Determinants in the Study of Thue's Method and Curves with Prescribed Singularities

Enrico Bombieri, David C. Hunt and Alfred J. van der Poorten

CONTENTS

Our investigations in the 1980's of Thue's method yielded determinants that we were only able to analyse successfully in part. We explain the context of our work, recount our experiences, mention our conjectures, and allude to a number of open questions.

1. INTRODUCTION

During 1983, and again in 1986, we attempted the explicit construction of the auxiliary polynomial required in Thue's method, as then recently refined by the first author. That led to our battling, and partly taming, some interesting determinants. Of course, the moment we finished we found an easier way. Nevertheless, the context warrants explaining, and our experiences seem worth recounting. Although our program has been only partially successful, we hope that telling its story may contribute to its eventual completion.

In passing, and perhaps peripheral to our principal objective, we noticed a host of fascinating facts and some curious identities. All are surely of intrinsic interest. We mention those facts and display our conjectured identities.

2. APPROXIMATION OF ALGEBRAIC NUMBERS BY ALGEBRAIC NUMBERS

Let $\alpha \neq 0$ be algebraic of degree r over $\mathbb{Q}$, with minimal polynomial over $\mathbb{Z}$ given by

$$\begin{aligned} f(X) &= a_0X^r + a_1X^{r-1} + \cdots + a_{r-1}X + a_r \\ &= a_0 \prod_\sigma (X - \sigma\alpha), \end{aligned}$$

Van der Poorten was supported in part by grants from the Australian Research Council and by a research agreement with Digital Equipment Corporation.

1991 Mathematics subject classification: 11J68, 11C20

1058-6458/96 $0.50 per page

where σ runs over all distinct embeddings of $\mathbb{Q}(\alpha)$ into $\mathbb{C}$.

We define the absolute logarithmic height $h(\alpha)$ of α by

$$\begin{aligned} h(\alpha) &= \log H(\alpha) \\ &= \frac{1}{r}\Bigl(\log|a_0| + \sum_\sigma \max(0, \log|\sigma\alpha|)\Bigr) \\ &= \frac{1}{r}\sum_v d_v \log^+ |\alpha|_v, \end{aligned}$$

with the latter sum over the appropriately normalised absolute values v of $\mathbb{Q}(\alpha)$; here d_v is the degree of the complete field $\mathbb{Q}(\alpha)_v$ over the complete field $\mathbb{Q}_v$. As usual, $\log^+ x = \max(0, \log x)$. It is also convenient to define $h(0) = 0$. The absolute values are normalized so that their restriction to the rational field $\mathbb{Q}$ coincide with the usual p-adic or real absolute values. The second definition of height, with the same normalization for the absolute values, makes sense in any number field K containing α, provided we take r to be the degree of K; the height so defined is independent of K, whence the name absolute height.

It is a nice exercise to verify that

$$\log H(\alpha) = \frac{1}{r}\int_0^1 \log|f(e^{2\pi i t})|\,dt.$$

It is a theorem of Kronecker, and not completely trivial, that $h(\alpha) = 0$ if and only if α is a root of unity or 0.

We also have the useful facts, which we shall use without further warning, that

$$\begin{gathered} h(1/\alpha) = h(\alpha), \\ h(\alpha_1 + \cdots + \alpha_n) \le h(\alpha_1) + \cdots + h(\alpha_n) + \log n, \\ h(\alpha\beta) \le h(\alpha) + h(\beta), \end{gathered}$$

the first equation being of course restricted to nonzero α.

The *product formula* says that either $\alpha = 0$ or $\sum_v \log|\alpha|_v = 0$. We easily deduce the *fundamental inequality* of diophantine approximation:

$$\alpha = 0 \quad \text{or} \quad \log|\alpha| \ge -rh(\alpha).$$

Then, if $p/q \in \mathbb{Q}$ and $\alpha \ne p/q$, we have a precise form of *Liouville's Theorem*:

$$|\alpha - p/q| \ge (2H(\alpha)\max(|p|,|q|))^{-r}.$$

Indeed, set $\beta = \alpha - p/q$. Then

$$h(\beta) \le h(\alpha) + h(p/q) + \log 2,$$

and the result follows from $h(1/\beta) = h(\beta)$ and the fundamental inequality.

It is worthwhile to endeavour to sharpen Liouville's inequality. Henceforth let $r \ge 3$. Consider the binary form

$$\begin{aligned} Y^r f(X/Y) &= F(X,Y) \\ &= a_0X^r + a_1X^{r-1}Y + \cdots + a_{r-1}XY^{r-1} + a_rY^r \end{aligned}$$

and an equation

$$F(X,Y) = m.$$

If $|\alpha - p/q| > q^{-\mu-\varepsilon}$ for $q > Q_0(\varepsilon)$, any $\varepsilon > 0$ and a fixed $\mu < r$, then the equation has only finitely many solutions in integers. In fact, suppose $q^r f(p/q) = F(p,q) = m$. Let α be the zero of f closest to p/q. Taking the logarithmic derivative we obtain

$$\left|\frac{f'(p/q)}{f(p/q)}\right| \le \sum_{f(\beta)=0} \frac{1}{|p/q - \beta|} \le \frac{r}{|p/q - \alpha|};$$

therefore $|\alpha - p/q| \le r\,|m|\,q^{-r}\,|f'(p/q)|^{-1}$. Since

$$f(p/q) = m\,q^{-r},$$

we have that p/q is close to α for sufficiently large q, and since $f'(\alpha) \ne 0$ we see that $f'(p/q)$ is bounded away from zero for large q. This contradicts the sharpened Liouville inequality.

A. Thue showed in 1909 that

$$|\alpha - p/q| > q^{-\mu-\varepsilon},$$

with $\mu = \frac{1}{2}r + 1$, for $q > Q_0(\varepsilon)$. This was improved by Siegel in 1921 to, in effect, $\mu = 2\sqrt{r}$. F. Dyson in 1947, and independently A. O. Gel'fond in 1949, obtained $\mu = \sqrt{2r}$. Finally, K. F. Roth in 1955 proved the best possible result $\mu = 2$, independent of $r > 2$.

But these results are *ineffective.* One proves that Q_0 exists, but there is no method for obtaining it explicitly.

More recently, there have been general effective results. The ideas introduced by A. Baker in 1966 led to $\mu = r - \eta$, effectively, thus obtaining the effective solution of the general Thue equation. Baker's methods are quite different from those described here, and $\eta > 0$ is in general rather small.

Loosely speaking, the reason for the ineffectivity in Thue's method and its developments is as follows: Thue's method, as refined by Gel'fond, leads to the result that there are two functions $Q_1 = Q_1(\alpha, \varepsilon)$ and $Q_2 = Q_2(\alpha, \varepsilon, q_1)$ such that if

$$|\alpha - p_1/q_1| < q_1^{-\mu_1} \quad \text{and} \quad |\alpha - p_2/q_2| < q_2^{-\mu_2}$$

for $q_i > Q_i$, then $\mu_1\mu_2 \le 2r + \varepsilon$. Taking $\mu_1 = \mu_2 > \sqrt{2r}$ we easily see that we cannot have an infinite sequence of solutions to $|\alpha - p/q| < q^{-\mu_1}$, which is the Dyson–Gel'fond Theorem.

Ineffectivity comes from the fact that Q_2 depends on the denominator q_1 of the first approximation: One must exhibit the approximation p_1/q_1 in order to determine Q_2. Unfortunately Q_1 is rather large, so large that no examples of the first inequality being satisfied with μ_1 usefully larger than 2 are known to us. Thus although Q_1 and Q_2 are explicit there is nothing to which to attach, or *anchor*, Thue's method to obtain effective results.

That Q_1 is large was inherent to Thue's original method. The basic argument can be summarized as follow.

Step 1. Construct an auxiliary polynomial $P \in \mathbb{Z}[X, Y]$, of bidegree (N_1, N_2) and with small coefficients, such that the initial Taylor coefficients of P at (α, α) vanish.

Step 2. Next, note that $P(p_1/q_1, p_2/q_2)$ is a rational number with denominator at most $q_1^{N_1}q_2^{N_2}$, hence either it is 0 or at least $1/(q_1^{N_1}q_2^{N_2})$ in absolute value. Then use the Taylor expansion of P at (α, α) and step 1 to deduce that $|\alpha - p_i/q_i| < q_i^{-\mu_i}$ implies that $P(p_1/q_1,\, p_2/q_2)$ is exceedingly small, and so, by the preceding remark, that $P(p_1/q_1,\, p_2/q_2) = 0$.

Step 3. Prove directly that, possibly replacing P by a partial derivative of rather small order, we have $P(p_1/q_1,\, p_2/q_2) \ne 0$, and deduce that we cannot have the two approximations $|\alpha - p_i/q_i| < q_i^{-\mu_i}$, where $i = 1, 2$, whose existence was supposed in step 2.

The basic idea used by Thue and other researchers to show the key step 3, namely $P(p_1/q_1,\, p_2/q_2) \ne 0$, was a two-dimensional version of the obvious fact that a polynomial g in one variable, with rational integral coefficients and with a rational root p/q of multiplicity m, has leading coefficient divisible by q^m. In particular, for some $h \le m$, there is a derivative $g^{(h)}(p/q) \ne 0$ as soon as $q > H(g)^{1/m}$, that is, if q is sufficiently large.

This approach to step 3 obviously required q_1 large, and Thue's method remained ineffective. In 1982, however, the first author noticed in Dyson's work a different method for showing that the auxiliary function does not vanish at the approximating point.

The old method relied on the principle that the height of a zero of P is controlled in terms of the height of P—hence the large size of Q_1. Dyson's observation was that the number of zeros, counted according to multiplicity, in a finite set of zeros of P is controlled by the bidegree of P and the number of distinct such zeros. The upshot is that there need be no lower bound on Q_1 at all.

Suppose we have constructed a polynomial

$$P(X, Y) = \sum_{j_1=0}^{N_1} \sum_{j_2=0}^{N_2} a_{j_1 j_2} X^{j_1} Y^{j_2} \in \mathbb{Z}[X, Y]$$

such that $P \not\equiv 0$ but lots of partial derivatives vanish:

$$P^{(i_1, i_2)}(\alpha_1, \alpha_2) = 0 \quad \text{for} \quad \theta_1 i_1/N_1 + \theta_2 i_2/N_2 < t.$$

Here $\mathbb{Q}(\alpha_1) = \mathbb{Q}(\alpha_2)$ is of degree r over $\mathbb{Q}$, and the θ_i and t are parameters satisfying $\theta_1\theta_2 = 1$, with $t < \theta_2 < t^{-1}$ and $\frac{1}{2}rt^2 < 1$.

Say we have constructed P with height $H(P)$—the height of the vector of coefficients of P—where

$$\log H(P) < A_1 N_1 \log H(\alpha_1) + A_2 N_2 \log H(\alpha_2).$$

Here A_1 and A_2 depend on the various parameters of the question. This is not difficult to achieve, and should be seen as little more than defining the A_i. In fact, we will see that the construction is a matter of somehow solving $\sim \frac{1}{2} r t^2 N_1 N_2$ linear equations in $\sim N_1 N_2$ unknowns.

Then carrying out the three steps outlined before (but using Dyson's method in step 3) yields what may be called the Thue–Siegel Principle [Bombieri 1982]:

Thue–Siegel Principle. *Suppose that β_1 and β_2 are rationals, that α_1 is algebraic of degree $r \geq 2$ and that $\alpha_2 \in \mathbb{Q}(\alpha_1)$. Let also t, τ be real numbers with $\sqrt{2 - rt^2} < \tau < t$. Then, if*

$$|\alpha_1 - \beta_1| < \left(c^{A_1} H(\alpha_1)^{A_1} H(\beta_1)\right)^{-\frac{2\theta_1}{t-\tau}}$$

and

$$c^{A_2} H(\alpha_2)^{A_2} H(\beta_2) > \left(c^{A_1} H(\alpha_1)^{A_1} H(\beta_1)\right)^{\frac{\tau}{rt^2+\tau^2-2}},$$

we have

$$|\alpha_2 - \beta_2| > \left(c^{A_2} H(\alpha_2)^{A_2} H(\beta_2)\right)^{-\frac{2\theta_2}{t-\tau}}.$$

Here c is a small absolute constant: $c = 3$ will do. There is no lower bound on $H(\beta_1)$.

We should mention that this introductory description readily allows the base field $\mathbb{Q}$ to be replaced by any fixed number field; moreover, the absolute value $|\ |$ may be deemed to stand for any properly normalized absolute value, archimedean or nonarchimedean. Versions involving several absolute values, useful for certain applications, can also be obtained rather easily.

By construction, P vanishes $\sim \frac{1}{2} t^2 N_1 N_2$ times at each of the r conjugate points $(\sigma\alpha_1, \sigma\alpha_2)$. All will be well if P does not vanish too often (that is, if not too many contiguous partial derivatives vanish) at (β_1, β_2). Say

$$P^{(i_1,i_2)}(\beta_1, \beta_2) \neq 0$$

for some (i_1, i_2) with

$$\frac{\theta_1 i_1}{N_1} + \frac{\theta_2 i_2}{N_2} \leq \tau.$$

In this context, Dyson's Lemma asserts that

$$\tfrac{1}{2} r t^2 N_1 N_2 + \tfrac{1}{2} \tau^2 N_1 N_2 < N_1 N_2 + \tfrac{1}{2} r {N_2}^2.$$

On the left is the number of zeros counted according to multiplicity. There are however just $r+1$ distinct such zeros. On the right we have, as it were, the "volume" $N_1 N_2$ of P, and an adjustment term $\frac{1}{2} r {N_2}^2$ depending on the number of distinct zeros. Of course it is the asymmetric form of that adjustment term that provides the real profit. That bounds τ, allowing one to employ $P^{(\tau)}$, so to speak, as one's auxiliary function in place of P, the point being that automatically $P^{(\tau)}(\beta_1, \beta_2) \neq 0$, yielding the key step 3. In the argument one eventually lets N_1, $N_2 \to \infty$ so that $N_2/N_1 \to 0$.

Ideally, also $t \nearrow \sqrt{2/r}$, $\theta_2 \searrow t$ and $\tau \searrow 0$. Indeed, if one could always perform the construction with $A_1 = 1$, allowing $\tau \searrow 0$, one would seemingly even have Roth's Theorem. But in practice one still has difficulties with the *anchor condition*

$$|\alpha_1 - \beta_1| < (c^{A_1} H(\alpha_1)^{A_1} H(\beta_1))^{-\frac{2\theta_1}{t-\tau}},$$

although even back in 1982 that could be arranged to hold in infinitely many number fields. The trouble was, and remains, that the only obvious general construction leads to

$$A_1 \approx \tfrac{1}{2} r t^2 / (1 - \tfrac{1}{2} r t^2),$$

which unfortunately increases to ∞ as $t \nearrow \sqrt{2/r}$.

That led Bombieri and Vaaler [1983] to restudy the problem of solving systems of linear equations and obtaining economical integer solution vectors. In vulgar terms, they show that one obtains a generalisation of Cramer's rule yielding solutions as quotients of determinants. In more highbrow terms, they find an upper bound for the geometric mean of the heights of a basis of solution vectors in terms of the height of the linear system. Whatever, in the present situation one finds oneself studying determinants of the maximal square minors of a matrix with typical entry

$$\binom{j_1}{i_1}\binom{j_2}{i_2}x_h^{j_1-i_1}y_h^{j_2-i_2}.$$

Here columns are indexed by the pairs (j_1, j_2) and rows by the triples (h, i_1, i_2). Constraints are $0 \le Z = j_1/N_1 \le 1$ and $0 \le W = j_2/N_2 \le 1$; and on setting $w = i_1/N_1$, $z = i_2/N_2$ also $\theta_1 z + \theta_2 w < t$. We found it congenial to write (x_h, y_h) for

$$(\sigma_h\alpha_1,\, \sigma_h\alpha_2),$$

so $h = 1, 2, \ldots, r$.

One's hope is to discover that these determinants have many factors in common. Then such factors may be 'cancelled' and the complementary factors remaining might allow one to proceed to the limit $t \nearrow \sqrt{2/r}$, so also $\tau \searrow 0$, leading to a small A_1.

Accordingly, during much of 1983, Bombieri and van der Poorten made a detailed and exhausting study of determinants

$$\left|\binom{j_1}{i_1}\binom{j_2}{i_2}x_h^{j_1-i_1}y_h^{j_2-i_2}\right|.$$

The need to bolster our flagging confidence eventually led us to designate David Hunt to assist and hit these objects with his computer.

Remark. This 'setting the scene' section, and the comments immediately following, are a reasonably faithful rendition of a lecture given by the third author at the annual meeting of the Japanese Mathematical Society in April, 1984. Since then, Dyson's Lemma has been explained [Viola 1985], trivialised [Beukers et al 1991], generalised [Vojta 1989], and extended to the case of many variables [Esnault and Viehweg 1984]. See also [Vojta 1991; Viola 1992]. The several-variable result [Esnault and Viehweg 1984] yields Roth's theorem in ineffective form—the chain of anchor conditions cannot be realised—but with significantly improved explicit parameters; see [Bombieri and van der Poorten 1988].

The new approximation method has acquired the name 'The Thue Principle'. There is a variety of applications; see, for example, [Bombieri and Mueller 1983]. More recently [Bombieri 1993] it has become apparent that an equivariant generalization of the Principle, with various refinements whose genesis is about to be described, does yield effective approximation in all algebraic extensions of number fields.

3. CONFLUENT VANDERMONDE DETERMINANTS

Thus, in 1983, the first and third authors attacked the determinants

$$\left|\binom{j_1}{i_1}\binom{j_2}{i_2}x_h^{j_1-i_1}y_h^{j_2-i_2}\right|,$$

with the hope that all such determinants might share substantial common factors in the 'variables' x_h and y_h, thereby significantly decreasing the critical parameter A_1.

Recall that the rows are given by the triples (i_1, i_2, h) with $z = i_1/N_1$ and $w = i_2/N_2$ lying in the triangle $\theta_1 z + \theta_2 w < t$. Because $h = 1, 2, \ldots, r$, there are r such triangles, each of 'volume' $\sim \frac{1}{2}t^2$ in the normalised unit square. So, as already noted, we must have $\frac{1}{2}rt^2 < 1$.

The columns are pairs (j_1, j_2). Setting $Z = j_1/N_1$ and $W = j_2/N_2$, the columns of each determinant in effect comprise a subset J of area $\frac{1}{2}rt^2$ of the 'column unit square'. Thus we are concerned with factors common to all the determinants as J ranges over all subsets with $|J| \sim \frac{1}{2}rt^2$. In the one-dimensional case our determinants reduce to $\left|\binom{j}{i}x_h^{j-i}\right|$, and one is faced by a confluent Vandermonde determinant. This is not all that dreadful. Suppose we have $rt = 1$, and study the full determinant Δ, corresponding to $\frac{1}{2}rt^2 = 1$ in the two-dimensional case. A naïve strategy to cope with these confluent objects is simply to consider an honest-to-goodness Vandermonde determinant $|x_{hi}^j|$, viewing variables x_{hi} with the same h as differing only by infinitesimals. One now divides by the infinitesimals, to wit by $\prod_{h=1}^{r}\prod_{i<i'}(x_{hi}-x_{hi'})$, nakedly revealing that

$$\Delta = \pm V_x^{(1+o(1))t^2N^2}$$

with $V_x = \prod_{h<h'}(x_h - x_{h'})$. The naïve strategy also exposes the detail of the symmetric factor additional to the difference product in the general case $rt < 1$. These matters are mentioned in the literature, and one can turn to [Muir 1911] and the treatise [Muir 1933] with advantage; see also [van der Poorten 1976]. Incidentally, in studying that general case it turns out to be a good strategy to pretend to still be in the easy full case Δ. Thus, rather than omitting $(1-rt)N$ columns, one simply acknowledges that $(1-rt)N$ of the rows are *ghosts*, present only as 0 's and a 1 in the column that was to have been omitted. The x_{hi} corresponding to ghost rows thus manifest themselves as a sort of δ-function, with all their powers infinitesimal, other than that $x_{hi}^{j(hi)} = 1$. Here the $j(hi)$-th column is the one reprieved by the spectre of the hi-th row.

In the two-dimensional case it is not at all obvious *a priori* that our determinants have any difference factors at all, even if $\frac{1}{2}rt^2 = 1$. The trouble is that x_h and y_h are tied together: if all pairs $(x_h, y_{h'})$ were to appear, our problem would factorise as a Kronecker product of matrices in one set of variables. But in general there need be no difference factors at all; yet, remarkably:

For $r = 3$, the full determinant (thus $\frac{1}{2}rt^2 = 1$)

$$\Delta = \left|\binom{j_1}{i_1}\binom{j_2}{i_2}x_h^{j_1-i_1}y_h^{j_2-i_2}\right|$$

factorises completely as a product of difference factors

$$c(\triangle, t)V_x^{(\frac{1}{4}t^2-\frac{1}{6}\theta_2t^3+o(1))N_1^2N_2}V_y^{(\frac{1}{4}t^2-\frac{1}{6}\theta_1t^3+o(1))N_1N_2^2}.$$

We say more about the constant $c(\triangle, t)$ below. Actually, we should be more careful in quoting those exponents, because we failed to say that our remark above presumes N_1, $N_2 \to \infty$. In the 'finite' case the exact exponent of V_x is, for example,

$$2N_1|\triangle| - \tfrac{1}{2}N_1(N_1+1)N_2 - \sum_{\triangle} i_1.$$

When $\frac{1}{2}rt^2 < 1$, as it must be in the actual construction of an auxiliary polynomial, one loses some difference factors... in fact, possibly as many as $(1-\frac{1}{2}3t^2)N_1^2N_2$ copies of V_x. The upshot appeared to be that we could prove that $1 \le A_1 \le 3$, depending on delicate properties of α_1. For $\alpha_1 = \sqrt[r]{a}$ our arguments plainly suggested $A_1 = 1$. As it happened, that case was understood [Bombieri and Mueller 1983], and our joy subsided as it rapidly dawned on us that we were choosing to forget the influence of $c(\triangle, t)$ on the actual value of A_1.

When above we defined the A_i in describing the height of the auxiliary polynomial P we might have constructed, we should have written something like

$$\log H(P) < B_1N_1\log H(\alpha_1) + B_2N_2\log H(\alpha_2) + C_1N_1 + C_2N_2,$$

emphasising that the A_i come in two parts, with the parts C_i independent of the $H(\alpha_i)$. Our remarks immediately above apply to B_1, rather than to A_1. Still, our temporary lapse had proved profitable. Our fervor to factorise Δ might have been greatly lessened had we been fully conscious that it was just a first easier step.

There is little purpose in trying to detail the contortions that led to our, now seen to be partial, success in the case $r = 3$. Once we had the factorisation of Δ we rapidly found simpler explanations for it. Still, it may be interesting to mention that, *en route*, as we discovered more and more batches of difference factors more and more closely comprising all, our confidence did falter. In truth, we had not anticipated the complete factorisation in the general cubic case, and feared we must have overshot. During a long afternoon, evening, and night the first author computed, by hand, an 18×18 determinant for $r = 3$ with $N_1 = 5$ and $N_2 = 2$ and found it to be divisible by all the difference factors to their appropriate multiplicities.

That had an amusing side effect in severely balking the attempts by the second author to debug his programs, for those programs persisted in claiming that this determinant vanished identically. It was chastening to realise, eventually, that indeed in this case $c(◺, t) = 0$.

4. QUESTIONS, AND SOME ANSWERS

Just why does the full determinant Δ factorise completely into difference factors when $r = 3$? One might argue as follows. The determinant Δ vanishes when some of the conditions placed on the curve $P(X, Y)$ are redundant. If one studies the configuration of singularities we impose, one sees singularities at $(1, 0, 0)$, $(0, 1, 0)$ and at the finite points (x_h, y_h). The infinite singularities are our imposition of the bidegree of P. Now the only *degenerate* configurations of the five points are when two of the finite points line up with a singularity at ∞—that is, exactly when $x_h = x_{h'}$ or $y_h = y_{h'}$, for some $h \neq h'$. We might have avoided some herculean calculations had we just chosen to employ pure thought!

When $r = 4$ there is an additional degenerate configuration. Indeed, the six points might all lie on a conic, suggesting the presence of the factor

$$B_{x,y} = \begin{vmatrix} 1 & x_1 & y_1 & x_1 y_1 \\ 1 & x_2 & y_2 & x_2 y_2 \\ 1 & x_3 & y_3 & x_3 y_3 \\ 1 & x_4 & y_4 & x_4 y_4 \end{vmatrix}.$$

We had already seen this numerically in the case $N_1 = 9$, $N_2 = 3$ and ◺ = ⁝∴, finding

$$\Delta = V_x^{20} B_{x,y}^{10} .$$

In this case Δ is a 40×40 determinant.

Of course, and so on. We started thinking about classical invariants. However, it remains an open problem how to predict the factorisation of Δ in any of the cases $r > 3$, even knowing the shapes of the possible factors. In the case $r = 3$ it suffices to know that the only factors are differences, because then symmetry and counting degrees suffices to write the factorisation. That is not the case once more complicated factors can appear.

On looking closely at some finite examples, we did realise that when $r = 3$ we could handle the easiest, almost degenerate, case $N_2 = 1$ completely. In this case the 'triangle' is just the points $(0, 0)$, $(0, 1)$, …, $(0, \frac{1}{3}(2N_1 + 1))$. It will be useful to set $N_1 = 3N + 1$. As it happened, we laboriously evaluated the determinants, thereby usefully honing our skills, but we then realised that we had already known the exact formula. The idea goes back to Hermite and is generalised by Mahler. There is a brief introductory summary in the obituary [van der Poorten 1991]. Mahler remarks that

$$\oint_{|\zeta|=R} \frac{(1-z)^{\zeta}}{\prod_{i=1}^{m} \prod_{k=1}^{\rho_i} (\zeta - \gamma_i - k + 1)} \, d\zeta = \sum_{i=1}^{m} A_i(z)(1-z)^{\gamma_i} = O(z^{\rho_1 + \cdots \rho_m - 1}) .$$

Here R is so large so that all the poles of the integrand are included in the contour. The $A_i(z)$ on the right are polynomials of degree not exceeding $\rho_i - 1$; that is readily seen on evaluating the integral by residues. Viewing the contour as being around ∞ immediately yields the alleged vanishing.

We need $m = 2$ and $\rho_1 = \rho_2 = N + 1$, and we take $\gamma_1 = 2N + 1$, $\gamma_2 = 0$. Now replace $1 - z$ by x and $A_i(1 - x) = u_i(x)$ and consider

$$p(x, y) = u_1(x) x^{2N+1} + u_2(x) y$$

as a polynomial of bidegree $(3N + 1, 1)$. We claim it is clear that

$$p^{(i,0)}(0, 0) = p^{(i,0)}(1, 1) = p^{(i,0)}(\infty, \infty) = 0$$

for $i = 0, 1, \ldots, 2N$. Given that, we need only some sleight of hand. Note, incidentally, by a more careful look at the integral, that up to a normalisation we have

$$u_1(x)=\sum_{j=0}^{N}u_{1j}x^j=\sum_{j=0}^{N}(-1)^{N-j}\binom{3N+1}{N-j}\binom{N+j}{j}x^j,$$
$$u_2(x)=\sum_{j=0}^{N}u_{2j}x^j=\sum_{j=0}^{N}(-1)^{N-j}\binom{3N+1}{j}\binom{2N-j}{N-j}x^j.$$

Then we assert that $N_2 = 1$ implies that, up to a normalisation, the required auxiliary polynomial of bidegree $(3N + 1,\, 1)$ is essentially

$$\begin{aligned}P(X,Y)&\\ =\sum_{j=0}^{N}&u_{1j}\Big(\frac{x_1-X}{x_1-x_2}\Big)^{2N+j+1}\Big(\frac{X-x_3}{x_2-x_3}\Big)^{N-j}\Big(\frac{Y-y_3}{y_2-y_3}\Big)\\ &+\sum_{j=0}^{N}u_{2j}\Big(\frac{x_1-X}{x_1-x_2}\Big)^{j}\Big(\frac{X-x_3}{x_2-x_3}\Big)^{3N+1-j}\Big(\frac{y_1-Y}{y_1-y_2}\Big).\end{aligned}$$

The correct vanishing at (x_1, y_1) and (x_3, y_3) is plain, and our construction was to guarantee that $P(X,Y) = O((X - x_2)^{2N+1}) + O(Y - y_2)$. Incidentally, the qualification 'essentially' is there because if P is to have coefficients in the base field it may be necessary to take an appropriate trace.

All this is a little better than it may look, because one sees that as $N_2 \to \infty$ an appropriate auxiliary polynomial of bidegree $((3N + 1)N_2,\, N_2)$ is given by $(P(X,Y))^{N_2}$. Specifically, this is the case where $\triangle$ is defined by $\theta_2 = t$, least favourable for applications. This is essentially the construction first essayed by Thue, before he instituted the dramatic innovation now known as Siegel's lemma, whereby one shows the existence of the auxiliary function rather than displaying it explicitly. Nonetheless, the Thue construction does not guarantee the existence of an anchor pair in every cubic extension. Whether that can be attained by the present methods remains unclear. It seemed worthwhile to see whether we could learn anything about the exact form of the auxiliary polynomial in cases different from $\theta_2 = t$.

5. SOME EXTRAORDINARY DETERMINANTS

Set $N_1 = 3N$. In the case $N_2 = 0$ our problem is to find a polynomial of degree $3N$ that vanishes N times at each of the points 0, 1 and ∞. That is trivial. The required polynomial is $X^N(X - 1)^N$, with coefficients of height $O(2^N)$.

In general, let N_1 and N_2 be positive integers with $N_1 \geq N_2$. In this context a 'triangle' is a sequence

$$\triangle = \{\, N_1 \geq l_0 > l_1 > \cdots > l_M = \cdots = l_{N_2} = 0\,\}$$

of integers l_i with sum $\sum l_i$ satisfying $3\sum l_i < (N_1 + 1)(N_2 + 1)$; but not *too much* less than. Our problem is to find a polynomial $P(X,Y) = \sum a_{j_1,j_2}X^{j_1}Y^{j_2}$ of bidegree (N_1, N_2) vanishing at $(0,0)$, $(1,1)$, and (∞,∞) 'on' the triangle; that is, so that $P^{(i_1,i_2)}(1,1) = 0$ for all derivatives with $i_1 < l_{i_2}$. The conditions at $(0,0)$ and (∞,∞) entail respectively that the coefficients a_{j_1,j_2} vanish for $j_1 < l_{j_2}$ and for $j_1 > N_1 - l_{N_2-j_2}$; that is, all but the coefficients a_{j_1,j_2} in the lozenge $\lozenge$ defined by $l_{j_2} \leq j_1 \leq N_1 - l_{N_2-j_2}$ must be zero. Once $N_2 > 0$ one cannot just write down a solution. It seems necessary to study the maximal minors of the matrix with entries

$$\binom{j_1}{i_1}\binom{j_2}{i_2},$$

with rows indexed by $(i_1, i_2) \in \triangle$ and columns by $(j_1, j_2) \in \lozenge$. These determinants are of height $O(c^{N_1^2N_2})$, but one can hope that they share so large a common factor that one obtains coefficients a_{j_1,j_2} of height just $O(c^{N_1})$. This is the case for the degenerate examples $N_2 = 0$ and $N_2 = 1$, when the entries are each just a single binomial coefficient. We carried out a number of sporadic calculations to convince ourselves that we were not being exaggeratedly hopeful. We did find very substantial common factors compatible with our hopes. But, of course, such computations cannot distinguish between an exponent N_1 and, say, $N_1 \log N_1$.

The first step in an exact construction seemed to be to understand the 'complete' case for which

$|\triangle| = |\triangledown|$, that is, $3|\triangle| = (N_1+1)(N_2+1)$. There is then just one determinant, Δ.

In 1986 we computed many examples of the cases

$$\triangle = \{2N_2 l, 2(N_2-1)l, \dots, 2l, 0\}$$

with $N_1 = 3lN_2 - 1$, and found that apparently

$$\begin{aligned}\Delta &= \pm\left(\frac{\prod_{k=0}^{l-1} k! \prod_{k=2l}^{3l-1} k!}{\left(\prod_{k=l}^{2l-1} k!\right)^2}\right)^{\binom{N_2+2}{3}} \\ &= \pm\left(\frac{(l!!)^3 3l!!}{(2l!!)^3}\right)^{\binom{N_2+2}{3}}.\end{aligned}$$

Here $l!!$ is a convenient shorthand for $\prod_{k=0}^{l-1} k!$. The remarks of Section 4 allow one to verify our computation in the case $N_2 = 1$, but for the rest these computational results remain an enigma. In particular, we do not even have an arm-waving argument to explain why one should get powers of the case $N_2 = 1$; or, even given that, why the exponent should be $\frac{1}{6}N_2(N_2+1)(N_2+2)$. Brute attempts to evaluate the determinants have not given any real insight into why the evaluation should hold.

Nevertheless, the computation does support our guess that only small primes—those not exceeding N_1—would appear in the evaluation. This guess is based on the following remark. The linear system of equations we want to solve corresponds to constructing a curve of bidegree (N_1, N_2) on the quadric $\mathbb{P}^1 \times \mathbb{P}^1$, with certain specific singularities at $(0,0)$, $(1,1)$, (∞,∞). In the 'complete' case, $\Delta \neq 0$ means that the number of conditions imposed by the singularities on the curve equals the number of parameters at our disposal, and there is no such curve. Suppose, however, that p is a prime that divides Δ. Then Δ vanishes modulo p, and there is such a curve defined over the finite field with p elements. It is now tempting to try to lift this curve to a curve in characteristic 0. Locally, the singularities are defined by derivations of order not exceeding N_1 (say $N_1 > N_2$), and there is no problem in extending these derivations if $p > N_1$. The points 0, 1, ∞ remain distinct modulo p for any p, and $\mathbb{P}^1 \setminus \{0,1,\infty\}$ has no moduli. Thus it is conceivable that if $p > N_1$ there should be no obstructions to lifting a solution mod p to one in characteristic 0, and this would mean that any prime divisor p of Δ satisfies $p \leq N_1$. It is also conceivable that congruences modulo powers of p may be related to lifting solutions over $\mathbb{Z}/p^n\mathbb{Z}$ to the Witt ring, but in any case we could not really understand why *three* singularities should be independent. On the other hand, the conjecture that only small primes appear in a factorization of Δ led quite naturally to the conjecture that combinatorial identities lie behind the problem.

Recently, the second author evaluated an extensive collection of Δ's in the case $N_2 = 2$, with $\triangle = \{bl, cl, 0\}$; here $\gcd(b,c) = 1$. It turns out to be convenient to set $l = 2a$. On the basis of those computations one may confidently conjecture that:

Conjecture. (i) *$\Delta(bl, cl, 0) = 0$ if b is even and both c and l are odd*;

(ii) *if any of these conditions does not hold and $2c < b$, then*

$$\pm\Delta(bl, cl, 0) = \frac{[(2b-c)a]^2[(2b-4c)a][(b+c)a]^2[(b-c)a]^6[ca]^6}{[2(b-c)a]^3[ba]^6[(b-2c)a]^2[2ca]^3},$$

where $[s] = s!! = \prod_{k=0}^{s-1} k!$ if s is an integer, and $[s]^2 = (s+\frac{1}{2})!!(s-\frac{1}{2})!!$ if $s \in \mathbb{Z} + \frac{1}{2}$;

(iii) *if any of the conditions in* (i) *does not hold and $2c \geq b$, then*

$$\pm\Delta(bl, cl, 0) = 2^{2(b-2c)a}\Delta(bl, (b-c)l, 0).$$

It appears to be an interesting, if not easy, undertaking to elucidate this problem, with its relations to algebraic geometry, hypergeometric functions, combinatorics and diophantine approximation. It may be that when we are eventually able to guess such a result for arbitrary N_2 we will have some hint as to how to prove these evaluations. That may then allow us to evaluate the cofactors and thence to construct the polynomials.

REFERENCES

[Beukers et al 1991] F. Beukers, A. J. van der Poorten and R. I. Yager, "Dyson's lemma without tears", *Indag. Math.* (new series) **2** (1991), 1–8.

[Bombieri 1982] E. Bombieri, "On the Thue–Siegel theorem", *Acta Math.* **148** (1982), 255–296.

[Bombieri 1993] E. Bombieri, "Effective diophantine approximation on $\mathbb{G}_m$", *Ann. Scuola Norm. Sup. Pisa*, Cl. Sci. (4) **20** (1993), 61–89.

[Bombieri and Mueller 1983] E. Bombieri and J. Mueller, "On effective measures of irrationality for $\sqrt[r]{a/b}$ and related numbers", *J. reine angew. Math.* **342** (1983), 173–196.

[Bombieri and van der Poorten 1988] E. Bombieri and A. J. van der Poorten, "Some quantitative results related to Roth's theorem", *J. Austral. Math. Soc.* A**45** (1988), 233–248; corrigenda, **48** (1990), 154–155.

[Bombieri and Vaaler 1983] E. Bombieri and J. D. Vaaler, "On Siegel's lemma", *Invent. Math.* **73** (1983), 11–32; addendum, **75** (1984), 377.

[Esnault and Viehweg 1984] H. Esnault and E. Viehweg, "Dyson's lemma for polynomials in several variables (and the theorem of Roth)", *Invent. Math.* **78** (1984), 445–490.

[Muir 1911] T. Muir, *Theory of determinants in the historical order of development*, Macmillan, London, 1911; reprinted by Dover, New York, 1960.

[Muir 1933] T. Muir, *A treatise on the theory of determinants*, Longmans, New York, 1933.

[van der Poorten 1976] A. J. van der Poorten, "Some determinants that should be better known", *J. Austral. Math. Soc.* A**21** (1976), 278–288.

[van der Poorten 1991] A. J. van der Poorten, "Obituary: Kurt Mahler 1903–1988", *J. Austral. Math. Soc.* **51** (1991), 343–380.

[Viola 1985] C. Viola, "On Dyson's lemma", *Ann. Scuola Norm. Sup. Pisa*, Cl. Sci. (4) **12** (1985), 105–135.

[Viola 1992] C. Viola, "On Dyson's lemma in several variables", in *Approximations diophantiennes et nombres transcendants*, Luminy 1990, de Gruyter, Berlin, 1992, 281–284.

[Vojta 1989] P. Vojta, "Dyson's lemma for products of two curves of arbitrary genus", *Invent. Math.* **98** (1989), 107–113.

[Vojta 1991] P. Vojta, "Siegel's theorem in the compact case", *Ann. Math.* **133** (1991), 509–548.

Enrico Bombieri, School of Mathematics, Institute for Advanced Study, Princeton, NJ, 08540, USA (eb@math.ias.edu)

David C. Hunt, School of Mathematics, University of New South Wales, Kensington, NSW 2031, Australia (dave@hydra.maths.unsw.edu)

Alfred J. van der Poorten, Centre for Number Theory Research, Macquarie University, Sydney, NSW 2109, Australia (alf@mpce.mq.edu.au)

Received January 31, 1995; accepted May 8