ANNALI DELLA
SCUOLA NORMALE SUPERIORE DI PISA
Classe di Scienze

ENRICO BOMBIERI
PAULA B. COHEN

Siegel's lemma, Padé approximations and jacobians

Annali della Scuola Normale Superiore di Pisa, Classe di Scienze 4e série, tome 25, nº 1-2 (1997), p. 155-178
<http://www.numdam.org/item?id=ASNSP_1997_4_25_1-2_155_0>

Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)
Vol. XXV (1997), pp. 155-178

Siegel's Lemma, Padé Approximations and Jacobians

ENRICO BOMBIERI - PAULA B. COHEN

with an appendix by UMBERTO ZANNIER

1. - Introduction

We consider a Padé approximation problem arising in the theory of algebraic functions of one variable. Let K be a number field and let C be a non-singular projective curve of genus g, defined over K.

Let $x \in K(C)$ be a non-constant rational function in the function field of C/K, hence defining a surjective rational morphism $x : C \longrightarrow \mathbb{P}^1$. Then $K(C)$ is a finite extension of $K(x)$ of degree $n = [K(C) : K(x)]$ equal to the degree of the function x. By the theorem of the primitive element, there is an element $y \in K(C)$ such that $K(C) = K(x)(y)$, and then $1, y, \dots, y^{n-1}$ form a basis of $K(C)$ as a $K(x)$-vector space. Since $1, y, \dots, y^n$ are linearly dependent over $K(x)$ we deduce that there is a polynomial

$$f(x, y) = A_0(x)y^n + A_1(x)y^{n-1} + \dots + A_n(x) \in K[x, y]$$

such that $f(x, y) = 0$ identically on C. This gives us a birational model of C as a plane curve of degree n in y, by means of the morphism $\pi : C \longrightarrow \mathbb{P}^2$ given by $\pi(P) = (1 : x(P) : y(P))$. Conversely, every such birational model of C arises in this way.

We assume that $x^{-1}(0)$ consists of n distinct points $Q_1, \dots, Q_n$, so that $x : C \longrightarrow \mathbb{P}^1$ is unramified over 0, and fix once and for all $Q \in x^{-1}(0)$. Then, in the above construction, the rational function y can be chosen so that y is a local uniformiser at Q, or in other words $y(Q) = 0$ and $f_y(0, 0) \neq 0$, where f_y denotes the partial derivative with respect to y.

Finally we shall assume, possibly by replacing K by a suitable finite extension, that all points Q_i are defined over K.

The rational function y on C may be viewed as an algebraic function of degree n of x, giving rise to n branches corresponding to the n roots of the equation $f(x, y) = 0$. Because of our assumptions, there is a unique such branch at $(0, 0)$, which we shall denote by

$$u(x) = a_1 x + a_2 x^2 + \dots .$$

Padé approximations associated to algebraic functions are of considerable importance in diophantine approximations and transcendence. Let $u_0(x), u_1(x), \dots, u_s(x)$ be $s+1$ functions holomorphic in a neighborhood of $x = 0$. Then for any given positive integers $m_0, m_1, \dots, m_s$ we can find $s+1$ polynomials $P_i(x)$, of degree at most m_i and not all 0, such that the function

$$P_0(x)u_0(x) + P_1(x)u_1(x) + \dots + P_s(x)u_s(x)$$

has a zero at the origin of order at least $m_0 + m_1 + \dots + m_s + s$ (one needs to solve a homogeneous linear system of $m_0 + m_1 + \dots + m_s + s$ equations in $m_0+m_1+\dots+m_s+s+1$ unknowns, the coefficients of the polynomials P_i). One then talks about a type I (or Latin) Padé approximation system for the vector $(u_0(x), u_1(x), \dots, u_s(x))$. The vector $(m_0, m_1, \dots, m_s)$ is called the weight of $(P_0(x), P_1(x), \dots, P_s(x))$.

In this paper we shall consider only the case of equal weights $(m, m, \dots, m)$.

The following well-known example, considered in detail by K. Mahler [9], is particularly interesting. Mahler studied the case in which $u_j(x) = (1-x)^{\omega_j}$ where the ω_j are rational numbers. For example, take

$$u_j(x) = (1-x)^{j/n} = 1 - \binom{j/n}{1}x + \binom{j/n}{2}x^2 - \dots$$

for $j = 0, 1, \dots, s$ and consider the corresponding Padé approximations with equal weights $(m, m, \dots, m)$. Then Mahler was able to prove, by means of an explicit construction, that in this case there is a unique solution up to multiplication by a scalar and, normalising the solution so that the polynomials have rational integral coefficients without a common divisor, the coefficients of the polynomials are bounded by $c(n,s)^m$ for a suitable constant $c(n,s)$. By using this information, Mahler was able to give new proofs of earlier theorems of Thue and Siegel about approximations of n-th roots by rationals.

An even more striking application of Padé approximation methods, again using the algebraic function $u(x) = (1-x)^{1/n}$, was given in A. Baker's paper [1] on rational approximations to $\sqrt[3]{2}$ and other numbers. There for the first time one obtained effective non-trivial lower bounds for best approximations by rationals to certain non-quadratic algebraic numbers.

It is a natural question to ask to what extent these ideas can be applied to algebraic functions other than $(1-x)^{\omega_j}$. Unfortunately, a key feature for approximation methods to succeed in arithmetical applications is to have an exponential bound c^m for the height of the coefficients of the approximating polynomials. Obtaining such a bound has proven to be quite elusive except in very special situations. In the general case with equal weights, Siegel's lemma gives us only a bound c^{m^2}.

In an important paper on Padé approximations D. V. Chudnovsky and G. V. Chudnovsky [6], Section 2, p. 92 and Section 10, Remark 10.3, p. 147, constructed Padé approximations in closed form for the vector $(1, y)$ with $y^2 =$

$4x^3 - g_2x - g_3$ in a neighborhood of $x = x_0$, and showed that the height of coefficients for Padé approximating polynomials grows as c^{m^2} unless the point (x_0, y_0) on the elliptic curve is a torsion point, in which case it has exponential growth c^m. They also noticed that in the more general case of a function $y^2 = p(x)$ the problem is related to the question, considered for the first time by Abel, of the periodicity of the continued fraction expansion of $\sqrt{p(x)}$.

A way out of the difficulty in controlling the height of Padé approximations is to weaken the requirement of having a zero of highest possible order at the origin, by asking instead (in the case of equal weights) for a zero of order $(s + 1 - \delta)(m + 1)$ with $\delta > 0$. In this case one talks about (m, δ)-Padé type I (or Latin) approximations, or briefly (m, δ)-Padé approximations. A standard application of Siegel's lemma now can be used to show that in every case there are (m, δ)-Padé approximations for the vector $(u_0(x), \ldots, u_s(x))$, with height bounded by $c^{m/\delta}$ for some constant c. This restored exponential bound suffices for several interesting applications but the quality of results so obtained always suffers because one needs to take δ very small, with a corresponding worsening of the height.

In this paper, we study (m, δ)-Padé approximations for $(1, u(x), \ldots, u(x)^{n-1})$, for a general algebraic function $y = u(x)$ of degree n satisfying the simple conditions stated before.

Our main result shows that if the curve $f(x, y) = 0$ has positive genus then the order of growth of the height of (m, δ)-Padé approximations for the set of functions $(1, u(x), \ldots, u(x)^{n-1})$ is not less than $c^{m/\delta}$ if the rational equivalence class of the divisor $nQ - (Q_1 + Q_2 + \ldots + Q_n)$ is not a torsion point of the Jacobian of the curve C (here $c > 1$ and Q is the point on C determining the algebraic function $u(x)$). If $\delta = 1/(m + 1)$ we get a lower bound c^{m^2} for the height of classical Padé approximations.

Although the method of proof owes a lot to the ideas in the papers [2], [3] and [7], we have chosen to make this paper essentially self-contained.

One may ask to what extent this result is optimal. In the Appendix, it is shown that if the divisor $nQ - (Q_1 + Q_2 + \ldots + Q_n)$ is a torsion point on the Jacobian of the curve C then $(1, y, \ldots, y^{n-1})$ admits (m, δ)-Padé approximations with δ as small as $O(1/m)$ and height growing only at exponential rate c^m.

It would be of definite interest to obtain results of this type for the case in which $s < n - 1$ and also for the case in which C has genus 0. We consider our results as a first step in this direction.

Acknowledgement. The first author thanks the Eidgenössische Technische Hochschule of Zürich for its hospitality and support during the preparation of this paper. The second and third authors were supported in part by the Institute for Advanced Study in Princeton.

2. – Padé approximations on algebraic curves

We recall the standard notion of height to which we are referring. Let M_K be the set of places of K. For $v \in M_K$, we denote by $d_v = [K_v : \mathbb{Q}_v]$ the corresponding local degree. We normalise the absolute values $|\ |_v$ of K by requiring that

$$|x|_v = \|x\|_v^{d_v/d}, \qquad x \in K_v$$

where $\|\ \|_v$ is the unique extension to K_v of the usual p-adic or archimedean valuation on $\mathbb{Q}_v$. We have the product formula

$$\prod_{v \in M_K} |x|_v = 1, \qquad x \in K, \quad x \neq 0.$$

The (absolute) Height of $x \in K$ is defined to be

$$H(x) = \prod_{v \in M_K} \max(1, |x|_v),$$

and the (absolute) height of $x \in K$ is given by the logarithm of the Height,

$$h(x) = \sum_{v \in M_K} \log^+ |x|_v,$$

where $\log^+ a = \log \max(1, a)$ for $a \geq 0$. These definitions do not depend on the field K containing x. For a vector $\mathbf{x} = (x_1, \ldots, x_m)$ in K^m and a place $v \in M_K$, we define

$$|\mathbf{x}|_v = \max(|x_1|_v, \ldots, |x_m|_v)$$

and

$$H(\mathbf{x}) = \prod_{v \in M_K} \max(1, |\mathbf{x}|_v).$$

This Height definition may be further extended to polynomials by taking the Height of the vector of coefficients, the corresponding height being obtained from the Height by taking the logarithm.

We restate the notion of (m, δ)-Padé approximations given in the introduction in a more geometrical way as follows. Let C, x, y, Q and $\pi : C \longrightarrow \mathbb{P}^2$ be as in the preceding section. Choose $s = n - 1$ and $u_j(x) = u(x)^j$. Then an (m, δ)-Padé approximation for $(1, u(x), \ldots, u(x)^{n-1})$ is a vector $(P_0(x), \ldots, P_{n-1}(x))$ of polynomials in $K[x]$, of degree at most m, for which the rational function

$$F = P(x, y) = \sum_{j=0}^{n-1} P_j(x) y^j$$

has a zero of order at least $[(n - \delta)(m + 1)]$ at the point Q.

We are interested in the behaviour of $h(P)$ for large m and small δ. Indeed, the main result of this paper is the following.

THEOREM 1. *Let C be a non-singular irreducible projective curve of genus $g \geq 1$ defined over a number field K and let x be an element of degree $n \geq 1$ of the function field $K(C)$. Suppose that x is unramified over $0 \in \mathbb{P}_1$, let Q be a point $Q \in x^{-1}(0)$ and suppose that K is so large that $Q \in K(C)$.*

Let $y \in K(C)$ be a rational function on C such that $K(C) = K(x, y)$ and which is a local uniformising parameter at Q.

Let δ be such that $n \geq \delta \geq 1/(m+1)$ and let F be a non-zero element of $K(C)$ of the form

$$F = P(x, y) = \sum_{j=0}^{n-1} P_j(x)y^j$$

with $P_j(x) \in K[x]$ polynomials of degree at most m, with a zero at Q of order

$$\mathrm{ord}_Q(F) \geq [(n-\delta)(m+1)].$$

Suppose that the linear equivalence class q^* of $x^{-1}(0) - nQ$ in the Jacobian of C is not a torsion point. Then there are two effectively computable positive constants c_1, c_2, depending only on K, C, x and y, such that

$$h(P) \geq \frac{c_1}{\delta}(m+1) - c_2 m$$

for all sufficiently large m.

REMARKS. We suppose $\pi(Q) = (0, 0)$ in the theorem only for notational convenience, since a translation of x and y affects the height of $P(x, y)$ by a quantity $O(m)$, which is independent of δ.

The two conditions that x is unramified over 0 and y is a uniformiser at Q can be dispensed with to some extent, but our proof will then require substantial modifications in places.

The dependence on δ in the lower bound for $h(P)$ provided by our theorem is sharp. Also, taking $\delta = 1/(m+1)$, we obtain a new proof and generalisation of the result of [5] stated in the introduction.

If $g \geq 2$ one can show that, except for finitely many possibilities for q^*, the constant c_1 admits a positive lower bound which depends only on the curve C and the degree n of the rational function x, and moreover the number of possible exceptions is bounded solely in terms of C and n. To see this, we note that the proof of Theorem 1 gives a constant c_1 of order $|\, q^* \,|^2$, where $|\ |^2$ is the canonical Néron-Tate height on J. The locus of points $q^* = \mathrm{cl}(x^{-1}(x(Q)) - nQ)$ for $Q \in C$ is a curve Γ in J whose degree with respect to a fixed polarisation of J is bounded as a function of C and n alone. Hence, except for a finite set of points q^* of cardinality bounded by a function of C and n, the height $|\, q^* \,|^2$ admits a positive lower bound in terms of C and n alone. This follows from a uniform version of Bogomolov's conjecture, which can be obtained combining the results of L. Szpiro, E. Ullmo and S. Zhang [12] together with a determinantal argument of Bombieri and Zannier [5].

3. – The upper bound

We recall the notion of Height of matrices as given in [4], p. 15. This is simply the Height of the vector of Plücker coordinates of the matrix, namely the vector of determinants of minors of maximal rank of the matrix. More explicitly, let X be an $M \times N$ matrix with coefficients in K and with $\operatorname{rank}(X) = R \leq M < N$. If $J \subset \{1, 2, \dots, N\}$ is a subset with $|J| = R$ elements we write

$$X_J = (x_{ij}), \quad i = 1, \dots, R, \quad j \in J$$

for the corresponding sub-matrix. For each place v of K, we define the local Height as follows:

$$H_v(X) = \begin{cases} \max_{|J|=R} |\det X_J|_v & \text{if } v \nmid \infty, \\ \left(\sum_{|J|=R} \|\det X_J\|_v^2 \right)^{d_v/2d} & \text{if } v \mid \infty. \end{cases}$$

Taking the product over all places, we obtain the global Height

$$H(X) = \prod_v H_v(X)$$

and the corresponding height $h(X)$ by taking the logarithm of $H(X)$.

We shall work with the following version of Siegel's lemma as given in [4], Corollary 11, p. 28.

SIEGEL'S LEMMA. *Let $M < N$ be positive integers, let $\mathcal{O}_K$ be the ring of integers and D_K the discriminant of K. Let $A = (a_{ij})$ be an $M \times N$ matrix over K of maximal rank* $\operatorname{rank}(A) = M$. *Then there exist $N - M$ linearly independent vectors $\mathbf{x}_1, \dots, \mathbf{x}_{N-M}$ in $\mathcal{O}_K^N$ which satisfy*

$$A\mathbf{x}_l = 0, \qquad l = 1, \dots, N - M$$

and

$$\prod_{l=1}^{N-M} H(\mathbf{x}_l) \leq |D_K|^{(N-M)/(2d)} H(A).$$

In particular, there is a non-trivial solution $\mathbf{x} \in \mathcal{O}_K^N$ of $A\mathbf{x} = 0$ satisfying

$$H(\mathbf{x}) \leq |D_K|^{1/(2d)} H(A)^{1/(N-M)}.$$

We also need an estimate for the coefficients of the McLaurin expansion of an uniformiser. Results of this type go back to Eisenstein. We have

LOCAL EISENSTEIN THEOREM. *Let $u(x) = a_1x + a_2x^2 + \dots \in K[[x]]$ be a formal power series solution of $f(x, y) = 0$ where $f(x, y) \in K[x, y]$ is a polynomial such that $f_y(0, 0) \neq 0$.*

Then, for every $v \in M_K$ the power series $u(x)$ has a positive radius of convergence r_v. Let

$$\rho_v = \frac{|f_y(0,0)|_v}{|\, f \,|_v} .$$

Then $\rho_v \leq 1$ and for $k = 2, 3, \dots$ we have the explicit bound

$$|\, a_k \,|_v \leq c(v)^k \rho_v^{-(2k-1)} ,$$

where $c(v) = 1$ if v is a finite place and $c(v) = |\, \frac{1}{2} \deg(f)^7 \,|_v$ otherwise. If $k = 1$ the same bound holds if v is a finite place, and it holds with an extra factor of $|\, 2 \,|_v$ if v is an infinite place.

In particular, we have $r_v \geq c(v)^{-1} \rho_v^2$ for every place v.

COROLLARY. *Let $u(x)$ be as before and $j \geq 1$. Then the coefficients of $u(x)^j = \sum_k a_{jk} x^k$ satisfy*

$$|\, a_{jk} \,|_v \leq c'(v)^k \rho_v^{-(2k-1)} ,$$

where $c'(v) = 1$ if v is a finite place and $c'(v) = |\, \deg(f)^7 \,|_v$ otherwise.

PROOF OF COROLLARY. We may assume $j \geq 2$. The coefficient of x^k in $u(x)^j$ is $\sum a_{\nu_1} \cdots a_{\nu_j}$ where the sum is over $\nu_1 + \ldots + \nu_j = k$ with $\nu_l \geq 1$. The number of such j-tuples is $\binom{k}{j} \leq 2^k$ and the result follows because $k \geq 2$.

PROOF OF THE LOCAL EISENSTEIN THEOREM. We write partial derivatives with respect to x and y by means of subscripts. If v is finite, we write $f(x, y) = \sum_{ij} b_{ij} x^i y^j$ and substitute the power series $u(x)$ for y. Since $f(x, u(x)) = 0$, equating to 0 the coefficient of x^k in $f(x, u(x))$ we get

$$\sum_{ij} b_{ij} \sum_{i+\nu_1+\ldots+\nu_j=k} a_{\nu_1} \cdots a_{\nu_j} = 0 ,$$

where as usual the empty products for $j = 0$ are meant to be 1.

The contribution of the terms with $\nu_l = k$ to this equation is

$$b_{01} a_k = f_y(0,0)\, a_k .$$

Since v is ultrametric, it follows that

$$|\, b_{01} \,|_v \,|\, a_k \,|_v \leq \left(\max_{ij} |\, b_{ij} \,|_v \right) \max{}' \,|\, a_{\nu_1} \cdots a_{\nu_j} \,|_v$$

where max$'$ runs over $(i, \nu_1, \dots, \nu_j)$ with $i + \nu_1 + \ldots + \nu_j = k$ and $0 < \nu_l < k$, because of our assumption that $a_0 = 0$. This implies that in max$'$ we have either $j \geq 2$ or $\nu_1 + \ldots + \nu_j \leq k - 1$. Hence if $C \geq 1$ is such that $|\, a_l \,|_v \leq C^{2l-1}$ for $l = 1, 2, \dots, k-1$, we obtain

$$|\, b_{01} \,|_v \,|\, a_k \,|_v \leq \left(\max_{ij} |\, b_{ij} \,|_v \right) C^{2k-2} .$$

The result follows by induction, taking

$$C = \frac{\max |\, b_{ij} \,|_v}{|\, b_{01} \,|_v} = \frac{1}{\rho_v}.$$

If instead v is infinite, we argue as follows. Let $|\ |$ denote the usual Euclidean absolute value and for a polynomial f let $\|f\|$ be the Gauss norm, namely the maximum of the coefficients of f.

Let us abbreviate $u^{(l)}(x) = \left(\frac{d}{dx}\right)^l u(x)$. By induction on l we establish that there is a polynomial $f_l(x, y)$ such that

$$u^{(l)}(x) = -\frac{f_l(x, u(x))}{f_y(x, u(x))^{2l-1}}$$

for $l = 1, 2, \ldots$. We have $f_1 = f_x$ and

$$f_{l+1} = (f_l)_x f_y^2 - (f_l)_y f_x f_y + (2l-1) f_l (f_{yy} f_x - f_{xy} f_y).$$

From this equation, we see that if $d = \deg f$ then f_l has degree at most $(2l-1)(d-1)$, and we can also estimate $\|f_{l+1}\|$ by

$$\|f_{l+1}\| \le \frac{1}{2} l d^7 \|f\|^2 \|f_l\|.$$

This can be seen as follows. Let $f_l = \sum b_{uv}^{(l)} x^u y^v$. Then the above formula yields

$$b_{hk}^{(l+1)} = \sum_{\substack{u+p+r=h+1 \\ v+q+s=k+2}} b_{uv}^{(l)} b_{pq} b_{rs} \big(uqs - vps + (2l-1)(p(s-1)s - pqs)\big).$$

We have

$$|\, uqs - vps + (2l-1)(p(s-1)s - pqs) \,| \le (2l-1)(d-1)(p+q)s,$$

and summing over r with $r+s \le d$ gives us $(2l-1)(d-1)(p+q)s(d-s+1)$. Summing over s gives $(2l-1)(d-1)d(d+1)(d+2)(p+q)/6$, and the sum over p and q with $p+q \le d$ gives $(2l-1)(d-1)d^2(d+1)^2(d+2)^2/18 \le ld^7/2$.

Since $\|f_1\| \le d\,\|f\|$, we obtain by induction

$$\|f_{l+1}\| \le l!\, d^{7l+1} 2^{-l} \|f\|^{2l+1}.$$

The required estimate for a_k follows from

$$a_k = -\frac{1}{k!} \frac{f_k(0,0)}{f_y(0,0)^{2k-1}}.$$

We apply Siegel's lemma to obtain a Padé approximation of the algebraic function u, normalised so that $u(0) = 0$. The result is as follows. Let $Q \in x^{-1}(0)$ be the point for which $\pi(Q) = (0,0)$ and suppose that K is so large that $Q \in C(K)$.

THEOREM 2. *Let m be a positive integer and let δ be a real number with $n > \delta \geq 1/(m+1)$. Then we can find polynomials $P_0(X), \ldots, P_{n-1}(X)$ in $K[X]$, not all zero and of degree at most m, such that the rational function on C given by*

$$F = P(x, y) = \sum_{j=0}^{n-1} P_j(x) y^j$$

has a zero at Q of order

$$\operatorname{ord}_Q(F) \geq [(n-\delta)(m+1)]$$

and moreover

$$h(P) \leq \frac{(n-\delta)^2}{\delta}(m+1)\{7 \log \deg(f) + 2h(f)\} + \frac{n-\delta}{2\delta} \log\big(n(m+1)\big) + O(1)$$

as $m \to \infty$. The implied constant in $O(\ )$ is bounded independently of δ.

PROOF. Writing the polynomials $P_j(x)$ as

$$P_j(x) = \sum_{l=0}^{m} p_{jl} x^l, \qquad j = 0, \ldots, n-1,$$

and setting $u(x)^j = \sum_k a_{jk} x^k$, the requirement $\operatorname{ord}_Q(F) \geq [(n-\delta)(m+1)]$ reduces to solving the linear system

$$\sum_{j=0}^{n-1} \sum_{l=0}^{\min(m,k)} p_{jl} a_{j,k-l} = 0, \qquad k = 0, \ldots, M-1$$

where $M = [(n-\delta)(m+1)]$. Let A_0 be the associated matrix

$$A_0 = (a_{j,k-l})$$

where the columns are indexed by $(j, l) \in [0, n-1] \times [0, m]$ and the rows by $k = 0, \ldots, M-1$. Let $N = n(m+1)$, let R be the rank of A_0 and A be an $R \times N$ sub-matrix of A_0 of rank R, obtained by eliminating if needed some rows of A_0. As $\delta(m+1) \geq 1$ we must have $R < N$.

One then applies Siegel's lemma to A, so that we know there is a solution $F = P(x, y)$ with

$$h(P) \leq \frac{1}{N-R} \log H(A) + O(1).$$

It remains to estimate $h(A) = \log H(A)$. For this, we recall the general inequality

$$H_v(A) \leq H_v(B) H_v(C)$$

valid for any matrix A written in block form as $A = \binom{B}{C}$, see for instance [4], (2.6), p. 15. In particular, in our case we have

$$H_v(A) \leq \prod H_v(A_i)$$

where A_i is the i-th row of A and the product runs over all rows of A. Thus, denoting by $\prod'$ a product over the R rows of A_0 occurring in A, we have

$$\begin{aligned} H_v(A) &\leq \prod_k{}' \left(\max(1, \mid N \mid_v^{1/2}) \max_{jl} \mid a_{j,k-l} \mid_v \right) \\ &\leq \left\{ \max(1, \mid N \mid_v^{1/2}) \max_{jk} \mid a_{jk} \mid_v \right\}^R . \end{aligned}$$

By the corollary to the Local Eisenstein Theorem we deduce, taking the product over all $v \in M_K$, that

$$H(A) \leq \left\{ \sqrt{N} \prod_v c'(v)^M (1/\rho_v)^{2M-1} \right\}^R .$$

Now we take the logarithm, obtaining

$$\begin{aligned} h(A) &\leq \frac{1}{2} R \log N + MR \sum_v \log c'(v) + (2M-1) R \sum_v \log \frac{1}{\rho_v} \\ &= \frac{1}{2} R \log N + MR \log(\deg(f)^7) \\ &\quad + (2M-1) R \left\{ \sum_v \log \mid f \mid_v - \sum_v \log \mid f_y(0,0) \mid_v \right\} \\ &= \frac{1}{2} R \log N + 7MR \log \deg(f) + (2M-1) R\, h(f). \end{aligned}$$

Theorem 2 follows from this estimate.

4. – The lower bound

We adopt the notations of the preceding section. Let L be a finite extension of K, let $R \in C(L)$ and write for simplicity $\xi = x(R)$, $\eta = y(R)$.

For the rest of this section, we shall suppose that

a) *We have* $F(R) = P(\xi, \eta) \neq 0$.

We are going to obtain a lower bound for $h(P)$ using the following idea. We apply the product formula to $F(R)$ and evaluate, for $w \in M_L$, the various quantities $\log \mid F(R) \mid_w$ in two ways.

One begins by showing that if $|\eta|_w$ is sufficiently small then $\eta = u(\xi)$. Since F vanishes to high order at $Q = \pi^{-1}(0,0)$, this shows that $|F(R)|_w$ must be very small, thus contributing a large negative term to the product formula.

If instead $|\eta|_w$ is not small then one estimates $|F(R)|_w$ trivially.

In the end, after applying the product formula one obtains a lower bound for $h(P)$.

The next lemma, already used in the paper [8] of P. Debes, identifies a set of places $w \in M_L$ such that $\eta = u(\xi)$ for ξ in a sufficiently small neighborhood of 0 in L_w.

LEMMA 1. *For $w \in M_L$, let $c''(w) = 1$ if w is a finite place and $c''(w) = |\deg(f)|_w^3$ otherwise. Then the following holds.*

a) *If $z \in L_w$ is such that $|z|_w < c'(w)^{-2}\rho_w^2$, we have $|u(z)|_w < c''(w)^{-1}\rho_w$.*

b) *Suppose that*

$$|\xi|_w < c'(w)^{-2}\rho_w^2, \qquad |\eta|_w < c''(w)^{-1}\rho_w .$$

Then $\eta = u(\xi)$.

PROOF. The lemma is trivial if $\deg(f) = 1$, hence we shall suppose $\deg(f) \geq 2$.

Statement a) follows immediately from the Local Eisenstein Theorem.

To prove b), it suffices to prove that the equation $f(\xi, y) = 0$ has at most one root in the disk $|y|_w < c''(w)^{-1}\rho_w$. In fact, by a) we have $|u(\xi)|_w < c''(w)^{-1}\rho_w$, which implies $u(\xi) = \eta$ by this uniqueness statement.

Hence suppose η and η' are two distinct roots such that

$$\max(|\eta|_w, |\eta'|_w) < c''(w)^{-1}\rho_w,$$

and note that *a fortiori* the same bound holds for $|\xi|_w$.

Writing $f(x,y) = \sum b_{ij}x^iy^j$, we have

$$\begin{aligned} 0 &= \frac{f(\xi,\eta) - f(\xi,\eta')}{\eta - \eta'} \\ &= b_{01} + {\sum}' b_{ij}\xi^i \left(\eta^{j-1} + \eta^{j-2}\eta' + \ldots + (\eta')^{j-1}\right) \end{aligned}$$

where $\sum'$ is over (i,j) with $j \geq 1$ and $i + j - 1 \geq 1$. Note also that $b_{01} = f_y(0,0)$.

If w is a finite place this gives

$$1 \leq \rho_w^{-1} \max{}' \left(|\xi|_w^i |\eta|_w^{j-1-l} |\eta'|_w^l\right)$$

where $\max'$ runs over (i,j) as above and over $0 \leq l \leq j-1$. This contradicts $\max(|\xi|_w, |\eta|_w, |\eta'|_w) < \rho_w$ and proves our assertion in this case.

If instead w is an infinite place a similar estimate gives

$$1 \leq | D |_w \rho_w^{-1} \max{}' \left(| \xi |_w^i | \eta |_w^{j-1-l} | \eta' |_w^l \right)$$

where

$$D = \sum_{ij}{}' j \leq \sum_{j=1}^{\deg(f)} j(\deg(f) - j + 1) \leq \deg(f)^3 .$$

Again, this leads to a contradiction if $\max(| \xi |_w, | \eta |_w, | \eta' |_w) <| D |_w^{-1} \rho_w$.

This completes the proof of the lemma.

We now apply the product formula to the non-zero algebraic number $F(R)$, obtaining

$$\sum_{w \in L} \log | F(R) |_w = 0 ,$$

and estimate $\log | F(R) |_w$ in two different ways according to whether or not Lemma 1 is applicable to the point (ξ, η).

LEMMA 2. *We have*

$$h(P) \geq [(n - \delta)(m + 1)] \sum_{w \in M_L} \min \left\{ \log^+ \frac{1}{| \xi |_w}, \log^+ \frac{1}{| \eta |_w} \right\}$$
$$- (m + 1 + n \deg(f))h(\xi) - \big(14 \log \deg(f) + 2h(f)\big) n(m + 1)$$
$$- \log(n(m + 1)) - 2n \deg(f) - nh(f) .$$

PROOF. For $w \in M_L$ we denote by $\mathbb{C}_w$ a completion of an algebraic closure of L_w, with an absolute value extending the absolute value $| \ |_w$ in L_w.

We define $\sigma_w = c'(w)^{-2} \rho_w^2$ and note that $\sigma_w \leq c''(w)^{-1} \rho_w \leq 1$.

Consider now the set S of places of M_L for which

$$| \xi |_w < \sigma_w, \qquad | \eta |_w < \sigma_w .$$

By Lemma 1, we have $\eta = u(\xi)$ and the McLaurin series for $g(z) = P(z, u(z))$ defines an analytic function of z in the open disk

$$| z |_w < \sigma_w, \qquad z \in \mathbb{C}_w .$$

By construction, the function $g(z)$ has a zero of order at least $M = [(n - \delta)(m + 1)]$ at the origin. Therefore by Schwarz's lemma we get

$$| F(R) |_w = | g(\xi) |_w \leq \left(\frac{| \xi |_w}{\sigma_w} \right)^M \sup_{|z|_w < \sigma_w} | g(z) |_w .$$

We estimate $g(z)$ in $|z|_w < \sigma_w \le 1$ using Lemma 1, which gives $|u(z)|_w < c''(w)^{-1}\rho_w \le 1$. Hence

$$|g(z)|_w = |P(z, u(z))|_w \le |P|_w \max(1, |n(m+1)|_w)$$

because the polynomial $P(x, y)$ has at most $n(m+1)$ monomials. Combining the last two displayed estimates we deduce

$$\log |F(R)|_w \le -M \log\left(\frac{\sigma_w}{|\xi|_w}\right) + \log |P|_w + \log^+ |n(m+1)|_w$$

for $w \in S$.

If instead $w \notin S$ we have trivially from $F(R) = P(\xi, \eta)$ the bound

$$\log |F(R)|_w \le \log |P|_w + \log^+ |n(m+1)|_w + (m+1)\log^+ |\xi|_w + n \log^+ |\eta|_w \,.$$

Now these two estimates and the product formula give

$$0 \le h(P) + \log(n(m+1)) - M \sum_{w \in S} \log\left(\frac{\sigma_w}{|\xi|_w}\right) + (m+1)h(\xi) + n\,h(\eta)\,.$$

A lower bound for the sum is obtained as follows. We have

$$\begin{aligned}\sum_{w\in S} \log\left(\frac{\sigma_w}{|\xi|_w}\right) &\ge \sum_{w\in S} \min\left\{\log\left(\frac{\sigma_w}{|\xi|_w}\right), \log\left(\frac{\sigma_w}{|\eta|_w}\right)\right\} \\ &= \sum_{w\in M_L} \min\left\{\log^+\left(\frac{\sigma_w}{|\xi|_w}\right), \log^+\left(\frac{\sigma_w}{|\eta|_w}\right)\right\} \\ &\ge \sum_{w} \min\left\{\log^+ \frac{1}{|\xi|_w}, \log^+ \frac{1}{|\eta|_w}\right\} + \sum_{w} \log \sigma_w \\ &\ge \sum_{w} \min\left\{\log^+ \frac{1}{|\xi|_w}, \log^+ \frac{1}{|\eta|_w}\right\} - 14 \log \deg(f) - 2h(f)\,.\end{aligned}$$

This gives

$$\begin{aligned}h(P) \ge M \sum_{w\in M_L} \min\left\{\log^+ \frac{1}{|\xi|_w}, \log^+ \frac{1}{|\eta|_w}\right\} \\ - (m+1)h(\xi) - (14\log\deg(f) + 2h(f))M - \log(n(m+1)) - n\,h(\eta)\,.\end{aligned}$$

Finally we have the easy estimate

$$h(\eta) \le \deg(f)\,h(\xi) + 2\deg(f) + h(f)$$

which follows from $f(\xi, \eta) = 0$. This completes the proof.

The following result identifies the sum appearing in Lemma 2 with a Weil height.

LEMMA 3. *For $R \neq Q$ and not a pole of both x and y, the sum*

$$h_Q(R) = \sum_w \min\left\{\log^+ \frac{1}{|\, x(R)\,|_w},\ \log^+ \frac{1}{|\, y(R)\,|_w}\right\}$$

is a Weil height on C relative to the divisor Q.

PROOF. We recall the notion of a Weil function and Weil height on a projective variety. Let X be a projective variety over a number field K and let D be a Cartier divisor on X with associated sheaf $\mathcal{O}(D)$. Let $\mathcal{L}$, $\mathcal{M}$ be base-point-free line sheaves on X such that $\mathcal{O}(D) \cong \mathcal{L} \otimes \mathcal{M}^{-1}$.

Let σ_D be a rational section of $\mathcal{O}(D)$ with divisor D and let also $s_0, s_1, \dots, s_l$ be sections of $\mathcal{L}$ without common zeros, and similarly for sections $t_0, t_1, \dots, t_m$ of $\mathcal{M}$. We shall refer to these data as a presentation $\mathcal{D}$ of the Cartier divisor D.

Let L be a finite extension of K which is a field of definition for the presentation $\mathcal{D}$ and let $P \in X(L)$. Then for $w \in M_L$ one defines a local height by

$$\lambda_{\mathcal{D}}(P, w) = \min_i \max_j \log \left| \frac{s_j}{\sigma_D t_i}(P) \right|_w .$$

The sum

$$h_{\mathcal{D}}(P) = \sum_w \lambda_{\mathcal{D}}(P, w)$$

is the Weil height of P associated to the presentation $\mathcal{D}$. This height is independent of the field of definition for $\mathcal{D}$ and P. If $\mathcal{D}'$ is another presentation of D, the quantity $h_{\mathcal{D}}(\ ) - h_{\mathcal{D}'}(\ )$ is uniformly bounded on $X(\overline{K})$.

Let Z_0, Z', Z'' be the divisors on C given by

$$x^{-1}(\infty) = Z_0 + Z', \qquad y^{-1}(\infty) = Z_0 + Z'', \qquad Z = Z_0 + Z' + Z'',$$

with Z' and Z'' without common points. Then we define

$$X = x^{-1}(0) - Q + Z'', \qquad Y = y^{-1}(0) - Q + Z'.$$

The divisors $Q + X$, $Q + Y$ and Z are linearly equivalent because $Q + X - Z$ and $Q + Y - Z$ are the divisors of zeros and poles of the rational functions x and y.

Now take $\mathcal{L} = \mathcal{O}(Z)$, $\mathcal{M} = \mathcal{O}(X)$ and let σ_D a section of $\mathcal{O}(Q)$ with divisor Q and s_0 a section of $\mathcal{L}$ with divisor Z. Then there are sections s_1 and s_2 of $\mathcal{L}$ such that

$$x = s_1/s_0, \qquad y = s_2/s_0$$

and the equations $s_1 = \sigma_D t_0$ and $s_2 = \sigma_D t_1$ define two regular sections t_0 and t_1 of $\mathcal{M}$. Since X and Y have no common points and Q is not in Z, we have obtained a presentation $\mathcal{Q}$ of Q.

With this presentation the functions $s_i/\sigma_D t_0$ are $1/x$, 1, y/x, and the functions $s_i/\sigma_D t_1$ are $1/y$, x/y, 1. A simple calculation now shows that, outside Q and the divisor Z_0, we have

$$\lambda_Q(R, w) = \min\left\{\log^+ \frac{1}{|\, x(R)\,|_w}, \log^+ \frac{1}{|\, y(R)\,|_w}\right\}.$$

This completes the proof.

LEMMA 4. *Let $x^{-1}(0) = Q_1 + \ldots + Q_n$ and let $h_{Q_j}(\ )$ denote a choice of a Weil height for Q_j. Then for any point $R \in C(\overline{K})$ with $F(R) \neq 0$ and not a pole of both x and y we have*

$$\frac{1}{m+1} h(P) \geq (n - c_5\delta)\, h_Q(R) - \sum_{j=1}^{n} h_{Q_j}(R) - O(1).$$

The constant c_5 and the constant implicit in the symbol $O(\ )$ are bounded independently of m, R and δ

PROOF. We apply Lemma 2, Lemma 3 and the formula

$$h(x(R)) = \sum_{j=1}^{n} h_{Q_j}(R) + O(1),$$

obtained by functoriality of heights applied to the morphism $x : C \longrightarrow \mathbb{P}^1$. The lemma follows noting that $\delta \geq 1/(m+1)$ and $h_{Q_i}(R) = O(h_Q(R)) + O(1)$.

5. – Proof of Theorem 1

We now reinterpret the result of Lemma 4 on the Jacobian J of C. For the relevant background on heights and Jacobians, see [10] and [11].

Let D_0 be a divisor on C of degree 1 such that $(2g-2)D_0$ is linearly equivalent to a canonical divisor k_C, which we can do because J is a divisible group. By extending the base field K, we may and shall assume that D_0 is defined over K.

We have an embedding $j : C \longrightarrow J$ given by $j(P) = \mathrm{cl}(P - D_0)$. The map j is then extended to arbitrary divisors D on C by $j(\sum a_i P_i) = \sum a_i j(P_i)$.

The divisor

$$\Theta = \underbrace{j(C) + \ldots + j(C)}_{g-1 \text{ times}}$$

is called the theta divisor associated to this embedding. The particular choice of D_0 ensures that $\Theta^- = [-1]^*\Theta = \Theta$, so that $\theta = \mathrm{cl}(\Theta)$ is even.

There is a surjective morphism

$$\pi : C^g \to J$$

given by

$$\pi(R_1, \dots, R_g) = j(R_1) + \dots + j(R_g)\,.$$

Over a Zariski dense open subset U of J this morphism is $g!$ to 1. The inverse image of a point

$$j(R_1) + \dots + j(R_g) \in U$$

consists of the point $(R_1, R_2, \dots, R_g)$ and those obtained from it by permutation of the coordinates, and the associated divisor $(R_1) + (R_2) + \dots + (R_g)$ on C is then called non-special.

For $\mathcal{L} \in \mathrm{Pic}(J)$, let $\widehat{h}_{\mathcal{L}}$ be the associated Néron-Tate height. The even ample class θ gives rise to a quadratic form

$$\mid a \mid^2 = 2\widehat{h}_\theta(a)$$

on $J(\overline{K})$ and to an associated bilinear form

$$\langle a, b\rangle = \widehat{h}_\theta(a+b) - \widehat{h}_\theta(a) - \widehat{h}_\theta(b)\,.$$

In what follows, the implied constants in the $O(1)$ notation will depend on the points $Q_1, \dots, Q_n$ but will be uniform in the varying points $R, R_1, \dots, R_g$. All points considered will be in $C(\overline{K})$.

For D a divisor of degree g on C, let $\varphi_{-j(D)} : C \to J$ be the composition of j with translation by $-j(D)$ on J. Then the divisor class $\varphi^*_{-j(D)}(\theta^-) = \varphi^*_{-j(D)}(\theta)$ equals that of $\mathcal{O}(D)$ on C. Setting $D = gQ$ this implies, by a simple computation, that for $R \in C$, $R \neq Q$, we have

$$\begin{aligned} h_Q(R) &= \frac{1}{g}\,\widehat{h}_{\theta^-}(j(R) - gj(Q)) + O(1) = \frac{1}{g}\widehat{h}_\theta(j(R) - gj(Q)) + O(1) \\ &= \frac{1}{2g}\left|j(R) - gj(Q)\right|^2 + O(1) = \frac{1}{2g}\,\mid j(R)\mid^2 - \langle j(Q), j(R)\rangle + O(1)\,. \end{aligned}$$

This gives

$$\begin{aligned} n\,h_Q(R) - \sum_{i=1}^{n} h_{Q_i}(R) &= \frac{n}{2g}\,\mid j(R)\mid^2 - \langle nj(Q), j(R)\rangle + O(1) \\ &\quad - \frac{n}{2g}\,\mid j(R)\mid^2 + \sum_{i=1}^{n}\langle j(Q_i), j(R)\rangle + O(1) \\ &= \langle q^*, j(R)\rangle + O(1) \end{aligned}$$

where q^* is the point $\sum_{i=1}^n j(Q_i) - nj(Q) \in J$. If we combine this equation with Lemma 4 we find

$$\frac{1}{m+1} h(P) \geq \langle q^*, j(R)\rangle - c_5\, \delta\, h_Q(R) - O(1)\,.$$

This inequality holds for every $R \in C(\overline{K})$ such that $F(R) \neq 0$, $R \neq Q$ and R not a pole of both x and y.

Let $W \subset J$ be the locus of special divisors and let $T \in J \setminus W$. Then we can write, uniquely up to a permutation,

$$T = j(R_1) + j(R_2) + \ldots + j(R_g), \qquad R_i \in C\,.$$

Now, we factor the morphism $C^g \xrightarrow{\pi} J$ as $C^g \xrightarrow{\phi} C^{(g)} \xrightarrow{\psi} J$, where $C^{(g)}$ is the g-fold symmetric product of C. The morphism ψ is birational and an isomorphism outside the inverse image $\widetilde{W} = \psi^{-1}(W)$ of the special locus W. Let $\mathcal{M}$, $\mathcal{N}$ be very ample line sheaves on $C^{(g)}$, J and let $s_0, \ldots, s_M$, $t_0, \ldots, t_N$ be bases of sections of $\mathcal{M}$, $\mathcal{N}$, giving projective coordinates on $C^{(g)}$ and J. Since ψ^{-1} is an isomorphism outside W, there are homogeneous polynomials $G_i(t_0, \ldots, t_N)$, $i = 0, \ldots, M$, all of the same degree, not all zero at $(t_0(T), \ldots, t_N(T))$ if $T \notin W$, which describe ψ^{-1} outside W.

More explicitly, let $(R_1, \ldots, R_g) \in C^g$ and write for simplicity $S = \phi(R_1, \ldots, R_g)$, $T = \psi(S)$. Then we have

$$(s_0(S) : \cdots : s_M(S)) = (G_0(t_0(T), \ldots, t_N(T)) : \cdots : G_M(t_0(T), \ldots, t_N(T))\,.$$

This gives $h_{\mathcal{M}}(S) \leq c\, h_{\mathcal{N}}(T) + O(1)$ for some constant c, depending on $\mathcal{M}$ and $\mathcal{N}$. We can take $\mathcal{N} = \mathcal{O}(3\Theta)$, therefore

$$h_{\mathcal{M}}(S) \leq 3c\, \widehat{h}_\theta(T) + O(1)\,.$$

On the other hand, ϕ is finite of degree $g!$, so that for any positive line sheaf $\mathcal{L}$ on C^g we have $h_{\mathcal{L}}(R_1, \ldots, R_g) \leq c'\, h_{\mathcal{M}}(S) + O(1)$, with c' depending on $\mathcal{L}$ and $\mathcal{M}$. Taking $\mathcal{L} = \mathcal{O}(Q \times C^{g-1} + C \times Q \times C^{g-2} + \ldots + C^{g-1} \times Q)$, we have

$$h_{\mathcal{L}}(R_1, \ldots, R_g) = \sum_{i=1}^{g} h_Q(R_i) + O(1)\,.$$

This shows that

$$\sum_{i=1}^{g} h_Q(R_i) \leq c_6 \widehat{h}_\theta(T) + O(1) = \frac{1}{2} c_6 \mid T \mid^2 + O(1)\,.$$

We apply the lower bound for $h(P)$ to the points R_i in the decomposition $T = j(R_1) + \ldots + j(R_g)$, assuming that $T \notin W$ and $F(R_i) \neq 0$, $R_i \neq Q$, R_i

not a pole of both x and y, and sum the inequalities so obtained. In view of the above discussion, we obtain

$$\frac{g}{m+1} h(P) \geq \langle q^*, T \rangle - c_5 \,\delta \sum_{i=1}^{g} h_Q(R_i) - O(1) \geq \langle q^*, T \rangle - c_7 \,\delta \mid T \mid^2 - O(1),$$

with $c_7 = \frac{1}{2} c_5 c_6$.

Now we remark that given any $T \in J$, we can translate T by a torsion point ζ such that $T' = T + \zeta$ satisfies all the conditions above. In fact, we need to verify that $T' \notin W$ and that $T' \notin \Theta + j(S)$ with S in the finite set consisting of Q, the zeros of F and the common poles of x and y. Since $W \bigcup (\Theta + j(S))$ has codimension 1 in J and torsion points are dense in J, our remark becomes obvious.

On the other hand, the bilinear form $\langle \, , \, \rangle$ is invariant by translation in J_{tors}, therefore we conclude that the above inequality holds for every $T \in J(\overline{K})$.

It is now easy to conclude the proof of the theorem. By assumption $q^* \notin J_{\text{tors}}$, so that $\mid q^* \mid \neq 0$. We choose T to be any representative for νq^*, with $\nu \in \mathbb{Q}$ at our disposal. We take

$$\nu = \frac{1}{2 c_7 \delta} + O(1)$$

and get

$$\frac{g}{m+1} h(P) \geq \frac{1}{4 c_7 \delta} \mid q^* \mid^2 - O(1),$$

completing the proof of our theorem.

REFERENCES

[1] A. Baker, *Rational approximations to $\sqrt[3]{2}$ and other algebraic numbers*, Quart. J. Math. Oxford **15** (1964), 375-383.

[2] E. Bombieri, *On G-functions*, in "Recent Progress in Analytic Number Theory", H. Halberstam and C. Hooley (ed.), Academic Press, 1981, Vol. 2, 1-67.

[3] E. Bombieri, *On Weil's "Théorème de Décomposition"*, Amer. J. Math. **105** (1983), 295-308.

[4] E. Bombieri – J. Vaaler, *On Siegel's lemma*, Invent. math. **73** (1983), 11-32; Addendum, *ibid* **75** (1984), 177.

[5] E. Bombieri – U. Zannier, *Heights of algebraic points on subvarieties of abelian varieties*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (IV) **23** (1996), 779-792.

[6] D. V. Chudnovsky – G. V. Chudnovsky, *Padé approximations to solutions of linear differential equations and applications to diophantine analysis*, Springer Lecture Notes in Math. **1052** (1984), 85-167.

[7] P. Debes, *Quelques remarques sur un article de Bombieri concernant le Théorème de Décomposition de Weil*, Amer. J. Math. **107** (1985), 39-44.

[8] P. Debes, *G-fonctions et théorème d'irréductibilité de Hilbert*, Acta Arithmetica **XLVII** (1986), 371-402.

[9] K. Mahler, *Ein Beweis des Thue-Siegelschen Satzes über die Approximation algebraischen Zahlen für binomische Gleichungen*, Math. Annalen **105** (1931), 267-276.

[10] D. Mumford, "Curves and Their Jacobians", The Univ. of Michigan Press, Ann Arbor, 1976.

[11] J.-P. Serre, "Lectures on the Mordell-Weil theorem, Aspects of Mathematics", Vieweg & Sohn, Braunschweig, 1989.

[12] L. Szpiro – E. Ullmo – S. Zhang, *Equidistribution des petits points*, Invent. Math. **127** (1997), 337-347.

School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540, USA
e-mail:eb@ias.edu

URA Géométrie - Analyse - Topologie
Université des Sciences et
Technologies de Lille
59655 Villeneuve d'Ascq cedex, France
e-mail:pcohen@gat.univ-lille1.fr

Appendix

UMBERTO ZANNIER

In this Appendix we describe examples of Padé approximation on curves of positive genus where, in the notation of the present paper, $\delta = O(1/m)$ but nevertheless the height of the polynomials involved grows linearly in m. By Theorem 1, we must have points $Q_1, \ldots, Q_n$, with $nj(Q) - \sum j(Q_i) \in J(C)_{\rm tors}$ for Q one of the Q_i. We show here that, conversely, such Padé approximations can be constructed whenever $nj(Q) - \sum j(Q_i) \in J(C)_{\rm tors}$. This shows that the condition $nj(Q) - \sum j(Q_i) \notin J(C)_{\rm tors}$ is essential for the validity of the theorem.

For $n = 2$, this condition is that $j(Q_1) - j(Q_2) \in J(C)_{\rm tors}$. The same condition arises in several different contexts. One, studied by N. H. Abel [1] and subsequently by A. Schinzel [3], deals with continued fractions of square roots. The second occurs in work of Y. Hellegouarch, D. L. McQuillan and R. Paysant-Le Roux [2] on unit norm equations over function fields. We mentioned in the Introduction of the present paper work of D. V. Chudnovsky and G. V. Chudnovsky on the case $n = 2$.

Abel was concerned with the integration "in terms of logarithms" of the differential $\rho(x)\mathrm{d}x/\sqrt{R(x)}$, for polynomials ρ, R. He had observed, in some cases, formulas for the corresponding indefinite integral of the type $\log \frac{y+\sqrt{R}}{y-\sqrt{R}}$, with y a rational function, and he sought general conditions for their existence. He found that such a formula exists for some ρ precisely when $\sqrt{R}$ admits a periodic continued fraction whose partial quotients are polynomials. This turned out to be equivalent to the solvability of a Pell equation[(1)]

$$U^2(X) - R(X)V^2(X) = 1$$

in nonzero polynomials U and V.

To see the connection with the present paper, suppose we have a solution (U, V) to this equation and that $R(X)$ has degree $2p \geq 4$ and no multiple root. The curve $Y^2 = R(X)$ has genus $p - 1$. We define (U_s, V_s) by

$$U_s(X) + YV_s(X) = \big(U_1(X) + YV_1(X)\big)^s .$$

Then (U_s, V_s) is also a solution and $\deg(U_s) = \deg(V_s) + p = s\deg(U_1) = sd$, with $d = \deg(U_1)$.

[(1)] The name Pell's equation is a misnomer originating with Euler. See L. E. Dickson, History of the Theory of Numbers, Chelsea 1952, vol. II, Ch. XII, p. 341 and ref. 62, p. 354.

Now set $x = 1/X$ and multiply by x^{2sd}. We obtain

$$U_s^*(x)^2 - R^*(x)V_s^*(x)^2 = x^{2sd}$$

where for a polynomial $W(x)$ we define $W^*(x) = x^{\deg(W)}W(1/x)$.

Let $y^* = \sqrt{R^*(x)} = x^p\sqrt{R(1/x)}$. For a suitable choice of a branch, $U_s^*(x) + y^*V_s^*(x)$ vanishes to order $2sd$ at $x = 0$. We therefore have a Padé approximation with $m = sd$, $n = 2$ and $(n - \delta)(m + 1) = 2sd$, that is $\delta = 2/(m+1)$. On the other hand, when $R \in \overline{\mathbb{Q}}[x]$ the polynomials U_s^*, V_s^* will have height bounded by $O(m)$. We are working now on the curve $(y^*)^2 = R^*(x)$, and, in the notation of this paper, Q_1, Q_2 are the distinct points $(0, \pm\sqrt{R^*(0)})$. By Theorem 1, the difference $j(Q_1) - j(Q_2)$ must be a torsion point on $J(C^*)$. Of course, this may be checked directly; in fact the functions $\varphi_\pm = U \pm yV$ may have poles only at infinity whence, noting that $\varphi_+\varphi_- = 1$, their zeros must also lie at infinity. This means that the poles and zeros of $\varphi^* = U^* + y^*V^*$ are in the set $\{Q_1, Q_2\}$, so the divisor of φ^* must be of the form $h\cdot((Q_1) - (Q_2))$, for some nonzero integer h, which implies that $h \cdot (j(Q_1) - j(Q_2)) = 0$.

Now we construct more general examples which show that Theorem 1 is sharp. Namely, the hypothesis $q^* \notin J(C)_{\mathrm{tors}}$ cannot in general be weakened.

For notational convenience, we identify the point Q with Q_1.

We introduce the two new functions $X = 1/x$, $Y = y/x$ and note that X has divisor of poles $(X)_\infty = Q_1 + \ldots + Q_n$ while Y is regular and not 0 at Q_1, because both x and y are uniformisers at Q_1.

Since $K(C) = K(x, y) = K(X, Y)$ we may write any $\varphi \in K(C)$ in the form

$$\varphi = a_0(X) + a_1(X)\,Y + \ldots + a_{n-1}(X)\,Y^{n-1}, \qquad a_i \in K(X)\,.$$

If in addition $\mathrm{div}_\infty(\varphi)$ has support contained in $\{Q_1, \ldots, Q_n\}$ then φ is integral over $K[X]$ and there is a polynomial $\Delta(X)$, independent of φ, such that $A_i(X) = a_i(X)\Delta(X) \in K[X]$ (it suffices to take Δ to be the discriminant of a minimal equation for Y over $K[X]$).

By assumption, $\mathcal{D} = h \cdot (nQ_1 - x^{-1}(0))$ is the divisor of a function $\varphi \in K(C)$. Then $\mathrm{div}(\varphi^s) = s\mathcal{D}$ and by the preceding remark we have

$$\Delta\,\varphi^s = A_{0s}(X) + A_{1s}(X)\,Y + \ldots + A_{n-1,s}(X)\,Y^{n-1}$$

for certain polynomials $A_{is}(X) \in K[X]$.

We claim that the polynomials A_{is} have maximum degree bounded by

$$\deg(A_{is}) \le hs + \frac{n(n-1)}{2}N + \frac{1}{2}\deg(\Delta)$$

where N is the degree of the rational function y on C. For the proof, let $Y_1, \ldots, Y_n$ be the conjugates of Y over $K(X)$, so that we may consider Y_i as Puiseux series in the uniformiser $1/X$ at Q_i. Since X is unramified at ∞, each

Y_i is in fact a Laurent series in $1/X$. We proceed in the same way for the conjugates φ_i of φ over $K(X)$ and obtain the equations

$$A_{0s}(X) + Y_i\, A_{1s}(X) + \ldots + Y_i^{n-1}\, A_{n-1,s}(X) = \Delta\, \varphi_i^s, \qquad i = 1, \ldots, n,$$

which we view as a linear system for the polynomials A_{is}. Solving by Cramer's Rule we get

$$A_{is}(X) = \sum_{j=1}^{n} \frac{V_{ij}}{V}\, \Delta\, \varphi_i^s,$$

where V is the Vandermonde determinant of the Y_j^i and where V_{ij} is the cofactor of Y_j^i in V. This gives

$$\deg(A_{is}) \le \deg_X(\Delta) - \deg_X(V) + s \max_i \big(-\operatorname{ord}_{Q_i}(\varphi)\big) + \max_{ij} \deg_X(V_{ij}).$$

Now $\deg_X(V) = \frac{1}{2}\deg_X(\Delta) \le \frac{1}{2}\deg(\Delta)$ because $\Delta = V^2$, also $\operatorname{ord}_{Q_i}(\varphi) \ge -h$ and $\deg_X(V_{ij}) \le (n(n-1)/2)\max_i \max\big(-\operatorname{ord}_{Q_i}(Y), 0\big)$. This proves our claim.

On the other hand,

$$\operatorname{ord}_{Q_1}(\Delta\,\varphi^s) = (n-1)hs + \operatorname{ord}_{Q_1}(\Delta) \ge (n-1)hs - \deg(\Delta).$$

Finally, consider the polynomial in x, y given by

$$P(x, y) = \sum_{i=0}^{n-1} \left(x^D A_{is}(1/x) x^{n-1-i}\right) y^i = x^{D+n-1} \Delta\, \varphi^s,$$

where we have abbreviated

$$D = hs + \frac{n(n-1)}{2} N + \deg(\Delta) \ge \max_i \deg(A_{is}).$$

Then P has degree at most $D + n - 1$ in x, and the associated function F on C vanishes at $(0, 0)$ to order

$$\begin{aligned}\operatorname{ord}_{Q_1}(\Delta\,\varphi^s) + \operatorname{ord}_{Q_1}(x^{D+n-1}) &= \operatorname{ord}_{Q_1}(\Delta\,\varphi^s) + D + n - 1\\ &\ge (n-1)hs - \deg(\Delta) + hs + \frac{n(n-1)}{2} N + \deg(\Delta) + n - 1\\ &\ge nhs + \frac{n(n-1)}{2} N.\end{aligned}$$

Setting $m = D + n - 1$ we have obtained an (m, δ)-Padé approximation with δ such that

$$(n - \delta)(D + n) \ge nhs + \frac{n(n-1)}{2} N,$$

giving

$$(m+1)\delta \le n(D+n) - nhs - \frac{n(n-1)}{2} N$$
$$\le \frac{n^3+n}{2} N + n \deg(\Delta),$$

so that $\delta = O(1/m)$.

The height of P is bounded by the height of A_{is}, and this is bounded linearly in s, and hence in m. We have

$$\Delta \varphi^s = \sum_{i=0}^{n-1} A_{is}(X) Y^i$$

and there are rational functions $b_{ji}(X) \in K(X)$ such that

$$\varphi Y^i = \sum_{j=0}^{n-1} b_{ji}(X) Y^j .$$

This gives the recurrence

$$A_{i,s+1}(X) = \sum_{j=0}^{n-1} b_{ij}(X) A_{js}(X)$$

and our claim follows by induction on s.

As a final remark, let C/K be an elliptic curve with Mordell-Weil group of rank at least $n-1$ and let $Q_1, \dots, Q_{n-1}$ be $n-1$ points in $C(K)$ generating a subgroup of rank $n-1$. Setting $Q_n = (n-1)Q_1 - Q_2 - \dots - Q_{n-1}$ we see that $(n-1)Q_1 - Q_2 - \dots - Q_n = 0$ generates all relations among $Q_1, \dots, Q_n$. This gives an example where the growth of the coefficients of Padé approximations changes from c^m to c^{m^2} if we replace $u(x)$ by any of its conjugates over $K(x)$.

REFERENCES

[1] N. H. Abel, *Über die Integration der Differential-Formel $\rho dx/\sqrt{R}$, wenn R und ρ ganze Funktionen sind*, Journal für die reine und angew. Math. **1** (1826), 185-221. Also *Sur l'intégration de la formule differentielle $\rho dx/\sqrt{R}$, R et ρ étant des fonctions entieres*, Oeuvres Complètes, Tome Prémier, Christiania 1881, 104-144.

[2] Y. Hellegouarch – D. L. McQuillan – R. Paysant-Le Roux, *Unités de certain sous-anneaux des corps de fonctions algébriques*, Acta Arith. **XLVIII** (1987), 9- 47.

[3] A. Schinzel, *On some problems of the arithmetical theory of continued fractions II*, Acta Arith. **VII** (1962), 287-298.

Istituto Universitario di Architettura
Dipartimento di Costruzione dell'Architettura
Santa Croce 191, 30135 Venezia, Italy
e-mail:zannier@brezza.iuav.unive.it