Thue equations with composite fields

by
Yuri Bilu* (Zürich)
and
Guillaume Hanrot† (Talence)

April 16, 1997

Abstract

We consider the Thue equation $F(x,y)=a$, where F is an irreducible form of degree $n \geq 3$. We describe a method of resolution which takes advantage of the fact that the number field generated by a root of $F(1,y)$ has small subfields. We illustrate this method by solving several real cyclotomic equations of degrees as large as 2505.

Résumé

Considérons l'équation de Thue $F(x,y)=a$, avec F une forme irréductible homogène de degré $n \geq 3$. Nous décrivons une méthode de résolution permettant de tirer profit de l'existence de petits sous-corps du corps de nombres engendré par une racine de $F(1,y)$. À titre d'illustration, nous résolvons l'équation cyclotomique réels en des degrés allant jusqu'à 2505.

1 Introduction

The history of numerical solution of Diophantine equations begun in 1969, when Baker and Davenport [1] solved completely a system of two Pell equations. They used the well-known fact that every "large" solution gives rise to a "very small" value of a linear form $\Lambda(b_1,b_2)=\log\alpha_0+b_1\log\alpha_1+b_2\log\alpha_2$ (where α_0, α_1 and α_2 are explicitly given algebraic numbers) at an integral point (b_1,b_2). Using Baker's theory of logarithmic forms, they obtained a huge (around 10^{400}) upper bound for $\max(|b_1|,|b_2|)$. After this, expanding $\log\alpha_2/\log\alpha_1$ into continued fraction, they showed that $|\Lambda(b_1,b_2)|$ cannot be too small when b_1 and b_2 run the integers below the huge bound. Therefore the system cannot have "large" solutions, while "small" solutions can be easily enumerated.

This idea was developed in various directions by Pethö, Tzanakis, de Weger, and many other authors. The subject became especially popular when Lenstra, Lenstra and

*Partially supported by SFB 343 "Diskrete Strukturen in der Mathematik", Bielefeld
†Partially supported by CNRS GDR AMI
1991 *Mathematics Subject Classification* Primary: 11Y50, Secondary: 11B37.

Lovasz [9] suggested a polynomially quick algorithm for finding an almost shortest vector in a lattice (referred to as LLL-algorithm in the sequel). The LLL-algorithm made it possible to extend the idea of Baker and Davenport to logarithmic forms in three or more variables, when continued fractions are not efficient anymore. See [16, 10] for a detailed description of the methods, history of the subject and extensive bibliography up to 1989.

In [4] we showed that one can solve Diophantine equations of Thue using only continued fractions (as Baker and Davenport did), and without involving the LLL-algorithm. This allowed us to solve completely Thue equations of rather high degree. In [5] we extended our method to superelliptic Diophantine equations (see also [3]).

In this paper we show that the method of [4] becomes especially efficient if the number field related to the Thue equation contains a small subfield of degree at least 3 over $\mathbb{Q}$. We shall see that in this case one has to deal mainly with the subfield rather than with the whole field. We were motivated by the fact that such equations often occur in practice, for instance in the classical problem of primitive divisors [15, 18].

Using our method we managed to solve many totally real Thue equations of extremely high degree (up to 2505). See Section 7 for the details.

Acknowledgments. We would like to thank Benne de Weger for useful discussions and suggestions.

2 Notations

We consider the Thue equation

$$F(x,y) = \mathcal{N}_{\mathbb{L}/\mathbb{Q}}(y-\alpha x) = a, \tag{1}$$

where $a = a_1/a_2$ is a rational number, α an algebraic number of degree $n \geq 3$, and $\mathbb{L} = \mathbb{Q}(\alpha)$. We put

$$f(y) = F(1,y) = \mathcal{N}_{\mathbb{L}/\mathbb{Q}}(y-\alpha).$$

We shall assume that the field $\mathbb{L}$ has a "small" subfield $\mathbb{K}$, of degree $m \geq 3$.

Let $\mathbb{K}$ have s real and $2t$ complex conjugate embeddings, where $s + 2t = m$. We number the embeddings $\sigma_1, \ldots, \sigma_m$ so that $\sigma_1, \ldots, \sigma_s$ are the real embeddings, and

$$\sigma_{s+i} = (\text{complex conjugation}) \circ \sigma_{s+i+t} \quad (1 \leq i \leq t). \tag{2}$$

We write $\mathbb{K}_i = \sigma_i(\mathbb{K})$. We shall assume that $s \geq 1$; in particular, $\mathbb{K}$ has no roots of unity distinct from ± 1 (in the case $s = 0$ the equation is trivial; see [4, Section 2], for instance).

Put $l = n/m$ and fix an ordering $\alpha_{11} = \alpha, \alpha_{12}, \alpha_{13}, \ldots, \alpha_{ml}$ of the conjugates of α over $\mathbb{Q}$ so that for a fixed i the numbers $\alpha_{i1}, \ldots, \alpha_{il}$ are conjugate over $\mathbb{K}_i$.

We use $O_1(\ldots)$ as a quantitative version of the usual $O(\ldots)$:

$$A = O_1(B)$$

means $|A| \leq B$.

For practical implementation of the method, one should be able to perform the following operations in the number field $\mathbb{K}$:

(U) find a system of fundamental units;

(N) given a fractional ideal I of the field $\mathbb{K}$, find a complete system of non-associate solutions of the norm equation

$$N_{\mathbb{K}/\mathbb{Q}}(\beta) = a, \quad \beta \in I\,. \tag{3}$$

(The units of $\mathbb{K}$ act on the solutions of (3) by multiplication. By *a complete system of non-associate solutions* of the equation (3) we mean any set of representatives of this action.) It is well known that any complete system of non-associate solutions is finite, and that problems (U) and (N) are effectively soluble [6, Ch. 2]. However, finding efficient algorithms for the practical resolution of these problems proved to be difficult, especially for fields of high degree. This is the main reason why the method is efficient only when the field $\mathbb{K}$ is "not very big". We do not discuss this problem, referring to [7, 13, 12].

The purpose of the present paper is to show that the Thue equation (1) can be practically solved in reasonable time as soon as the problem (U) is solved and the problem (N) is solved with $I = \mathcal{N}_{\mathbb{L}/\mathbb{K}}\Big((1,\alpha)\Big)$.

Thus, fix once and for all a system $\eta_1, \ldots, \eta_r$ of basic units of the field $\mathbb{K}$, where $r = s + t - 1$, and a complete system M of non-associate solutions of (3). In the important particular case when $|a| = 1$ and α is an algebraic integer, we have $\mathrm{M} = \{1\}$.

Since $\mathbb{K}$ has no root of unity except ± 1, for any solution $\beta \in I$ of the equation (3) there exist $\mu \in \pm\mathrm{M}$ and $b_1, \ldots, b_r \in \mathbb{Z}$ such that $\beta = \mu\eta_1^{b_1}\cdots\eta_r^{b_r}$. Here $\pm\mathrm{M} = \{\pm\mu\colon \mu \in \mathrm{M}\}$.

3 General background

3.1 The numbers φ_i

Fix a solution $(x,y) \in \mathbb{Z}^2$ of the equation (1).

Proposition 3.1.1 *Put*

$$\begin{aligned}
X_0 &= \begin{cases} \left(\dfrac{2^{n-1}|a|}{\min\limits_{\alpha_{ik}\notin\mathbb{R}} |f'(\alpha_{ik})| \cdot \min\limits_{\alpha_{ik}\notin\mathbb{R}} |\operatorname{Im}\alpha_{ik}|}\right)^{1/n}, & \textit{if } \mathbb{L} \textit{ is not totally real,}\\ 1, & \textit{if } \mathbb{L} \textit{ is totally real,}\end{cases}\\
c_1 &= \frac{2^{n-1}|a|}{\min\limits_{(i,k)}|f'(\alpha_{ik})|}, \qquad c_2 = \min_{(i,k)\neq(i',k')}|\alpha_{ik}-\alpha_{i'k'}|\,,\\
c_3 &= 1.39c_1c_2^{-1}, \qquad X_1 = \max\Big(X_0,\ \big(2c_1c_2^{-1}\big)^{1/n}\Big)
\end{aligned}$$

(in the definition of X_0 both the minima run over the non-real conjugates of α). Let (x,y) be an integer solution of (1).

(i) If $|x| > X_0$ then for some real conjugate $\alpha_{i_0k_0}$ we have

$$|y/x - \alpha_{i_0k_0}| \le c_1|x|^{-n}. \tag{4}$$

(ii) If $|x| > X_1$ then

$$y - \alpha_{ik}x = (\alpha_{i_0k_0} - \alpha_{ik})\, x e^{O_1\left(c_3|x|^{-n}\right)} \qquad \Big((i,k) \ne (i_0,k_0)\Big). \tag{5}$$

Proof For (i) see [16, Lemma 1.1]. To prove (ii), write

$$y - \alpha_{ik}x = (\alpha_{i_0k_0} - \alpha_{ik})\, x \left(1 + \frac{y/x - \alpha_{i_0k_0}}{\alpha_{i_0k_0} - \alpha_{ik}}\right). \tag{6}$$

Since $|x| \ge X_1$, we have

$$\left|\frac{y/x - \alpha_{i_0k_0}}{\alpha_{i_0k_0} - \alpha_{ik}}\right| \le \frac{1}{2}.$$

But $1 + z = e^{O_1(1.39|z|)}$ if the complex number z satisfies $|z| \le 1/2$ (see [16, p. 106]). Therefore (5) is a consequence of (6).

In concrete examples the constant X_1 is very small, and solutions satisfying $|x| \le X_1$ can be easily enumerated. From now on, we assume that $|x| > X_1$, so that (4) and (5) holds for some (i_0, k_0). Fix this (i_0, k_0) and put

$$\begin{aligned} \varphi_i &= \prod_{k=1}^{l}(y - \alpha_{ik}x) \qquad (1 \le i \le m), \\ \psi_i &= \prod_{k=1}^{l}(\alpha_{i_0k_0} - \alpha_{ik}) \qquad (1 \le i \le m, \quad i \ne i_0), \end{aligned}$$

Then $\varphi := \varphi_1 = \mathcal{N}_{\mathbb{L}/\mathbb{K}}(y - \alpha x)$ and $\varphi_i = \sigma_i(\varphi)$. Also, immediately from (5) we deduce that

$$\varphi_i = \psi_i x^l e^{O_1\left(c_4|x|^{-n}\right)} \quad (i \ne i_0), \tag{7}$$

where $c_4 = lc_3$. Since $\varphi_1 \cdots \varphi_m = a$, we also obtain

$$\varphi_{i_0} = \psi_{i_0} x^{(1-m)l} e^{O_1\left(c_5|x|^{-n}\right)}, \tag{8}$$

where $\psi_{i_0} = a\left(\prod_{i \ne i_0}\psi_i\right)^{-1}$ and $c_5 = (m-1)c_4$. We unify (7) and (8) in

$$\varphi_i = \psi_i x^{\rho_i} e^{O_1\left(c_5|x|^{-n}\right)} \quad (1 \le i \le m), \tag{9}$$

where

$$\rho_i = \begin{cases} l, & i \ne i_0, \\ (1-m)l, & i = i_0. \end{cases}$$

We conclude this subsection with the following important property of the numbers φ_i.

Proposition 3.1.2 *Among the $m-1$ numbers*

$$\varphi_i/\psi_i \quad (i \ne i_0), \tag{10}$$

two at least are distinct.

Proof Assume that the numbers (10) are all equal, and write this as

$$P_i(\theta)/P_i(\alpha_{i_0k_0}) = P_{i'}(\theta)/P_{i'}(\alpha_{i_0k_0}) \quad (i, i' \neq i_0), \tag{11}$$

where $\theta = y/x$ and $P_i(T) = \prod_{k=1}^{l}(T - \alpha_{ik})$. Note that the polynomials P_i are pairwise distinct.

Let $\sigma_{i_01}, \ldots, \sigma_{i_0l} : \mathbb{L} \to \mathbb{C}$ be the extensions of σ_{i_0} to $\mathbb{L}$ defined by $\sigma_{i_0k}(\alpha) = \alpha_{i_0k}$. Then for any $k \in \{1, \ldots, l\}$, the map $\tau_k := \sigma_{i_0k}\sigma_{i_0k_0}^{-1}$ permutes the polynomials P_i, where $i \neq i_0$, and stabilizes θ, a rational number. Hence, acting on (11) by $\tau_1, \ldots, \tau_l$, we obtain

$$P_i(\theta)/P_i(\alpha_{i_0k}) = P_{i'}(\theta)/P_{i'}(\alpha_{i_0k}) \quad (i, i' \neq i_0, \quad 1 \leq k \leq l).$$

Fix distinct $i, i' \in \{1, \ldots, m\} \setminus \{i_0\}$ (this is possible since $m \geq 3$). Then

$$P_i(\theta)/P_{i'}(\theta) = P_i(\alpha_{i_0k})/P_{i'}(\alpha_{i_0k}) \quad (1 \leq k \leq l).$$

Put $\phi = P_i(\theta)/P_{i'}(\theta)$. Then the polynomial $P_i(T) - \phi P_{i'}(T)$ has $l+1$ distinct roots $\theta, \alpha_{i_01}, \ldots, \alpha_{i_0l}$. Since its degree does not exceed l, it is identically zero. Since its leading coefficient is $1 - \phi$, we have $\phi = 1$. Thus, $P_i = P_{i'}$, a contradiction. The proposition is proved.

3.2 The numbers b_i

Since $\mathcal{N}_{\mathbb{K}/\mathbb{Q}}(\varphi) = a$, we have

$$\varphi = \mu\eta_1^{b_1} \cdots \eta_r^{b_r}, \tag{12}$$

where $\mu \in \pm\mathrm{M}$ and $b_1, \ldots, b_r \in \mathbb{Z}$. Put $\eta_{ij} = \sigma_i(\eta_j)$ and $\mu_i = \sigma_i(\mu)$, and let $A = [a_{ij}]_{1 \leq i,j \leq r}$ be the inverse of the matrix

$$[\log|\eta_{ij}|]_{1 \leq i,j \leq r}\,. \tag{13}$$

(The matrix (13) is non-degenerate, because its determinant is $\pm\min(1, 2^{1-t})$ times the regulator of the field $\mathbb{K}$.) Since

$$\log|\varphi_i| = \log|\mu_i| + b_1\log|\eta_{i1}| + \cdots + b_r\log|\eta_{ir}| \quad (i \neq i_0), \tag{14}$$

it follows from (12) that

$$b_i = \sum_{j=1}^{r} a_{ij}\log|\varphi_j/\mu_j| = \delta_i\log|x| + \lambda_i + O_1\left(c_6|x|^{-n}\right) \quad (1 \leq i \leq r), \tag{15}$$

where

$$\delta_i = \sum_{j=1}^{r}\rho_j a_{ij}, \quad \lambda_i = \sum_{j=1}^{r} a_{ij}\log|\psi_j/\mu_j|\,, \quad c_6 = c_5\sum_{j=1}^{r}|a_{ij}|. \tag{16}$$

In particular, we obtain the following.

Proposition 3.2.1 *If* $|x| \geq X_2 := \max\left(X_1, (2 \cdot 10^{10}c_6)^{1/n}\right)$ *then*

$$B := \max(|b_1|, \ldots, |b_r|) \leq c_7\log|x| + c_8, \tag{17}$$

where

$$c_7 = \max_{1 \leq i \leq r}|\delta_i|, \quad c_8 = \max_{1 \leq i \leq r}|\lambda_i| + 10^{-10}.$$

3.3 A large upper bound for B

In this subsection we obtain a huge upper bound for B using Baker's theory. We apply a result of Baker and Wüstholz [2, p. 20], formulating it in a form convenient for the present paper.

Theorem 3.3.1 (Baker–Wüstholz) *Let $\beta_0, \ldots, \beta_r$ be complex algebraic numbers distinct from 0 and 1, and $b_1, \ldots, b_{r+1}$ rational integers. Also, let*

$$\begin{aligned} d &\geq [\mathbb{Q}(\beta_0, \ldots, \beta_r) : \mathbb{Q}], && (18)\\ h_i &\geq \max\left(h(\beta_i), d^{-1}|\log \beta_i|, d^{-1}\right) \qquad (0 \leq i \leq r), && (19) \end{aligned}$$

where $h(\ldots)$ is the absolute logarithmic height. Then either

$$\Lambda := \log \beta_0 + b_1 \log \beta_1 + \cdots + b_r \log \beta_r + b_{r+1}\pi i = 0, \tag{20}$$

or

$$|\Lambda| \geq \exp(-c_9 \log B'). \tag{21}$$

Here $B' = \max(|b_1|, \ldots, |b_r|, |b_{r+1}|, e)$, and

$$c_9 = 18\pi \cdot 32^{r+4}(r+3)!(r+2)^{r+3}d^{r+3}\log(2d(r+2))h_0 \cdots h_r.$$

Remark 3.3.2 The parameters n, $h'(\alpha_1), \ldots, h'(\alpha_n)$, $h'(L)$ of the original theorem in [2] respectively correspond to $r+2$, $h_0, \ldots, h_r$, π/d, $\log B''$ in Theorem 3.3.1.

We have slightly modified the statement in [2], to allow inequalities in (18) and (19). It is often much easier (and quicker) to find an upper bound for the degree of a number field or for the height of an algebraic number, than to compute them exactly.

The following lemma is the case $h = 1$ of [11, Lemma 2.2]

Lemma 3.3.3 *Let z and C_1 be positive real numbers and C_2 an arbitrary real number. If $z \leq C_1 \log z + C_2$, then $z \leq 2(C_1 \log C_1 + C_2)$.*

By Proposition 3.1.2, there exist $i_1, i_2 \in \{1, \ldots, m\} \setminus \{i_0\}$ such that

$$\frac{\psi_{i_2}\varphi_{i_1}}{\psi_{i_1}\varphi_{i_2}} \neq 1. \tag{22}$$

On the other hand, as follows from (7),

$$\frac{\psi_{i_2}\varphi_{i_1}}{\psi_{i_1}\varphi_{i_2}} = e^{O_1\left(2c_4|x|^{-n}\right)}. \tag{23}$$

Combining (22) and (23) with (12), we obtain

$$1 \neq \beta_0\beta_1^{b_1}\cdots\beta_r^{b_r} = e^{O_1(2c_4|x|^{-n})}, \tag{24}$$

where

$$\beta_0 = \frac{\psi_{i_2}}{\psi_{i_1}}\frac{\sigma_{i_1}(\mu)}{\sigma_{i_2}(\mu)}, \quad \beta_j = \frac{\sigma_{i_1}(\eta_j)}{\sigma_{i_2}(\eta_j)} \quad (1 \leq j \leq r).$$

To compute the constant c_9 in this setting, we need to estimate $h(\beta_0), \dots,\ h(\beta_r)$. This can be done using the well-known inequalities $h(a \pm b) \le h(a) + h(b) + \log 2$ and $h(ab^{\pm 1}) \le h(a) + h(b)$;

$$\begin{aligned} h(\beta_0) &\le 2h(\mu) + 2l(2h(\alpha) + \log 2), \\ h(\beta_j) &\le 2h(\eta_j) \quad (1 \le j \le r). \end{aligned}$$

Denote by log the principal branch of the complex logarithm, that is $-\pi < \mathrm{Im}\, \log z \le \pi$. Then

$$0 < |\log \beta_0 + b_1 \log \beta_1 + \cdots + b_r \log \beta_r + b_{r+1} \cdot \pi i| \le 2c_4 |x|^{-n}, \tag{25}$$

for some $b_{r+1} \in \mathbb{Z}$. Comparing the imaginary parts, we obtain

$$|b_{r+1}| \le 1 + |b_1| + \cdots + |b_r| + 2\pi^{-1} c_4 X_1^{-n} \le 1 + 0.45l + Br, \tag{26}$$

because by the definition of c_4 and X_1 we have

$$2\pi^{-1} c_4 X_1^{-n} \le 2 \cdot 1.39 l (2\pi)^{-1} < 0.45 l.$$

Therefore

$$B' \le \max\left(e, c_{10} + c_{11} \log |x|\right), \tag{27}$$

where $c_{10} = rc_8 + 1 + 0.45l$ and $c_{11} = rc_7$.

As follows from (25), (21) and (27), either we have $B' = e$, or

$$\exp\left(c_{12}\left(c_{13} - B'\right)\right) \ge \exp\left(-c_9 \log B'\right),$$

where $c_{12} = nc_{11}^{-1}$ and $c_{13} = c_{10} + c_{12}^{-1} \log(2c_4)$. Hence either we have $B' = e$, or

$$B' \le c_{12}^{-1} c_9 \log B' + c_{13}.$$

In view of Lemma 3.3.3, this implies that

$$B \le B' \le B_0 := \max\left(e, 2\left(c_{12}^{-1} c_9 \log\left(c_{12}^{-1} c_9\right) + c_{13}\right)\right).$$

4 Reduction of Baker's bound

In practice, the value of B_0 is too large for directly enumerating all possible $(b_1, \dots, b_r)$. However, B_0 may be significantly reduced using continued fractions. As it was already pointed in the introduction, a method of reduction was suggested by Baker and Davenport [1]. This method was developed by Tzanakis and de Weger [16], Pethö [10] and others. In this paper we use another modification of the Baker-Davenport method, suggested in [3, 4, 5].

The algorithm of reduction depends on whether $r = 1$ or $r \ge 2$.

4.1 The case $r \ge 2$

Define i_1 by the condition

$$|\delta_{i_1}| = \max_{1 \le i \le r} |\delta_i| = c_7. \tag{28}$$

(Clearly, $\delta_{i_1} \neq 0$, because the matrix A is non-degenerate). Further, put

$$\overline{\delta}_i = \delta_{i_1}^{-1}\delta_i, \quad \overline{\lambda}_i = \delta_{i_1}^{-1}\left(\delta_i\lambda_{i_1} - \delta_{i_1}\lambda_i\right) \quad (1 \leq 1 \leq r)$$

By the choice of i_1 we have $|\overline{\delta}_i| \leq 1$ for every i. Using (15), we obtain the following:

$$\begin{aligned} b_i &= \delta_i \log|x| + \lambda_i + O_1(c_6|x|^{-n}) \\ &= \delta_i\delta_{i_1}^{-1}\left(b_{i-1} - \lambda_{i_1} + O_1(c_6|x|^{-n})\right) + \lambda_i + O_1(c_6|x|^{-n}) \\ &= \overline{\delta}_i b_{i_1} - \overline{\lambda}_i + O_1\left((1+|\overline{\delta}_i|)c_6|x|^{-n}\right) \\ &= \overline{\delta}_i b_{i_1} - \overline{\lambda}_i + O_1\left(2c_6|x|^{-n}\right). \end{aligned} \tag{29}$$

Fix $i_2 \neq i_1$ and put $\overline{\delta} = \overline{\delta}_{i_2}$ and $\overline{\lambda} = \overline{\lambda}_{i_2}$. Then we can rewrite (29) as

$$\left|b_{i_2} - \overline{\delta} b_{i_1} + \overline{\lambda}\right| \leq 2c_6|x|^{-n}. \tag{30}$$

Let $\kappa > 2$ be a not very large number (at the end of this subsection we discuss the practical choice of κ). By the theorem of Dirichlet, there exists a positive integer $q \leq \kappa B_0$ such that

$$\|q\overline{\delta}\| \leq (\kappa B_0)^{-1}, \tag{31}$$

where $\|\cdot\|$ is the distance to the nearest integer. In practice q can be quickly found from the continuous fraction expansion of $\overline{\delta}$. Multiplying (30) by q, we obtain

$$\left\|\pm b_{i_1}\|q\overline{\delta}\| + q\overline{\lambda}\right\| \leq 2c_6\kappa B_0|x|^{-n}, \tag{32}$$

where "$\pm$" should be "$+$" if $q\delta$ is smaller than the nearest integer and "$-$" otherwise.

It follows from (31) that $|b_{i_1}| \cdot \|q\overline{\delta}\| \leq \kappa^{-1}$. Therefore (32) implies that

$$\|q\overline{\lambda}\| - \kappa^{-1} \leq 2c_6\kappa B_0|x|^{-n}. \tag{33}$$

If $\|q\overline{\lambda}\| > \kappa^{-1}$, which is heuristically plausible when κ is large enough, then

$$|x| \leq \left(\frac{2c_6\kappa B_0}{\|q\overline{\lambda}\| - \kappa^{-1}}\right)^{1/n}. \tag{34}$$

Together with (17) this yields a new estimate for B:

$$B \leq c_{15}\log B_0 + \log\frac{c_{14}\kappa}{\|q\overline{\lambda}\| - \kappa^{-1}}, \tag{35}$$

where $c_{14} = 2c_6e^{c_8}$ and $c_{15} = c_7/n$. In particular, when $\|q\overline{\lambda}\| \geq 2\kappa^{-1}$, we have an estimate

$$B \leq c_{15}\log B_0 + \log\left(c_{14}\kappa^2\right) \tag{36}$$

(compare this with the lemma from [1, Section 3]).

We took as a starting value $\kappa = 10$, and tried the first reduction. If $\|q\lambda\| < 2\kappa^{-1}$, then we changed κ by 10κ and repeated the process.

The reduced bound for B can be reduced again, using the same procedure, etc. Since in the case of Thue equation the constant c_{15} is usually rather small, the reduction is very efficient.

4.2 The case $r = 1$

In this case the method of reduction is more or less the same as in the Tzanakis-de Weger paper [16]. We include some details for the sake of completeness.

Since $\mathbb{K}$ has a real embedding, we have $m = 3$, and $\mathbb{K}$ has one real embedding σ_1 and a pair of complex conjugate embeddings σ_2, σ_3. We have $i_0 = 1$ and $\{i_1, i_2\} = \{2, 3\}$; for instance, let it be $i_1 = 2$ and $i_2 = 3$.

Now (24) can be rewritten as

$$1 \neq \beta_0 \beta_1^{b_1} = e^{O_1(2c_4|x|^{-n})}, \tag{37}$$

where

$$\beta_0 = \frac{\psi_3}{\psi_2}\frac{\sigma_2(\mu)}{\sigma_3(\mu)}, \quad \beta_1 = \frac{\sigma_2(\eta_1)}{\sigma_3(\eta_1)}.$$

Since σ_2 and σ_3 are complex conjugate, one has $|\beta_0| = |\beta_1| = 1$. Also, β_1 is not a root of unity; otherwise, it would be $\sigma_2\left(\eta_1^N\right) = \sigma_3\left(\eta_1^N\right)$ for some positive integer N. It would follow that η_1^N is a Dirichlet unit of the field $\mathbb{Q}$, which means $\eta_1^N = \pm 1$, a contradiction.

Now rewrite (37) as

$$0 < \|\lambda + b_1\delta\| \leq 2c_4|x|^{-n}$$

with $\delta = \arg \beta_1 / 2\pi$ and $\lambda = \arg \beta_0 / 2\pi$, and continue as in the case $r \geq 2$.

4.3 Pathological reduction

In [4, Subsection 4.6] (see also [8]) we described various cases of "pathological" reduction: "semirational" and "totally rational" cases when $r \geq 2$, and multiplicative dependence of β_0 and β_1 when $r = 1$. The method of reduction in the pathological cases is similar to the described above, and even more efficient. Since the "pathologies" occur in practice very seldom, we find it possible to omit in this paper the detailed analysis of them; if needed, it can be copied from [4] with insignificant changes.

5 Enumerating small b_i

Even when the upper bound for B is reduced, enumerating all possible $(b_1, \ldots, b_r)$ can require extensive computations. One can imagine several ways to overcome this difficulty:

- using the continued fraction expansions of α, see [10, 16] for the details;
- sieving modulo several primes, as in [17, 14], for instance;
- using Fincke–Pohst algorithm for finding all short vectors in a lattice, as in [19, 17], for instance.

We use a method suggested in [3], with some modifications introduced in [5].

For $1 \leq i \leq r$ put $b_i' = \overline{\delta}_i b_{i_1} - \overline{\lambda}_i$, where i_1 is defined from (28). Then $|b_i - b_i'| \leq 2c_6|x|^{-n}$. Since $X_2 \geq (2 \cdot 10^{10} c_6)^{1/n}$, we obtain

$$|b_i - b_i'| < 10^{-10} \quad (1 \leq i \leq r) \tag{38}$$

as soon as $|x| > X_2$. In particular,

$$\|b_i'\| < 10^{-10} \quad (1 \leq i \leq r), \tag{39}$$

and b_i is the nearest integer to b_i'.

Now we proceed as follows. Denote by B_0' the reduced bound for B. For every integer b such that $|b| \leq B_0'$, we put $b_{i_1} = b$, and compute the real numbers b_i' as above. Then for every i we verify whether $\|b_j'\| < 10^{-10}$ or not. This condition trivially holds for $i = i_1$, but for $i \neq i_1$ it need not. *If it is false for at least one i, then there is no solution x with $|x| > X_2$ such that $b_{i_1} = b$*, and we go to the next b.

The heuristic probability that the integer b passes this severe test is $(2 \cdot 10^{-10})^{1-r}$, quite a small number (when $r \geq 2$). For those very few b that survive after the test, we use the second test, based on the following lemma. (We define $z^{1/l}$ by $-\pi/l < \arg z^{1/l} \leq \pi/l$.)

Lemma 5.1 *For $i \neq i_0$ put $\omega_i = (\varphi_i/\psi_i)^{1/l}$. Assume that $|x| > X_3 := \max\left(X_2, (1.3 \cdot 10^{10} c_4)^{1/(n-1)}\right)$. If l is odd then*

$$|x - \omega_i| < 10^{-10}. \tag{40}$$

If l is even then we have either (40) or

$$|x + \omega_i| < 10^{-10}. \tag{41}$$

Proof We shall assume that l is odd; the case of even l is done similarly.

For $|z| \leq 1/2$ we have

$$e^z = 1 + O_1(c_{16} z), \tag{42}$$

where $c_{16} = 2(e^{1/2} - 1) \leq 1.3$. (This follows from the Schwarz lemma, applied to the function $e^z - 1$ in the disc $|z| \leq 1/2$.) On the other hand,

$$\omega_i = x e^{O_1(c_4 |x|^{-n})}, \tag{43}$$

as follows from (7). Now (40) is an immediate consequence of (42) and (43). This proves the lemma.

The second test is performed as follows. Fix $i \neq i_0$ and compute φ_i from (12), where b_i are the nearest integers to b_i'. Having φ_i, one can compute ω_i and check whether $\|\omega_2\| \leq 10^{-10}$ or not. If this fails then we go to the next b. Otherwise, we compute x as the nearest integer to ω_2 and check whether it corresponds to a solution (x, y) of our equation[1]. However, this option never happened in our computations.

Remark 5.2 In the process of reduction one obtains an upper bound not only for B but for x as well, due to (34). Quite often, especially when n is large, this bound does not exceed X_3 (or does exceed X_3 but is still reasonable). In this case enumerating small b_i becomes superfluous.

[1] If l is even then one has also to check the nearest integer to $-\omega_i$.

6 The algorithm

Now we can summarize the contents of the previous sections in a formal algorithm. Before giving it, we notice that the numbers $\overline{\delta}_i$ depend only on i_0, and are independent of μ and k_0. Therefore the reduction and final enumeration can be performed simultaneously for all possible pairs (k_0, μ), when i_0 is fixed. To do this, one must unify the constants depending on k_0 and μ, by putting $\widehat{c}_8 := \max c_8(k_0, \mu)$, and redefining c_{10} by substituting $\widehat{c}_8$ instead of c_8. (This would also affect the definitions of c_{13}, c_{14}, and B_0.)

At the starting point we are given the following data, which will be referred to as "the data":

- approximate values of α and all its conjugates;
- a system of fundamental units of the field $\mathbb{K}$ (for each unit approximate values of all its conjugates being required);
- the set M (again, for every $\mu \in \mathrm{M}$ we have to know approximate values of all its conjugates).

Here "approximate" means, depending on the situation, from fifty to one thousand decimal digits for both the real and imaginary part. If it turns out in course of solution that the precision is not sufficient, then the data should be recomputed with higher precision, and the algorithm re-executed from a suitable point (see Step 7).

Now we are in a position to describe the algorithm.

1. Compute matrix A, with the highest possible precision.
2. Compute constants c_1–c_6 and X_1–X_3 with low precision (two decimal digits OK).
3. Set $I_0 \leftarrow 1$.
4. Set $i_0 \leftarrow I_0$.
5. Compute the numbers $\overline{\delta}_i$, with highest possible precision, and the constants c_7, $\widehat{c}_8$, c_9–c_{15}, and B_0 with low precision.
6. For every pair $(k_0, \mu) \in \{1, \ldots, l\} \times \pm\mathrm{M}$ such that $\alpha_{i_0 k_0} \in \mathbb{R}$ compute the corresponding set of $\overline{\lambda}_i$. If $\alpha_{i_0 k_0} \notin \mathbb{R}$ for all k_0 then go to Step 9.
7. Find a reduced bound for B, as described in Section 4.
   If it turns out that the precision of $\overline{\delta}_i$ is not sufficient, then:
   (a) recompute the data with a suitable precision;
   (b) set $I_0 \leftarrow i_0$;
   (c) go to Step 4.
8. Enumerate small b_i, as described in Section 5.
9. Set $i_0 \leftarrow i_0 + 1$. If $i_0 \leq s$ then go to Step 5.
10. For any $x \in \mathbb{Z}$ such that $|x| \leq X_3$ check whether x corresponds to a solution of (1).
11. Collect all solutions obtained at Steps 8 and 10.
12. End.

7 The real cyclotomic equation

As an example, we consider the *real cyclotomic equation*

$$F_P(x,y) := \prod_{k=1}^{(P-1)/2} \left(y - x \cdot 2\cos\tfrac{2\pi k}{P}\right) = \pm 1,\ \pm P, \tag{44}$$

where $P > 12$ is prime number. This equation occurs in study of primitive divisors of Lucas and Lehmer numbers, see [15, 18].

7.1 The field $\mathbb{K}$

Since the field $\mathbb{L}$ is abelian, for any m dividing $(P-1)/2$ there exists a subfield $\mathbb{K}$ of degree m. Thus, our method would be inefficient only if $n := (P-1)/2$ has no small divisors distinct from 1 and 2, which happens quite seldom.

Thus, put

$$m = \begin{cases} 4, & \text{if} \quad P \equiv 17 \pmod{24}, \\ \text{the least odd prime divisor of } n & \text{otherwise.} \end{cases}$$

Since the group $\mathrm{Gal}(\mathbb{L}/\mathbb{Q})$ is cyclic, there exists a single subfield $\mathbb{K}$ of $\mathbb{L}$ of degree m. The following lemma was used to compute a generator of $\mathbb{K}$ over $\mathbb{Q}$.

Lemma 7.1.1 *Let a be a primitive root modulo P. Then the algebraic integer*

$$\xi_0 = \sum_{k=0}^{n/m} 2\cos\left(\frac{2a^{mk}\pi}{P}\right)$$

generates the field $\mathbb{K}$ over $\mathbb{Q}$. The conjugates of ξ_0 over $\mathbb{Q}$ are the numbers

$$\xi_i = \sum_{k=0}^{n/m} 2\cos\left(\frac{2a^{mk+i}\pi}{P}\right) \qquad (0 \le i \le m-1). \tag{45}$$

Proof One verifies immediately that $\mathrm{Gal}(\mathbb{L}/\mathbb{K})$ stabilizes every ξ_i, and that $\mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ acts on the set $\{\xi_0, \ldots, \xi_{m-1}\}$ transitively. This means that $\xi_0, \ldots, \xi_{m-1} \in \mathbb{K}$ and that ξ_i are pairwise conjugate over $\mathbb{Q}$.

It remains to prove that ξ_0 generates $\mathbb{K}$. We shall use the following general observation.

Let $k \subseteq K \subseteq L$ be a tower of fields of characteristic zero, and assume that $\alpha \in L$ generates L over k. Then the numbers

$$\mathrm{Tr}_{\mathrm{L/K}}(\alpha^j) \qquad (1 \le j \le [L\colon K]) \tag{46}$$

generate K over k.

(To prove this, notice that

(i) the field K is generated over k by the coefficients of the minimal polynomial of α over K, and

(ii) these coefficients can be expressed as polynomials in the numbers (46) with integral coefficients.)

Now it is easy to complete the proof of the lemma. Since $a^n = a^{(P-1)/2} \equiv -1 \pmod{P}$, one can rewrite (45) as

$$\xi_i = \sum_{k=0}^{(P-1)/m} \zeta^{a^{mk+i}} \qquad (0 \le i \le m-1), \tag{47}$$

where ζ is a primitive P-th root of unity. It follows immediately that $\mathrm{Tr}_{\mathbb{Q}(\zeta)/\mathbb{K}}(\zeta^j) \in \{\xi_0, \dots, \xi_{m-1}\}$ for any $j \not\equiv 0 \pmod{P}$. Hence $\mathbb{K} = \mathbb{Q}(\xi_0, \dots, \xi_{m-1})$. Since $\mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ is cyclic, at least one of the numbers ξ_i generates $\mathbb{K}$. Since they are pairwise conjugate, any of them generates $\mathbb{K}$. The lemma is proved.

The set M is $\{1\}$ when the right-hand side is ± 1, and consists of a single element when the right-hand side is $\pm P$ (because P totally ramifies in $\mathbb{K}$).

7.2 Computing the constants, etc.

The following lemmas show how to compute quickly c_1 and c_2.

Lemma 7.2.1 ([8, Lemme 3.6]) *Let l be an integer in $[1, (P-1)/2]$, and*

$$\Psi(l) = \prod_{k=1, k\neq l}^{(P-1)/2} \left|2\cos(2k\pi/P) - 2\cos(2l\pi/P)\right|.$$

Put

$$p_0 = \left\lfloor \mathrm{Acos}\left(\frac{\sqrt{3}}{3}\right) \frac{P}{\pi} \right\rfloor.$$

Then

$$\min_{1 \le l \le (P-1)/2} |\Psi(l)| = \min(|\Psi(p_0)|, |\Psi(p_0+1)|).$$

More precisely, this minimum is equal to $|\Psi(p_0)|$ if and only if

$$\sin(2p_0\pi/P)\sin(p_0\pi/P) \ge \sin(2(p_0+1)\pi/P)\sin((p_0+1)\pi/P).$$

Since $\min \Psi(l) = \min_{i,k} |g'(\alpha_{ik})|$, this lemma allows one to compute c_1 very quickly.

Lemma 7.2.2 ([8, Lemme 3.7]) *We have $c_2 = 4\sin(\pi/P)\sin(2\pi/P)$.*

All the constants are expressed in terms of the roots of $F(1, y)$ rather than its coefficients. Nevertheless, it is useful to have the following "closed" expression for $F(x, y)$, in particular, for enumerating the solutions with $|x| \le X_3$ (Step 10 of the algorithm).

Lemma 7.2.3 ([8, Lemme 3.8]) *Let $\phi_P(x) = \frac{x^P - 1}{x-1}$ be the P-th cyclotomic polynomial. Then*

$$F_P(x, y) = \left(\frac{2x^2}{y + \sqrt{y^2 - 4x^2}}\right)^{(P-1)/2} \phi_P\left(\frac{y + \sqrt{y^2 - 4x^2}}{2x}\right).$$

7.3 Numerical results

The computations were done on a PC Pentium Pro 200MHz, by a program written in C, using the PARI library version 1.915. We give in this table the value of the main constants for a few primes; the program, complete numerical details, and results for many other values of P are available from the second author.

The last two columns of the following table contain respectively the total time of computation, and the time to compute and certify (using PARI) the fundamental units of $\mathbb{K}$ (both the times are in seconds). Compare the 4.3 seconds for the case $p = 67$ with the 28 minutes of [4].

p	m	c_6	c_7	c_8	c_{14}	c_{15}	B_0	B_0 red.	X_3	*Time*	*Time* (FU)
67	3	$8.01 \cdot 10^{12}$	4.65	2.00	$1.18 \cdot 10^{14}$	0.141	$2.05 \cdot 10^{28}$	7	7	4.3	1.0
311	5	$5.56 \cdot 10^{51}$	12.0	2.10	$9.01 \cdot 10^{52}$	0.077	$4.08 \cdot 10^{45}$	20	4	58.1	43.4
977	4	$4.73 \cdot 10^{153}$	76.2	7.89	$2.50 \cdot 10^{157}$	0.156	$2.58 \cdot 10^{42}$	80	2	60.5	16.5
997	3	$3.04 \cdot 10^{155}$	6.89	2.00	$4.48 \cdot 10^{156}$	0.014	$4.58 \cdot 10^{36}$	8	2	39.0	4.7
5011	3	$9.46 \cdot 10^{761}$	57.8	3.27	$4.95 \cdot 10763$	0.024	$5.77 \cdot 10^{40}$	46	2	479.8	6.6

We found that (for all P above) the solutions of the equation $F_P(x, y) = \pm 1$ are

$$(0, \pm 1), \quad (\pm 1, 0), \quad (\pm 1, \pm 1), \quad (\pm 1, \mp 1), \quad (\pm 1, \mp 2),$$

and the solutions of the equation $F_P(x, y) = \pm P$ are $(\pm 1, \pm 2)$.

Combining this with [18, Lemma 1], we obtain the following result.

Corollary 7.3.1 *The* 311-*th*, 977-*th*, 997-*th*, 5011-*th terms of any Lucas or Lehmer sequence have a primitive divisor.*

We intend to go further in that direction in a forthcoming paper.

References

[1] A. Baker, H. Davenport, *The equations* $3x^2 - 2 = y^2$ *and* $8x^2 - 7 = z^2$, Quat. J. Math. Oxford (2) 20 (1969), 129–137.

[2] A. Baker and G. Wüstholz, *Logarithmic forms and group varieties*, J. reine angew. Math. 442 (1993), 19–62.

[3] Yu. Bilu, *Solving superelliptic Diophantine equations by the method of Gelfond–Baker*, preprint 94-09, Mathématiques Stochastiques, Univ. Bordeaux 2, 1994.

[4] Yu. Bilu, G. Hanrot, *Solving Thue equations of high degree*, J. Number Th., 60 (1996), 373–392.

[5] Yu. Bilu, G. Hanrot, *Solving superelliptic Diophantine equations by Baker's method*, Compositio Math., to appear.

[6] Z.I. Borevich, I.R. Shafarevich, *Number Theory*, Academic Press, New York, 1966.

[7] H. Cohen, *A Course in Computational Algebraic Number Theory*, Graduate Texts in Math. 138, Springer, 1993.

[8] G. Hanrot, *Résolution effective d'équations diophantiennes : algorithmes et applications,* Thèse, Université Bordeaux 1, 1997.

[9] A.K. Lenstra, H.W. Lenstra jr., L. Lovász, *Factoring polynomials with rational coefficients,* Math. Ann. 261 (1982), 515–534.

[10] A. Pethö, *Computational Methods For the Resolution of Diophantine Equations*, in: R.A. Mollin (ed.), *Number Theory: Proc. First Conf. Can. Number Th. Ass., Banff, 1988*, de Gruyter, 1990, 477–492.

[11] A. Pethö, B.M.M. de Weger, *Products of Prime powers in Binary Recurrence Sequences Part I: The Hyperbolic Case, with an Application to the Generalized Ramanujan–Nagell Equation*, Math. Comp. 47 (1987), 713–727.

[12] M.E. Pohst, *Computational Algebraic Number Theory*, DMV Seminar 21, Birkhäuser, Basel, 1993.

[13] M.E. Pohst, H. Zassenhaus, *Algorithmic Algebraic Number Theory*, Cambridge Univ. Press, 1989.

[14] N. Smart, *The solution of triangularly connected decomposable form equation*, Math. Comp. 64 (1995), 819–840.

[15] C. Stewart, *Primitive divisors of Lucas and Lehmer numbers*, in: A. Baker and D.W. Masser (ed.), *Transcendence Theory: Advances and Applications*, Academic Press, 1977.

[16] N. Tzanakis, B.M.M. de Weger, *On the Practical Solution of the Thue Equation*, J. Number Th. 31 (1989), 99–132.

[17] N. Tzanakis, B.M.M. de Weger, *How to explicitly solve a Thue–Mahler Equation*, Compositio Math. 84 (1992), 223–288.

[18] P. Voutier, *Primitive divisors of Lucas and Lehmer sequences*, Math. Comp. 64 (1995), 869–888.

[19] B.M.M. de Weger, *Solving exponential diophantine equations using lattice basis reduction algorithms*, J. Number Th. 26 (1987), 325–367.

Bilu:
Forschungsinstitut für Mathematik
ETH-Zentrum
CH-8092 Zürich, SWITZERLAND
e-mail: `yuri@math.ethz.ch`

Hanrot:
Algorithmique Arithmétique Expérimentale (A2X)
UMR CNRS 9936
Université Bordeaux 1
351, cours de la Libération
F-33405 Talence Cedex, FRANCE
e-mail: `hanrot@math.u-bordeaux.fr`