Compositio Mathematica 66: 37–56 (1988)

Some diophantine equations with many solutions

P. ERDÖS,[1] C.L. STEWART[2] & R. TIJDEMAN[3]
[1]*Mathematics Institute of the Hungarian Academy of Sciences, Reáltanoda u. 13-15, Budapest, H-1053 Hungary;* [2]*Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1;* [3]*Mathematical Institute, University of Leiden, 2300 RA Leiden, The Netherlands*

Received 15 May 1987

§0. Introduction

Let $F(X, Y) \in \mathbb{Z}[X, Y]$ be an irreducible binary form of degree $n \geqslant 3$. In 1909, Thue proved that for each integer m the equation $F(x, y) = m$ has only finitely many solutions in integers x, y. Mahler extended Thue's result by proving that the number of solutions of $F(x, y) = m$ can be bounded in terms of F and the prime divisors of m. Several bounds for the number of solutions have been given. Let $S = \{p_1, \ldots, p_s\}$ be a set of prime numbers. Evertse solved an old conjecture of Siegel by proving that if F has non-zero discriminant, then the number of coprime pairs $x, y \in \mathbb{Z}$ such that $F(x, y)$ is composed of primes from S does not exceed $\exp(n^3(4s + 7))$. In the proofs of these results S-unit equations are used. An example of an S-unit equation is the equation $x + y = z$ in coprime positive integers x, y, z each composed of primes from S. Evertse also showed that this equation has at most $\exp(4s + 6)$ solutions. These results played a key role in the solution of an old conjecture of Erdös and Turán. Györy, Stewart and Tijdeman showed that if A and B are finite sets of k and l positive integers, respectively, and $k \geqslant l \geqslant 2$, then there exist a in A and b in B such that the greatest prime factor of $a + b$ exceeds $C \log k \operatorname{loglog} k$ where C is some positive constant.

In this paper we want to prove that there are Diophantine equations of above mentioned types which have surprisingly many solutions, thereby showing that some of the above results are not far from being the best possible ones. In §1 we consider the problem of Erdös and Turán. It follows from Theorem 1 that the bound $C \log k \operatorname{loglog} k$ cannot be replaced by $(\log k \operatorname{loglog} k)^2$. In §2 we turn to S-unit equations. We show in Theorem 4 that the equation $x + y = z$ can have more than $\exp((s/\log s)^{1/2})$ solutions in coprime positive integers x, y, z each composed of primes

From S. Finally, we deal with Thue–Mahler equations in §3. It follows from Theorem 5 that Evertse's bound $\exp(n^3(4s + 7))$ cannot be replaced by $\exp(n^2 s^{1/n}/\log s)$, not even when F is just a polynomial in one variable.

§1. Prime powers of sums of integers

For any integer $n > 1$ let $\omega(n)$ denote the number of distinct prime factors of n and let $P(n)$ denote the greatest prime factor of n. For any set X let $|X|$ denote the cardinality of X. In 1934 Erdös and Turán [15] proved that if A is a finite set of positive integers with $|A| = k$, then, for $k \geqslant 2$,

$$\omega\left(\prod_{a,a' \in A} (a + a')\right) > C_1 \log k.$$

They conjectured (cf. [14] p. 36) that for every h there is an $f(h)$ so that if A and B are finite sets of positive integers with $|A| = |B| = k \geqslant f(h)$ then

$$\omega\left(\prod_{a \in A, b \in B} (a + b)\right) > h.$$

György, Stewart and Tijdeman [23] proved the conjecture. They showed that the following much stronger assertion is an easy consequence of a result of Evertse [16]. Let A and B be finite sets of positive integers. Put $k = |A|$, $l = |B|$. If $k \geqslant l \geqslant 2$, then

$$\omega\left(\prod_{a \in A, b \in B} (a + b)\right) > C_2 \log k \tag{1}$$

where C_2 is an effectively computable positive constant. On combining this result with the prime number theorem we obtain that there exist a in A and b in B such that

$$P(a + b) > C_3 \log k \operatorname{loglog} k \tag{2}$$

where C_3 is an effectively computable positive constant. Other lower bounds for $\max_{a \in A, b \in B} P(a + b)$ have been given by Balog and Sárközy [2], Sárközy and Stewart [36, 37], György, Stewart and Tijdeman [23, 24] and Stewart and Tijdeman [44]. For surveys of these results we refer to Stewart [43] and Stewart and Tijdeman [44].

In this paragraph we want to show that (1) and (2) are not far from best possible, when l is small. It follows from Theorem 1 that when $l = 2$ the

right hand sides of (1) and (2) cannot be replaced by $((1/8) + \varepsilon)(\log k)^2$ loglog k and $((1/4) + \varepsilon)(\log k \log\log k)^2$, respectively, for any $\varepsilon > 0$. Theorem 1 deals with values of l which are $o(\log k)$. It follows from this theorem that the right hand sides of (1) and (2) cannot be replaced by $(\log k)^l$ when $l > 2$ loglog k. In Theorem 2 we consider values of l of the form $\delta \log k$ with $0 < \delta < 1$. It follows from this theorem that, even for such l, the right hand side of (2) cannot be replaced by $k^{1-\varepsilon}$ for every $\varepsilon > 0$ and $k \geqslant k_0(\delta, \varepsilon)$. We conjecture, however, that for $l > \log k$ and for every $\varepsilon > 0$ (2) can indeed be replaced by

$$P(a + b) > k^{1-\varepsilon}$$

for $k \geqslant k_1(\varepsilon)$. The trivial example $a_i = i, b_j = j - 1$ shows that the right-hand side of (2) cannot be replaced by $k + l$.

THEOREM 1. *Let $0 < \varepsilon < 1$. Let $f\colon \mathbb{R}_{>1} \to \mathbb{R}$ be a function such that $f(x) \to \infty$ as $x \to \infty$ and that $f(x)/\log x$ is monotone and non-increasing. Let k and l be positive integers such that k exceeds some effectively computable number depending only on ε and f and that $2 \leqslant l \leqslant (\log k)/f(k)$. Then there exist distinct positive integers $a_1, \ldots, a_k$ and distinct non-negative integers $b_1, \ldots, b_l$ such that*

$$P\left(\prod_{i=1}^{k}\prod_{j=1}^{l}(a_i + b_j)\right) < \left((1 + \varepsilon)\frac{\log k}{l}\log\left(\frac{\log k}{l}\right)\right)^l.$$

It is not hard to derive upper bounds for the numbers $a_1, \ldots, a_k$ and $b_1, \ldots, b_l$ in Theorem 1 from the proof.

Theorem 1 follows from Lemma 3 which is derived from Lemmas 1 and 2. Lemma 1 is a combinatorial result which is fundamental for all the results in this paper.

LEMMA 1. *Let N be a positive integer and let W be a non-empty subset of $\{1, \ldots, N\}$. Let l be an integer with $1 \leqslant l \leqslant |W|$. Then there is a set of non-negative integers B with $0 \in B$ and $|B| = l$ and a set A such that*

$$A + B \subseteq W \quad \text{and} \quad |A| \geqslant \binom{|W|}{l} \Big/ \binom{N-1}{l-1}.$$

Proof. There are $\binom{|W|}{l}$ l-element subsets of W. To each such subset $\{w_1, \ldots, w_l\}$ with $w_1 < \cdots < w_l$, associate the $(l-1)$-element subset $\{w_2 - w_1, \ldots, w_l - w_1\}$ of $\{1, \ldots, N-1\}$. Thus there is some $(l-1)$-element subset $\{b_1, \ldots, b_{l-1}\}$ associated to at least $k \geqslant \binom{|W|}{l}/\binom{N-1}{l-1}$ l-element subsets of W. Let $a_1, \ldots, a_k$ denote the least elements of these k different l-element subsets associated to $\{b_1, \ldots, b_{l-1}\}$. Thus $a_1, \ldots, a_k$ are distinct members of W. The lemma follows with $A = \{a_1, \ldots, a_k\}$, $B = \{0, b_1, \ldots, b_{l-1}\}$. □

let $\psi(x, y)$ be the number of positive integers not exceeding x which are free of prime divisors larger than y.

LEMMA 2. *Let x be a positive integer and u a real number with $u \geqslant 3$. There exists an effectively computable constant C such that*

$$\psi(x, x^{1/u}) \geqslant x \exp\left\{-u\left(\log u + \operatorname{loglog} u - 1 + C\left(\frac{\operatorname{loglog} u}{\log u}\right)\right)\right\}$$

Proof. See Theorem 3.1 of Canfield, Erdös, Pomerance [7].

For any real number z we shall denote the greatest integer less than or equal to z by $\lfloor z \rfloor$ and the least integer greater than or equal to z by $\lceil z \rceil$.

LEMMA 3. *Let c and δ be real constants with $c \geqslant 1$ and $0 < \delta < 1$. Let f be as in Theorem 1. Let N and l be positive integers such that N exceeds an effectively computable number depending only on c, δ and f and that $2 \leqslant l \leqslant (\log N)/f(N)$. Put*

$$m = \left\lceil \exp\left\{(1-\delta)\frac{\log N}{\log((\log N)/l)}\left(1 + \frac{\log c}{l}\right)\right\}\right\rceil$$

$$\text{and} \quad t = \left\lfloor c\left(\frac{\log N}{l}\right)^l \right\rfloor$$

Then there exist distinct integers $a_1, \ldots, a_m$ in $\{1, \ldots, N\}$ and there exist integers $b_1, \ldots, b_l$ with $0 = b_1 < b_2 < \cdots < b_l < N$ such that

$$P\left(\prod_{i=1}^{m}\prod_{j=1}^{l}(a_i + b_j)\right) \leqslant t.$$

Proof. Put $W = \{n \leqslant N\colon P(n) \leqslant t\}$. Then $|W| = \psi(N, t)$. For N sufficiently large we have

$$\psi(N, t) \geqslant \psi(N, \lfloor c(f(N))^l \rfloor) \geqslant c(f(N))^l/2 > l. \tag{3}$$

By Lemma 1 there exist sets A and B such that $0 \in B$, $|B| = l$, and $P(a + b) \leqslant t$ for all $a \in A$ and $b \in B$. It only remains to prove that $|A| \geqslant m$. By (3), for large N we have $l < (\psi(N, t))^{1/3}$. Hence, by Lemma 1,

$$|A| \geqslant \binom{\psi(N, t)}{l} \Big/ \binom{N-1}{l-1} \geqslant \frac{(\psi(N, t) - l)^l}{lN^{l-1}} \geqslant \frac{(\psi(N, t))^l}{lN^{l-1}}(1 + o(1)). \tag{4}$$

Here and later in the proof, $o(1)$ refers to $N \to \infty$. Put $x = N$, $y = t$, $u = (\log x)/\log y$ and $v = (\log N)/l$. Then $v \to \infty$ as $N \to \infty$. Hence,

$$\begin{aligned} u &= \frac{\log N}{\log \lfloor cv^l \rfloor} = \frac{\log N}{l \log v + \log c + o(1)} \\ &= \frac{\log N}{l \log v}\left(1 - \frac{\log c + o(1)}{l \log v}\right) = \frac{\log N}{l \log v} - \frac{v(\log c + o(1))}{l(\log v)^2} \\ &= (1 + o(1))\frac{v}{\log v}. \end{aligned}$$

This yields $\log u = \log v - \operatorname{loglog} v + o(1)$ and $\operatorname{loglog} u = \operatorname{loglog} v + o(1)$. Now, by Lemma 2,

$$\psi(N, t) \geqslant N \exp\left\{\left(-\frac{\log N}{l \log v} + \frac{v(\log c + o(1))}{l(\log v)^2}\right)(\log v - 1 + o(1))\right\}.$$

Hence, by (4),

$$|A| \geqslant (1 + o(1))\frac{N}{l} \exp\left\{\left(-\frac{\log N}{\log v} + \frac{(\log N)(\log c + o(1))}{l(\log v)^2}\right) \times (\log v - 1 + o(1))\right\}$$

$$= (1 + o(1))\frac{1}{l}\exp\left\{\frac{\log N}{\log v} + \frac{(\log N)(\log c)}{l \log v} + o\left(\frac{\log N}{\log v}\right)\right\}$$

$$= \frac{1}{l}\exp\left\{(1 + o(1))\frac{\log N}{\log v}\left(1 + \frac{\log c}{l}\right)\right\}.$$

Observe that

$$l < \log N = \exp\left\{o\left(\frac{\log N}{\log v}\right)\right\}.$$

Hence $|A| \geqslant m$ for N sufficiently large. □

Proof of Theorem 1. Put $w = (\log k)/l$. Then $w \to \infty$ as $k \to \infty$. We are going to apply Lemma 3 with $N = \lfloor \exp((1 + \varepsilon)(\log k)(\log w)) \rfloor$, $c = 1$ and $\delta = \varepsilon/5$. It follows that, for k sufficiently large, $k \leqslant N$ and $l \leqslant (\log k)/f(k) \leqslant (\log N)/f(N)$. Further

$$t = \left\lfloor \left(\frac{\log N}{l}\right)^{l} \right\rfloor \leqslant \left((1 + \varepsilon)\frac{\log k}{l}\log\left(\frac{\log k}{l}\right)\right)^{l}.$$

It therefore only remains to prove that $k \leqslant m$. We have, for k sufficiently large,

$$m \geqslant \exp\left\{\left(1 - \frac{\varepsilon}{4}\right)\frac{(1 + \quad)(\log k)(\log w)}{\log w + \operatorname{loglog} w + \log(1 + \varepsilon)}\right\}$$

$$\geqslant k^{\left(\left(1 + \frac{\varepsilon}{2}\right)\frac{\log w}{(1 + \varepsilon/2)\log w}\right)} = k.$$ □

For the statement of Theorem 2 we shall require the Dickman function $\varrho(u)$. $\varrho(u)$ is a positive, continuous, non-increasing function defined recursively by

$$\varrho(u) = 1 \quad \text{for } 0 \leqslant u \leqslant 1,$$

and, for $N = 1, 2, \ldots,$

$$\varrho(u) = \varrho(N) - \int_N^u v^{-1}\varrho(v - 1)\,dv \quad \text{for } N < u \leqslant N + 1.$$

Thus, in particular, $\varrho(u) = 1 - \log u$ for $1 \leqslant u \leqslant 2$. In general there is no known simple closed form for $\varrho(u)$ (cf. Appendix of [8]) and several authors [8], [29] have studied the problem of numerically approximating $\varrho(u)$. As for explicit bounds, it is easy to show that $\varrho(u) \leqslant 1/\Gamma(u+1)$ for $u \geqslant 1$, see for example Lemma 4.7 of [35], and Buchstab [6] proved that for $u \geqslant 6$,

$$\varrho(u) > \exp(-u(\log u + \operatorname{loglog} u + 6(\operatorname{loglog} u)/\log u)). \tag{5}$$

Further, de Bruijn [4] obtained the following asymptotic result,

$$\varrho(u) = \exp\left(-u\left(\log u + \operatorname{loglog} u - 1 + \frac{\operatorname{loglog} u}{\log u} - \frac{1}{\log u} + O\left(\left(\frac{\operatorname{loglog} u}{\log u}\right)^2\right)\right)\right). \tag{6}$$

THEOREM 2. *Let ε and θ be real numbers with $0 < \varepsilon < 1$ and $0 < \theta < 1$. Let k and l be positive integers with $2 \leqslant l \leqslant \theta \log k$ such that k exceeds a number which is effectively computable in terms of ε and θ. Then there exist distinct positive integers $a_1, \ldots, a_k$ and distinct non-negative integers $b_1, \ldots, b_l$ such that*

$$P\left(\prod_{i=1}^{k}\prod_{j=1}^{l}(a_i + b_j)\right) < k^{h(\theta)+\varepsilon}, \tag{7}$$

where

$$h(\theta) = \min_{u \geqslant 1}\left(\frac{1 - \theta \log \varrho(u)}{u}\right).$$

For any real number θ with $0 < \theta \leqslant 1$ define $f_\theta(u)$ for $u > 0$ by $f_\theta(u) = (1 - \theta \log \varrho(u))/u$. Since $\varrho(u)$ is continuous and $0 < \varrho(u) \leqslant 1$ for $u \geqslant 0$, $f_\theta(u)$ is also continuous and positive for $u > 0$. Further, by (6) $f_\theta(u)$ tends to infinity with u. Thus the minimum of $f_\theta(u)$ for $u \geqslant 1$ is attained and so $h(\theta)$ is well defined. If we evaluate $f_\theta(u)$ at $u = 1/\theta$ and apply Buchstab's inequality (5), we find that, for $\theta \leqslant 1/6$,

$$h(\theta) \leqslant \theta(1 + \log(1/\theta) + \operatorname{loglog}(1/\theta) + (6 \operatorname{loglog}(1/\theta)/\log(1/\theta))).$$

Plainly $f_0(1) = 1$ so $h(\theta) \leqslant 1$ for $0 < \theta \leqslant 1$. In fact, if $\theta < 1$ then $h(\theta) < 1$. To see this recall that $\varrho(u) = 1 - \log u$ for $1 \leqslant u \leqslant 2$. Thus, if $u = 1 + \delta$ with $0 < \delta < 1/2$, then

$$f_0(1 + \delta) = (1 - \theta \log (1 - \log (1 + \delta)))/(1 + \delta)$$

$$= (1 + \theta(\delta + O(\delta^2)))/(1 + \delta),$$

and so for δ sufficiently small $f_0(1 + \delta) < 1$, whence $h(\theta) < 1$, for $0 < \theta < 1$. Thus, for $2 \leqslant l \leqslant \theta \log k$ and $0 < \theta < 1$, (7) is an estimate which is better by a power than the trivial estimate $k + l$ which is realized when $a_i = i$, $b_j = j - 1$. Certainly (7) also holds with $P(\prod_{i=1}^{k} \prod_{j=1}^{l} (a_i + b_j))$ replaced by $\omega(\prod_{i=1}^{k} \prod_{j=1}^{l} (a_i + b_j))$ and in this case the trivial upper bound is $\pi(k + l)$. We conjecture that there does not exist a positive real number γ with $\gamma < 1$ and arbitrarily large integers l and k with $l > \log k$ for which there exist distinct positive integers $a_1, \ldots, a_k$ and distinct non-negative integers $b_1, \ldots, b_l$ such that

$$\omega\left(\prod_{i=1}^{k} \prod_{j=1}^{l} (a_i + b_j)\right) < (\pi(k + l))^{\gamma}.$$

We are able, however, to make some improvements on the trivial estimate $\pi(k + l)$ for $l > \log k$. In particular, there exist positive real numbers β_0 and β_1 such that for all sufficiently large integers k there exist positive integers $a_1, \ldots, a_k$ and $b_1, \ldots, b_l$ with $l > (1 + \beta_0) \log k$ for which

$$P\left(\prod_{i=1}^{k} \prod_{j=1}^{l} (a_i + b_j)\right) < (1 - \beta_1)(k + l), \tag{8}$$

hence, by the prime number theorem, for which

$$\omega\left(\prod_{i=1}^{k} \prod_{j=1}^{l} (a_i + b_j)\right) \leqslant \pi((1 - \beta_1)(k + l)) = (1 - \beta_1 + o(1))\pi(k + l).$$

To prove (8) we shall require the following result of independent interest. Let $2 = p_1, p_2, p_3, \ldots$ be the sequence of consecutive prime numbers.

LEMMA 4. *There are effectively computable positive real numbers β and n_0 so that if $n > n_0$ then*

$$\sum_{\substack{n < p_k < p_{k+1} < 2n \\ p_{k+1} - p_k > (1+\beta)\log n}} (p_{k+1} - p_k) > \beta n.$$

Proof. Let θ be a positive real number. Let L be the number of indices k with $n < p_k < 2n$ and $(1 - \theta) \log n < p_{k+1} - p_k < (1 + \theta) \log n$. It follows from Lemmas 1 and 2 of [13] that there exists an effectively computable positive constant c such that $L \leqslant c\theta n/\log n$. By the prime number theorem, the number of indices k with $n < p_k < 2n$ and $p_{k+1} - p_k \leqslant (1 - \theta) \log n$ is at most $(1 + o(1))n/\log n - L$. This implies that

$$\sum_{\substack{n<p_k<p_{k+1}<2n \\ p_{k+1}-p_k<(1+\theta)\log n}} (p_{k+1} - p_k) < \left\{(1 + o(1)) \frac{n}{\log n} - L\right\} \times (1 - \theta) \log n + L(1 + \theta) \log n$$

$$= (1 + o(1))(1 - \theta)n + 2L\theta \log n \leqslant \{1 - \theta + 2c\theta^2 + o(1)\}n. \tag{9}$$

On the other hand, since $p_{k+1} - p_k = o(n)$ for $n < p_k \leqslant 2n$ or $n < p_{k+1} \leqslant 2n$, see for example [26], we have

$$\sum_{n<p_k<p_{k+1}<2n} p_{k+1} - p_k = (1 + o(1))n. \tag{10}$$

A comparison of (9) and (10) reveals that

$$\sum_{\substack{n<p_k<p_{k+1}<2n \\ p_{k+1}-p_k\geqslant(1+\theta)\log n}} p_{k+1} - p_k > (\theta - 2c\theta^2 + o(1))n.$$

For θ sufficiently small and n sufficiently large we have $(\theta - 2c\theta^2 + o(1))n > (\theta/2)n$ and the result now follows on taking $\beta = \theta/2$. □

Suppose now that k is a positive integer, put $\beta_2 = \beta/2$ and $\beta_3 = ((\beta/4)/(1 + \beta))\beta$ and let n be that integer for which $(1 + \beta_3)n \leqslant k < (1 + \beta_3)(n + 1)$. Let T be the set of integers $1, \ldots, n$ together with the integers m with $n < m < 2n - (1 + \beta_2) \log n$ for which the closed interval $[m, m + (1 + \beta_2) \log n]$ contains no prime numbers. If j is a subscript such that $n < p_j < p_{j+1} < 2n$ and $p_{j+1} - p_j > (1 + \beta) \log n$, then all integers in the open interval $(p_j, p_j + (p_{j+1} - p_j)(\beta/2)/(1 + \beta))$ belong to T. Hence, by Lemma 4, T has cardinality at least

$$n + \left(\frac{\beta/2}{1 + \beta} + o(1)\right) \left(\sum_{\substack{n<p_j<p_{j+1}<2n \\ p_{j+1}-p_j>(1+\beta)\log n}} (p_{j+1} - p_j)\right)$$

$$> n + \beta n \left(\frac{\beta/2}{1 + \beta} + o(1)\right),$$

and plainly this exceeds k if k is sufficiently large. Thus we can choose $a_1, \ldots, a_k$ from T. Put $l = \lfloor (1 + \beta_2) \log n \rfloor$ and $b_j = j$ for $j = 1, \ldots, l$. Note that $l = (1 + \beta_2 + o(1)) \log k$ and that by construction,

$$P\left(\prod_{i=1}^{k}\prod_{j=1}^{l}(a_i + b_j)\right) \leqslant n + l = \left(\frac{1}{1 + \beta_3} + o(1)\right)k,$$

hence (8) follows directly.

Let h be a positive integer. We can prove, by appealing to a result of Maier (see the main theorem of [34]) and employing a similar construction to the one given above, that there exists a positive number c_h, which is effectively computable in terms of h, and arbitrarily large integers k and l with $l > c_h(\log k \log\log k \log\log\log\log k)/(\log\log\log k)^2$ for which there exist distinct positive integers $a_1, \ldots, a_k$ and distinct positive integers $b_1, \ldots, b_l$ with

$$\omega\left(\prod_{i=1}^{k}\prod_{j=1}^{l}(a_i + b_j)\right) < \pi(k) - h.$$

On the other hand, perhaps for each positive number ε there exists a number $k_0(\varepsilon)$ such that if $k > k_0(\varepsilon)$ and $l > (\log k)^{2+\varepsilon}$ then

$$\omega\left(\prod_{i=1}^{k}\prod_{j=1}^{l}(a_i + b_j)\right) \geqslant \pi(k + l),$$

for any distinct positive integers $a_1, \ldots, a_k$ and distinct positive integers $b_1, \ldots, b_l$. If it is true this conjecture will be very deep.

For the proof of Theorem 2 we shall require the following two lemmas.

LEMMA 5. *Let u be a real number with $u \geqslant 1$. Then*

$$\psi(x, x^{1/u}) \sim x\varrho(u).$$

Proof. This result is due to Dickman [12], see also de Bruijn [5].

LEMMA 6. *Let δ and u be real numbers with $0 < \delta < 1$ and $1 \leqslant u$. Let N and l be positive integers such that N exceeds a number which is effectively computable in terms of δ and u and such that $2 \leqslant l \leqslant \log N$. Put*

$$m = \left\lceil \frac{N}{l}((1 - \delta)\varrho(u))^l \right\rceil \quad \text{and} \quad t = N^{1/u}.$$

Then there exist distinct integers $a_1, \ldots, a_m$ *in* $\{1, \ldots, N\}$ *and there exist integers* $b_1, \ldots, b_l$ *with* $0 = b_1 < b_2 < \cdots < b_l < N$ *such that*

$$P\left(\prod_{i=1}^{m}\prod_{j=1}^{l}(a_i + b_j)\right) \leqslant t.$$

Proof. Put $W = \{1 \leqslant n \leqslant N: P(n) \leqslant t\}$. Then $|W| = \psi(N, t)$. For N sufficiently large we have

$$\psi(N, t) \geqslant t > l^3.$$

By Lemma 1 there exist sets A and B of non-negative integers such that $|A| \geqslant \binom{|W|}{l}/\binom{N-1}{l-1}$, $0 \in B$, $|B| = l$ and $A + B \subseteq W$, so in particular $P(a + b) \leqslant t$ for all $a \in A$ and $b \in B$. It remains to prove that $|A| \geqslant m$. We have, as in (4),

$$|A| \geqslant \frac{(\psi(N, t))^l(1 + o(1))}{lN^{l-1}}.$$

Thus, by Lemma 5,

$$|A| \geqslant \frac{(N\varrho(u)(1 + o(1)))^l}{lN^{l-1}},$$

and the result follows. □

Proof of Theorem 2. Suppose that $f_0(u) = (1 - \theta \log \varrho(u))/u$ attains its minimum value for $u \geqslant 1$ at $u = u_0$. We apply Lemma 6 with $N = \lceil kl((1 - (\varepsilon/2))\varrho(u_0))^{-l}\rceil$, $\delta = \varepsilon/2$ and $u = u_0$. Since $\varrho(u_0) \leqslant 1$ we have $N \geqslant k$, whence $2 \leqslant l \leqslant \theta \log k \leqslant \log N$. Further

$$m = \left\lceil \frac{N}{l}((1 - (\varepsilon/2))\varrho(u_0))^l \right\rceil \geqslant k.$$

Thus, for k sufficiently large in terms of ε and θ, there exist distinct integers $a_1, \ldots, a_k$ from $\{1, \ldots, N\}$ and integers $b_1, \ldots, b_l$ with $0 = b_1 < b_2 < \cdots < b_l < N$ with

$$P\left(\prod_{i=1}^{k}\prod_{j=1}^{l}(a_i + b_j)\right) \leqslant N^{1/u_0}.$$

Furthermore,

$$N^{1/u_0} \leqslant 2 \exp\left(\frac{\log k}{u_0}\left(1 + \frac{\text{loglog } k}{\log k} - \theta \log (1 - (\varepsilon/2)) - \theta \log \varrho(u_0)\right)\right).$$

Since $-\theta \log (1 - (\varepsilon/2)) < -\log (1 - (\varepsilon/2)) < 3\varepsilon/4$, we have, for k sufficiently large in terms of ε and θ,

$$N^{1/u_0} \leqslant k^{h(0)+\varepsilon},$$

as required. □

§2. S-unit equations with many solutions

Let $S = \{p_1, \ldots, p_s\}$ be a set of prime numbers. Let a, b and c be non-zero integers. Then the equation

$$ax + by = cz \tag{11}$$

in integers x, y and z which are all composed of primes from S is called an S-unit equation (over $\mathbb{Q}$). Usually S-unit equations are defined over algebraic number fields or other finitely generated domains. An extensive survey on these equations has been given by Evertse et al. [18].

It follows from the work of Mahler [31] (cf. Lang [27]) that the S-unit equation (11) has only finitely many solutions in coprime integers x, y, z. Mahler dealt explicitly with the case $a = b = c = 1$. An upper bound for the number of solutions in this case was given by Lewis and Mahler [28]. Their bound depends on S. Evertse [16] proved for general a, b, c that the S-unit equation (11) has at most $3 \times 7^{2s+3}$ solutions in coprime integers x, y, z (see also Silverman [41]). Generically the number of solutions of equation (11) is much smaller. S-unit equations split in a natural way into equivalence classes (cf. [18, 19]) in such a way that it is a trivial matter to compute all the solutions of an S-unit equation if one knows the solutions of an equivalent S-unit equation. Further the number of solutions of equivalent S-unit equations are equal. The number of equivalence classes is infinite, but Evertse et al. [19] proved that, with the exception of only finitely many equivalence classes, the number of solutions of the S-unit equation (11) in coprime positive integers x, y, z is at most two. By contrast, it follows from Theorem 4 that the S-unit equation $x + y = z$ can have at least as many as $\exp((4 + o(1))(s/\log s)^{1/2})$ coprime positive solutions and hence

Evertse's upper bound is not far from the best possible bound. On the basis of a heuristic computation we think that the truth is in between. We conjecture that if ε is any positive real number and S the set of the first s primes then the number of solutions of the S-unit equation $x + y = z$ in coprime positive integers x, y and z is at least $\exp(s^{(2/3)-\varepsilon})$ for $s > C_1(\varepsilon)$ and, on the other hand, if S is any set of s primes, then the number of solutions is at most $\exp(s^{(2/3)+\varepsilon})$ for $s > C_2(\varepsilon)$. Theorem 3 shows that in Theorem 4 one of x and y can be fixed at the cost of replacing $\exp((4-\varepsilon)(s/\log s)^{1/2})$ by $\exp((2-\varepsilon)(s/\log s)^{1/2})$.

THEOREM 3. *Let $2 = p_1, p_2, \ldots$ be the sequence of prime numbers. Let ε be a positive real number. There exists a positive number $s_0(\varepsilon)$ which is effectively computable in terms of ε such that if s is an integer with $s > s_0(\varepsilon)$ then there exist positive integers k_1 and k_2 with*

$$k_1 < \exp(2(s \log s)^{1/2}), \quad k_2 < \exp((s \log s)^{1/2}),$$

such that the equation

$$x - y = k_1$$

has at least $\exp((4-\varepsilon)(s/\log s)^{1/2})$ solutions in positive integers x and y with $P(xy) \leqslant p_s$ and such that the equation

$$x - y = k_2$$

has at least $\exp((2-\varepsilon)(s/\log s)^{1/2})$ solutions in coprime positive integers x and y with $P(xy) \leqslant p_s$.

THEOREM 4. *Let ε be a positive real number. There exists a number $s_0(\varepsilon)$, which is effectively computable in terms of ε, such that if s is an integer larger than $s_0(\varepsilon)$ then there exists a set S of prime numbers with $|S| = s$ for which the equation*

$$x + y = z$$

has at least $\exp((4-\varepsilon)(s/\log s)^{1/2})$ solutions in coprime positive integers composed of primes from S.

Proof. Let $0 < \varepsilon < 1$ and let s be so large that the following arguments hold true. Apply Lemma 3 with $c = 1$, the positive number δ to be chosen

later, $f(x) = (\log x)/2$, $N = \lfloor \exp((2 - \delta)(s \log s)^{1/2}) \rfloor$ and $l = 2$. Then there exists an integer m_1 with

$$m_1 \geqslant \exp\left\{(1 - \delta)\frac{4(1 - \delta)(s \log s)^{1/2}}{(1 + \delta)\log s}\right\}$$

and there exist integers $a_1, \ldots, a_{m_1}$ in $\{1, \ldots, N\}$ and b in $\{1, \ldots, N - 1\}$ such that

$$P\left(\prod_{i=1}^{m_1} a_i(a_i + b)\right) \leqslant \left(1 - \frac{\delta}{2}\right)^2 s \log s.$$

Taking $x_i = a_i + b$, $y_i = a_i$ for $i = 1, \ldots, m_1$, we obtain a positive integer $k_1 = b$ with $k_1 < \exp(2(s \log s)^{1/2})$ and m_1 solutions x_i, y_i of the equation $x - y = k_1$ with $P(x_i y_i) \leqslant (1 - \delta/2)^2 s \log s \leqslant p_s$, the last inequality by the prime number theorem. Choosing δ so small that $4(1 - \delta)^2(1 + \delta)^{-1} > 4 - \varepsilon$, we obtain

$$m_1 > \exp\left\{(4 - \varepsilon)\left(\frac{s}{\log s}\right)^{1/2}\right\}$$

and the first assertion follows.

For the second statement apply Lemma 3 with $c = 4$, δ to be chosen later, $f(x) = (\log x)/2$, $N = \lfloor \exp((1 - \delta)(s \log s)^{1/2}) \rfloor$ and $l = 2$. Then there exists an integer m_2 with

$$m_2 \geqslant \exp\left\{(1 - \delta)\frac{2(1 - \delta)(s \log s)^{1/2}}{(1 + \delta)\log s}(1 + \log 2)\right\} \tag{12}$$

and there exist integers $a_1, \ldots, a_{m_2}$ in $\{1, \ldots, N\}$ and b in $\{1, \ldots, N - 1\}$ such that

$$P\left(\prod_{i=1}^{m_2} a_i(a_i + b)\right) \leqslant \left(1 - \frac{\delta}{2}\right)^2 s \log s \leqslant p_s.$$

Taking $x_i = a_i + b, y_i = a_i$ for $i = 1, \ldots, m_2$ we obtain a positive integer $k = b$ with $k < \exp((s \log s)^{1/2})$ and m_2 solutions x_i, y_i of the equation $x - y = k$ with $P(x_i, y_i) \leqslant p_s$. Let $d_i = gcd(x_i, y_i)$ for $i = 1, \ldots, m$. Then $d_i | k$ and $x_i/d_i, y_i/d_i$ is a solution of $x - y = k/d_i$. The number of possible values of k/d_i is at most the number of positive divisors of k. Since

$k \leqslant N$, the number of divisors of k does not exceed $\exp((1 + (\delta/2))(\log 2)(\log N)/\log\log N)$ (see Theorem 317 of Hardy and Wright [25]). Thus there exist positive integers d and $k_2 = k/d$ such that the equation $x - y = k_2$ has at least

$$m_2 \exp\left(-(1 + \delta)(1 - \delta)\left(\frac{s}{\log s}\right)^{1/2} 2 \log 2\right)$$

solutions $x_i/d, y_i/d$. Observe that all these solutions are coprime and distinct. Choose δ so small that $2(1 - \delta)^2(1 + \delta)^{-1}(1 + \log 2) - (1 + \delta)(1 - \delta)2 \log 2 > 2 - \varepsilon$. Then it follows from (12) that the number of solutions in coprime positive integers of the equation $x - y = k_2$ is at least $\exp((2 - \varepsilon)(s/\log s)^{1/2})$. Since these solutions x, y satisfy $0 < y < x < \exp((s/\log s)^{1/2})$ and $P(xy) \leqslant p_s$, and moreover $k_2 \leqslant k < \exp((s/\log s)^{1/2})$, this completes the proof.

Proof of Theorem 4. Let $0 < \delta < 1$. By Theorem 3 there exists a number $s_0^*(\delta)$ which is effectively computable in terms of δ such that if $s_1 > s_0^*(\delta)$ then there exists a positive integer k_1 with $k_1 < \exp(2(s_1/\log s_1)^{1/2})$ such that the equation $x - y = k_1$ has at least $\exp((4 - \delta))(s_1/\log s_1)^{1/2})$ solutions in positive integers x and y with $P(xy) \leqslant p_{s_1}$. We infer that the number of prime factors of k_1 does not exceed $4(s_1/\log s_1)^{1/2}$. Put

$$S = \{p \mid p \text{ prime and } p \leqslant p_{s_1} \text{ or } p \mid k_1\}.$$

Then $|S| \leqslant s_1 + 4(s_1/\log s_1)^{1/2} < (1 + \delta)s_1$ for s_1 sufficiently large. We now choose s_1 by $s_1 = \lfloor s(1 + \delta)^{-1} \rfloor$. Then $|S| \leqslant s$. By making δ sufficiently small with respect to ε we obtain that the number of solutions, in positive integers x, y composed of primes from S, of the equation $k_1 + y = x$ is at least $\exp((4 - \varepsilon)(s/\log s)^{1/2})$ and, by dividing out the common factor, we obtain this many distinct solutions in coprime positive integers x, y and z of the equation $x + y = z$ such that each of x, y and z are composed of primes from S. □

§3. Thue–Mahler equations with many solutions

Let $F(X, Y)$ be a binary form with integer coefficients of degree $n \geqslant 3$ and let $S = \{p_1, \ldots, p_s\}$ be a set of prime numbers. The equation

$$F(x, y) = p_1^{z_1} \ldots p_s^{z_s} \tag{13}$$

in non-negative integers $x, y, z_1, \ldots, z_s$ is called a Thue–Mahler equation. It becomes a Thue equation if $z_1, \ldots, z_s$ are all fixed. Mahler [30, 31] proved that if F is irreducible, then equation (13) has at most c^s solutions with x and y coprime where the number c depends only on F. Lewis and Mahler [28] derived explicit upper bounds for the number of coprime solutions of (13) in terms of F and S when F is a binary form with non-zero discriminant. Evertse [16] succeeded in deriving an upper bound for the number of coprime solutions which depends only on n and s. He showed that if the binary form F is divisible by at least three pairwise linearly independent forms in some algebraic number field, then the number of solutions of (13) in non-negative integers $x, y, z_1, \ldots, z_s$ with $gcd(x, y) = 1$ is at most

$$2 \times 7^{n^3(2s+3)}$$

(see also Mahler [33] and Silverman [38, 39]). Upper bounds for the solutions themselves were provided by Coates [9, 10], Sprindzhuk [42], Györy [20] and others.

One may wonder how many solutions equation (13) can have. Theorem 5 shows that Evertse's bound cannot be replaced by $\exp(s^{1/n}/\log s)$. There is a wide gap between the bound of Evertse and the one we have given, but we expect that the bound $\exp(s^{1/n}/\log s)$ is much closer to the truth than $\exp(s)$, say. In fact, Theorem 5 already applies to the Ramanujan–Nagell equation

$$F(x) = F(x, 1) = p_1^{z_1} \ldots p_s^{z_s}.$$

However, the polynomial F is not explicitly stated in Theorem 5. In this context, it is worthwhile to note the following immediate consequence of Theorem 4, which even gives a slightly better estimate than that of Theorem 5.

COROLLARY. *Let $\varepsilon > 0$. For $s > s_0(\varepsilon)$ there exists a set $S = \{p_1, \ldots, p_s\}$ of prime numbers such that the equation*

$$xy(x + y) = p_1^{z_1} \ldots p_s^{z_s}$$

has at least $\exp((4 - \varepsilon)(s/\log s)^{1/2})$ *solutions in non-negative integers $x, y, z_1, \ldots, z_s$ with $gcd(x, y) = 1$.*

The situation becomes entirely different for the Thue equation

$$F(x, y) = k \tag{14}$$

where F is a binary form as above and k is a non-zero integer. Upper bounds for the number of solutions of (14) have been given by Davenport and Roth [11], Silverman [38, 39], Evertse and Győry [17], Bombieri and Schmidt [3] and others. Upper bounds for the solutions themselves were provided by Baker [1], Győry and Papp [21, 22] and others. On the other hand, Silverman [40], extending work of Mahler [32], has shown that there exist infinitely many cubic binary forms, each with non-zero discriminant, such that the number of solutions of equation (14) exceeds $C(\log |k|)^{2/3}$ for infinitely many integers k where C is some positive constant. However, it may be that there exists a number C_1, depending only on n, such that equation (14) has at most C_1 solutions in coprime integers x, y.

We now proceed with Theorem 5.

THEOREM 5. *Let ε be a positive number. Let $2 = p_1, p_2, \ldots$ be the sequence of prime numbers and let l be an integer with $l \geqslant 2$. There exists a number $s_0(\varepsilon, l)$ which is effectively computable in terms of ε and l such that if s is an integer with $s \geqslant s_0(\varepsilon, l)$, then there exists a monic polynomial $F(X)$ of degree l with distinct roots and with rational integer coefficients for which the equation*

$$F(x) = p_1^{z_1} \ldots p_s^{z_s} \tag{15}$$

has at least

$$\exp\left\{(l^2 - \varepsilon)\frac{s^{1/l}}{(\log s)^{(l-1)/l}}\right\} \tag{16}$$

solutions in non-negative integers $x, z_1, \ldots, z_s$.

Proof. We assume that s is so large that the following arguments hold true. Apply Lemma 3 with $c = 1$, $f(x) = (\log x)/l$, $N = \lfloor \exp\{(l - \delta)(s \log s)^{1/l}\} \rfloor$ and the positive number δ to be chosen later. Then there exists an integer m with

$$m \geqslant \exp\left\{(1 - \delta)\frac{(l - 2\delta)(s \log s)^{1/l}}{\log((1 - (\delta/l))(s \log s)^{1/l})}\right\}$$

$$\geqslant \exp\left\{(1 - \delta)^3 \frac{l^2(s \log s)^{1/l}}{\log s}\right\}$$

and positive integers $a_1, \ldots, a_m$ and non-negative integers $b_1, \ldots, b_l$ such that

$$P\left(\prod_{i=1}^{m}\prod_{j=1}^{l}(a_i + b_j)\right) \leqslant \left(1 - \frac{\delta}{l}\right)^l s \log s.$$

By the prime number theorem, the right hand side of (17) does not exceed p_s, hence all the numbers $a_i + b_j$ are composed of $p_1, \ldots, p_s$. Put $F(X) = (X + b_1) \ldots (X + b_l)$. Then we have m solutions of the Diophantine equation (15) in non-negative integers $x, z_1, \ldots, z_s$. Choose δ so small that $(1 - \delta)^3 l^2 > l^2 - \varepsilon$. Then the number of solutions of equation (15) is at least (16). □

REMARK. The polynomial F mentioned in Theorem 5 has the special property that all its zeros are rational integers. The problem of finding a comparable lower bound for the number of solutions of (15) remains open if, for instance, F is irreducible over the rationals.

Acknowledgement

The research of the second author was supported in part by Grant A3528 from the Natural Sciences and Engineering Research Council of Canada.

References

1. A. Baker, Contributions to the theory of Diophantine equations. I. On the representation of integers by binary forms, *Philos. Trans. Roy. Soc. London Ser. A* 263 (1967/68) 173–191.
2. A. Balog and A. Sárközy, On sums of sequences of integers, II, *Acta Math. Hungar.* 44 (1984) 169–179.
3. E. Bombieri and W.M. Schmidt,On Thue's equation, *Invent. Math.* 88 (1987) 69–81.
4. N.G. de Bruijn, The asymptotic behaviour of a function occurring in the theory of primes, *J. Indian Math. Soc.* (N.S.) 15 (1951) 25–32.
5. N.G. de Bruijn, On the number of positive integers $\leqslant x$ and free of prime factors $> y$, *Nederl. Akad. Wetensch. Proc. Ser. A* 54 (1951) 50–60.
6. A.A. Buchstab, On those numbers in an arithmetical progression all prime factors of which are small in order of magnitude (Russian), *Dokl. Akad. Nauk. SSSR* 67 (1949) 5–8.
7. E.R. Canfield, P. Erdös and C. Pomerance, On a problem of Oppenheim concerning 'Factorisatio Numerorum', *J. Number Th.* 17 (1983) 1–28.
8. J.-M.-F. Chamayou, A probabilistic approach to a differential-difference equation arising in analytic number theory, *Math. Comp.* 27 (1973) 197–203.
9. J. Coates, An effective p-adic analogue of a theorem of Thue, *Acta Arith.* 15 (1968/69) 279–305.

10. J. Coates, An effective p-adic analogue of a theorem of Thue II, The greatest prime factor of a binary form, *Acta Arith.* 16 (1969/70) 399–412.
11. H. Davenport and K.F. Roth, Rational approximations to algebraic numbers, *Mathematika* 2 (1955) 160–167.
12. K. Dickman, On the frequency of numbers containing prime factors of a certain relative magnitude, *Ark. Mat. Astr. Fys.* 22 (1930) A10, 1–14.
13. P. Erdös, The difference of consecutive primes, *Duke Math. J.* 6 (1940) 438–441.
14. P. Erdös, Problems in number theory and combinatorics, *Proc. 6th Manitoba Conference on Numerical Mathematics*, Congress Numer. 18, Utilitas Math., Winnipeg, Man. (1977) 35–58.
15. P. Erdös and P. Turán, On a problem in the elementary theory of numbers, *Amer. Math. Monthly* 41 (1934) 608–611.
16. J.-H. Evertse, On equations in S-units and the Thue–Mahler equation, *Invent. Math.* 75 (1984) 561–584.
17. J.-H. Evertse and K. Győry, On unit equations and decomposable form equations, *J. reine angew. Math.* 358 (1985) 6–19.
18. J.-H. Evertse, K. Győry, C.L. Stewart and R. Tijdeman, S-unit equations and their applications, *New Advances in Transcendence Theory* (to appear).
19. J.-H. Evertse, K. Győry, C.L. Stewart and R. Tijdeman, On S-unit equations in two unknowns, *Invent. Math.* (to appear).
20. K. Győry, Explicit upper bounds for the solutions of some Diophantine equations, *Ann. Acad. Sc. Fenn. Ser. AI* 5 (1980) 3–12.
21. K. Győry and Z.Z. Papp, Effective estimates for the integer solutions of norm form and discriminant form equations, *Publ. Math. Debrecen* 25 (1978) 311–325.
22. K. Győry and Z.Z. Papp, Norm form equations and explicit lower bounds for linear forms with algebraic coefficients, *Studies in Pure Math. to the Memory of Paul Turán*, Birkhäuser, Basel, pp. 245–257.
23. K. Győry, C.L. Stewart and R. Tijdeman, On prime factors of sums of integers I, *Comp. Math.* 59 (1986) 81–88.
24. K. Győry, C.L. Stewart and R. Tijdeman, On prime factors of sums of integers III, *Acta Arith.* (to appear).
25. G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, 5th edn., Oxford (1979).
26. A.E. Ingham, On the difference between consecutive primes, *Quarterly J. Math. Oxford* 8 (1937) 255–266.
27. S. Lang, Integral points on curves, *Publ. Math. I.H.E.S.* 6 (1960) 27–43.
28. D.J. Lewis and K. Mahler, On the representations of integers by binary forms, *Acta Arith.* 6 (1960) 333–363.
29. J. van de Lune and E. Wattel, On the numerical solution of a differential – difference equation arising in analytic number theory, *Math. Comp.* 23 (1969) 417–421.
30. K. Mahler, Zur Approximation algebraischer Zahlen, I: Über den grössten Primteiler binärer Formen, *Math. Ann.* 107 (1933) 691–730.
31. K. Mahler, Zur Approximation algebraischer Zahlen, II: Über die Anzahl der Darstellungen grösser Zahlen durch binäre Formen, *Math. Ann.* 108 (1933) 37–55.
32. K. Mahler, On the lattice points on curves of genus 1, *Proc. London Math. Soc.* (2) 39 (1935) 431–466.
33. K. Mahler, On Thue's equation, *Math. Scand.* 55 (1984) 188–200.
34. H. Maier, Chains of large gaps between consecutive primes, *Adv. in Math.* 39 (1981) 257–269.
35. K.K. Norton, Numbers with small prime factors, and the least k-th power non-residue, *Memoirs of the American Math. Soc.* 106 (1971).

36. A. Sárközy and C.L. Stewart, On divisors of sums of integers I, *Acta Math. Hungar.* 48 (1986) 147–154.
37. A. Sárközy and C.L. Stewart, On divisors of sums of integers II, *J. reine angew. Math.* 365 (1986) 171–191.
38. J.H. Silverman, Integer points and the rank of Thue elliptic curves, *Invent. Math.* 66 (1982) 395–404.
39. J.H. Silverman, Representations of integers by binary forms and the rank of the Mordell-Weil group, *Invent. Math.* 74 (1983) 281–292.
40. J.H. Silverman, Integer points on curves of genus 1, *J. London Math. Soc.* 28 (1983) 1–7.
41. J.H. Silverman, *Quantitative results in Diophantine geometry*, Preprint, M.I.T. (1984).
42. V.G. Sprindzhuk, Estimation of the solutions of the Thue equation (Russian), *Izv. Akad. Nauk. SSSR Ser. Mat.* 36 (1972) 712–741.
43. C.L. Stewart, Some remarks on prime divisors of integers, *Séminaire de Théorie des Nombres, Paris 1984–85, Progress in Math.* 63, Birkhauser, Boston, etc. (1986) 217–223.
44. C.L. Stewart and R. Tijdeman, On prime factors of sums of integers II, *Diophantine Analysis, LMS Lecture Notes* 109, Cambridge Univ. Press (1986) 83–98.