every* p/q *with* $H(p/q) \geq q_0(\alpha)$ *we have*

$$\left|\alpha - \frac{p}{q}\right| > H\left(\frac{p}{q}\right)^{-39.2574}.$$

We have computed some examples with specific values of m and r, with the goal of finding non-trivial effective results for equations with not exceedingly large coefficients. The following example represents the result of our search for the case $\vartheta=1$, after carrying out the majorizations with great precision.

Example 4. *Let* K *be the field generated by the root* $x>1$ *of the equation*

$$x^{3216} - 2469528\, x^{3215} + 1 = 0.$$

There is an effectively computable absolute constant $\mu<3216$ *with the following property. Let* α *be such that* $Q(\alpha)=K$. *Then there is an effectively computable* $q_0(\alpha)$ *such that for every* p/q *with* $H(p/q) \geq q_0(\alpha)$ *we have*

$$\left|\alpha - \frac{p}{q}\right| > H\left(\frac{p}{q}\right)^{-\mu}.$$

If we use the more precise estimate in the remark at the end of the proof of Lemma 6 and if we choose r very large and also ϑ very large, of order $\sqrt{r}$, then we can obtain effective results for equations with small coefficients.

Example 5. *Let* K *be the field generated by the root* $x>1$ *of the equation*

$$x^r - mx^{r-1} + 1 = 0,$$

where $m \geq 2561$ and where $r \geq r_0$ for some effectively computable constant r_0. Let also α be such that $Q(\alpha)=K$. There is an effectively computable $q_0(\alpha)$ such that for every p/q with $H(p/q) \geq q_0(\alpha)$ we have

$$\left|\alpha-\frac{p}{q}\right| > H\left(\frac{p}{q}\right)^{-(1-10^{-4})r}.$$

V.3. It should be obvious by now that results of the same kind apply to a class of equations of a fairly general type. Rather than prescribing α and trying to find a good approximation β, examples may be found by choosing β first and deforming it slightly into an algebraic number of much higher degree. A typical situation would be to consider a polynomial $f(x, m)$ depending on a parameter m and with bounded height and fixed degree and deform the equation $f(x, m)=0$ (the equation for β) into the equation $f(x, m)=R(x)$ where $R(x)$ is a rational function with bounded height with a zero of very high order at ∞; then we can take $f(x, m)=R(x)$ as our equation for α.

On the other hand, it would be nice if one could utilize for the purpose of obtaining effective results the remarkable approximations investigated by Stark [S] in the case of certain cubic fields. It appears however that our procedures are not sufficiently refined and new ideas may be needed in order to achieve this goal.

Further improvements of our results may come from a sharpening of Lemma 6 (this could lead to the best possible exponent 2, in every case), but it is also possible that the consideration of several approximations $|\alpha_i-\beta_i|_v$, $i=1, 2, \ldots, m$ for suitably independent pairs (α_i, β_i), is needed, as is the case with Roth's theorem. Here the difficulties are of two kinds. In the hypothesis of Roth's lemma, one needs $\lambda(\beta_1)$ large and $\lambda(\beta_{i+1})/\lambda(\beta_i)$ also large in order to obtain a useful result. In Roth's lemma, $\lambda(\beta_1)$ large means

$$\lambda(\beta_1) > Ar\lambda(\alpha_1)$$

with $A>1$, which is too strong a condition for our purposes. On the other hand, Dyson's lemma has the advantage of being free from considerations of heights. Unfortunately, the extension of Dyson's lemma to more than two variables is still lacking, as well as a formulation of Roth's lemma which does not require the β_i's to be of rapidly increasing heights; we have no contribution to offer here for the solution of these problems.

References

[Ba1] BAKER, A., Rational approximation to $\sqrt[3]{2}$ and other algebraic numbers. *Quart. J. Math. Oxford*, 15 (1964), 375–383.

[Ba2] — Simultaneous rational approximations to certain algebraic numbers. *Proc. Camb. Phil. Soc.*, 63 (1967), 693–702.

[Ba3] — The theory of linear forms in logarithms. In *Transcendence Theory: Advances and Applications*. Baker, A., & Masser, D. W. ed. Academic Press 1977.

[Ba4] — Recent advances in transcendence theory. *Proceedings of International Conference on Number Theory,* Moscow 1971, 67–69.

[Bo] BOMBIERI, E., On G-functions. In *Recent Progress in Analytic Number Theory*, vol II. H. Halberstam & C. Hooley ed., Academic Press 1981.

[D] DYSON, F., The approximation to algebraic numbers by rationals. *Acta Math.*, 79 (1947), 225–240.

[F] FELDMAN, N. I.. An effective refinement of the exponent in Liouville's theorem (Russian). *Izv. Akad. Nauk SSSR Ser. Mat.*, 35 (1971), 973–990. Also *Math. USSR-Izv.*, 5 (1971), 985–1002.

[G] GELFOND, A. O., *Transcendental and algebraic numbers*. English translation by L. F. Boron. Dover Publications Inc., New York 1960.

[H] HYYRÖ, S., Über rationale Näherungswerte algebraischer Zahlen. *Ann. Acad. Sci. Fenn. Ser. A I. Math.*, 376 (1965), 1–15.

[L] LANG, S., *Diophantine geometry*. Interscience Publishers Inc., New York–London 1962.

[M] MAHLER, K., Inequalities for ideal bases in algebraic number fields. *J. Austral. Math. Soc.*, IV (1964), 425–448.

[S] STARK, H. M., An explanation of some exotic continued fractions found by Brillhart. *Computers in Number Theory* (Proc. Sci. Res. Council Atlas Symp. No. 2 Oxford 1969), pp. 21–35. Academic Press, London 1971.

[T] THUE, A., *Selected mathematical papers*. Universitet-forlaget Oslo–Bergen–Tromsø, 1977.

[W] WEIL, A., Arithmetic on algebraic varieties. *Ann. of Math.*, 53 (1951), 412–444.

Received September 21, 1981