ILLINOIS JOURNAL OF MATHEMATICS
Volume 43, Number 1, Spring 1999

CONGRUENCE PROPERTIES OF THE Ω-FUNCTION ON SUMSETS

J. RIVAT, A. SÁRKÖZYAND C. L. STEWART

ABSTRACT. In this article we investigate the behaviour of the omega function, which counts the number of prime factors of an integer with multiplicity, as one runs over those integers of the form $a+b$ where a is from a set A and b is from a set B. We prove, for example, that if A and B are sufficiently dense subsets of the first N positive integers and k is a positive integer then the number of pairs (a, b) for which the omega function of $a+b$ lies in a given residue class modulo k is roughly the total number of pairs divided by k.

1. Notation

Throughout this paper, we shall use the following notation: $c_1, c_2, \ldots$ denote positive absolute constants. $\mathbb{Z}$, $\mathbb{N}$ and $\mathbb{N}_0$ denote the set of integers, positive integers and non-negative integers respectively. The cardinality of a set S is denoted by $|S|$. $\lfloor x \rfloor$ and $\{x\}$ denote the integer part and the fractional part of x and $\|x\|$ denotes the distance from x to the nearest integer: $\|x\| = \min(\{x\}, 1-\{x\})$. We write $e^{2\pi i\alpha} = e(\alpha)$. If $f(n) = O(g(n))$, then we write $f(n) \ll g(n)$; if the implied constant depends on a certain parameter c, then we write $f(n) \ll_c g(n)$. $\mathcal{A}, \mathcal{B}, \ldots$ denote subsets of $\mathbb{N}_0$ and $\mathcal{A}+\mathcal{B}$ denotes the set of the non-negative integers n that can be represented in the form $n = a+b$ with $a \in \mathcal{A}$, $b \in \mathcal{B}$. $\omega(n)$ denotes the number of distinct prime factors of n and $\Omega(n)$ denotes the number of prime factors of n counted with multiplicity. $\lambda(n)$ is the Liouville function: $\lambda(n) = (-1)^{\Omega(n)}$. The divisor function is denoted by $\tau(n)$.

2. The results

In 1988, Sárközy[11] proved the following result. If $H > 0$, $N > N_0(H)$, $\mathcal{A}, \mathcal{B} \subset \{1, 2, \ldots, N\}$ and $|\mathcal{A}|, |\mathcal{B}| > N(\log N)^{-H}$, then both equations

$$\lambda(a+b) = +1 \qquad (a \in \mathcal{A}, b \in \mathcal{B})$$

Received November 11, 1996.
1991 Mathematics Subject Classification. Primary 11N64; Secondary 11B05.
Research of the second-named author partially supported by grants from the Hungarian National Foundation for Scientific Research and C.E.E. The paper was written while the author was visiting the Université Claude Bernard Lyon 1.
Research of the third-named author supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

and

$$\lambda(a'+b')=-1 \qquad (a'\in\mathcal{A}, b'\in\mathcal{B})$$

can be solved. In this paper our goal is to study two problems motivated by this theorem.

The first problem to study is the following related conjecture of Sárközy[12] which partly generalizes the result above.

CONJECTURE 1. *If $k\in\mathbb{N}$, $\varepsilon>0$, $N\in\mathbb{N}$, $N>N_1(\varepsilon,k)$, $\mathcal{A},\mathcal{B}\subset\{1,2,\dots,N\}$ and $|\mathcal{A}|,|\mathcal{B}|>\varepsilon N$ then there are $a\in\mathcal{A}$, $b\in\mathcal{B}$ with $k\mid\Omega(a+b)$.*

The conjecture in this form follows readily from a recent result of Tenenbaum [15]. However, by using an extension of Selberg's formula [13] (which is the crucial tool in Tenenbaum's paper as well) in order to attack this problem directly, we shall be able to obtain a sharper result than the one that can be derived from Tenenbaum's result.

THEOREM 1. *For all $k\in\mathbb{N}$, $h\in\mathbb{Z}$ there exist effectively computable positive numbers c_1, $E=E(k)$ and N_2 such that if $N>N_2$ and $\mathcal{A},\mathcal{B}\subset\{1,2,\dots,N\}$, then*

$$\left|\left|\{(a,b):a\in\mathcal{A},b\in\mathcal{B},\Omega(a+b)\equiv h \bmod k\}\right|-\frac{|\mathcal{A}||\mathcal{B}|}{k}\right| < c_1(|\mathcal{A}||\mathcal{B}|)^{1/2}\frac{N}{(\log N)^E}. \quad (1)$$

In particular, for $(|\mathcal{A}||\mathcal{B}|)^{1/2}>c_2kN(\log N)^{-E}$ there are $a\in\mathcal{A}$, $b\in\mathcal{B}$ with $\Omega(a+b)\equiv h \bmod k$.

How far is Sárközy's result [11] quoted above (and also Theorem 1) from being best possible? In other words, how large can subsets $\mathcal{A}$, $\mathcal{B}$ of $\{1,2,\dots,N\}$ be with the property that $\Omega(a+b)$ is of the same parity for all $a\in\mathcal{A}$, $b\in\mathcal{B}$? Two principles used in [4] and [10] give the following two theorems easily.

THEOREM 2. *Let $N\in\mathbb{N}$, $l\in\mathbb{N}$, $l<\log N$. There is an effectively computable constant N_3 such that if $N>N_3$ then there exist $\mathcal{A},\mathcal{B}\subset\{1,2,\dots,N\}$ such that $|\mathcal{B}|=l$,*

$$|\mathcal{A}|>\frac{N}{l\,3^l}$$

and $\Omega(a+b)$ is even for all $a\in\mathcal{A}$, $b\in\mathcal{B}$.

Indeed the result is an easy consequence of the following lemma of Erdős, Stewart and Tijdeman [4] (see also [10]).

LEMMA 1. *If $N \in \mathbb{N}$, $\mathcal{D} \subset \{1, 2, \ldots, N\}$, $\mathcal{D} \neq \emptyset$, $l \in \mathbb{N}$ and $l \leq |\mathcal{D}|$, then there are sets $\mathcal{E} \subset \mathcal{D}$ and $\mathcal{F} \subset \mathbb{N}_0$ such that*

$$\mathcal{E} + \mathcal{F} \subset \mathcal{D}, |\mathcal{E}| \geq \frac{\binom{|\mathcal{D}|}{l}}{\binom{N-1}{l-1}}, |\mathcal{F}| = l.$$

We apply the lemma with $\mathcal{D} = \{n : 2 \leq n \leq N, 2 \mid \Omega(n)\}$, and use the fact that, by a well-known consequence of the Prime Number Theorem, $|\mathcal{D}| = (\frac{1}{2} + o(1))N$. We choose $\mathcal{A}$ as the set $\mathcal{E}$ in the lemma shifted down by 1 and $\mathcal{B}$ as the set $\mathcal{F}$ in the lemma shifted up by 1. It is easy to see that these sets $\mathcal{A}$ and $\mathcal{B}$ satisfy the conditions in Theorem 2.

In the special case $\mathcal{A} = \mathcal{B}$, Theorem 7 in [10] yields:

THEOREM 3. *For all large N there is a set $\mathcal{A} \subset \{\lfloor N/4 \rfloor, \lfloor N/4 \rfloor + 1, \ldots, N\}$ such that $|\mathcal{A}| > \log\log N$ and for each pair a, a' with $a, a' \in \mathcal{A}$ and $a \neq a'$ we have $\Omega(a + a') = 2$.*

(See the remark at the end of [10] on the role of the condition $a \neq a'$ in this result.)

In [11], Sárközy also studied the special case of $\mathcal{B} = -\mathcal{A}$ $(= \{-a : a \in \mathcal{A}\})$, and he showed that there is an $\mathcal{A} \subset \{1, 2, \ldots, N\}$ such that $|\mathcal{A}| \gg \log N$ and $\Omega(a - a')$ is even for all $a, a' \in \mathcal{A}$, $a \neq a'$. The proof is based on a Ramsey theorem of Erdős and Szekeres from graph theory.

A common feature of all these lower bound results is that in each case the bound is provided by a counting argument of combinatorial nature so that the proofs are purely existence proofs. One might like to prove results of a more constructive type. To make a small step in this direction, one may start out from the fact that a "large" set of integers with even Ω values is the set of the squares. Thus one might like to look for "dense" sets $\mathcal{A}, \mathcal{B}$ such that $a + b$ is a square for all $a \in \mathcal{A}$, $b \in \mathcal{B}$.

Indeed, let $f(N)$ be the greatest integer f such that there are sets $\mathcal{A}, \mathcal{B}$ with

$$\mathcal{A}, \mathcal{B} \subset \{1, 2, \ldots, N\}, f = |\mathcal{A}| \geq |\mathcal{B}| \geq 2 \tag{2}$$

and

$$a + b \text{ is a square for all } a \in \mathcal{A}, b \in \mathcal{B}. \tag{3}$$

(We exclude the trivial case $|\mathcal{B}| = 1$.) If (3) holds then we shall say that the pair $\mathcal{A}$, $\mathcal{B}$ has the property P. We shall prove the following result.

THEOREM 4. *Let ϵ be a positive real number. There exists a positive number $N_4(\epsilon)$ which is effectively computable in terms of ϵ such that if N exceeds $N_4(\epsilon)$ then*

$$\exp\left((\log 2 - \varepsilon)\frac{\log N}{\log\log N}\right) < f(N) < \exp\left((\log 2 + \varepsilon)\frac{\log N}{\log\log N}\right). \tag{4}$$

(We remark that it is easy to establish the weaker lower bound for $f(N)$ obtained by replacing $\log 2$ with $\frac{1}{2}\log 2$ in the exponent on the left hand side of inequality (4).)

Again one might like to study the special case $\mathcal{A} = \mathcal{B}$, i.e, sums $a+a'$. If $a = a'$ is allowed, then $a+a = x^2$ forces a to be of the form $a = 2y^2$ which makes the problem much easier; thus we shall exclude this case. Correspondingly, we shall say that a set of integers Z has property Q if $z + z'$ is a square whenever z and z' are distinct elements of Z. Sets Z with property Q have been studied extensively. Independently, Erdős(Problem 40 in [3]) and Moser (Problem 94 in [14]) asked whether there are arbitrary large sets with property Q. For any positive integer N, let $g(N)$ denote the greatest integer g such that there is a set $\mathcal{A} \subset \{1, 2, \ldots, N\}$ with $|A| = g$ and with property Q. J. Lagrange [7] found a set Z of property Q with $|Z| = 6$. This example is

$$Z = \{-15863902, 17798783, 21126338, 49064546, 82221218, 447422978\}. \quad (5)$$

The same example was found by Nicolas [9] using computers. Moreover, Lagrange gave a parametric representation of 6-tuples $(z_1, z_2, z_3, z_4, z_5, z_6)$ with the property that at least 14 of the 15 sums $z_i + z_j$ with $1 \le i < j \le 6$ are squares. His formula for z_1 is

$$z_1 = 2x^8 - 24x^7 + 100x^6 - 24x^5 - 334x^4 + 48x^3 + 400x^2 + 192x + 32$$

(and the other five z's are represented by similar polynomials in x). One obtains the set (5) by taking $x = 5/7$. Based on these considerations, he conjectured that there are infinitely many 6-tuples with property Q. However, since then computer searches by M. Deléglise, Nicolas, Rivat and others have failed to produce further 6-tuples with property Q, either along these lines or by using any other approach. In fact we do not know of any 6-tuple of positive integers with property Q.

Lagrange's method produces 6-tuples $(z_1, z_2, z_3, z_4, z_5, z_6)$ which depend on two parameters x and m and he specializes to the case where $m = 1$ to obtain his parametric representation in terms of x. The general representation is given below.

$$\begin{aligned}
z_1 &= 2m^{10}x^8 - \left(8m^{10} + 16m^8\right)x^7 + \left(20m^{10} + 80m^6\right)x^6 \\
&\quad + \left(8m^{10} - 32m^8 - 64m^6 + 64m^4\right)x^5 \\
&\quad - \left(14m^{10} + 96m^6 + 224m^2\right)x^4 \\
&\quad - \left(16m^8 - 64m^6 - 128m^4 + 128m^2\right)x^3 \\
&\quad + \left(80m^6 + 320m^2\right)x^2 + \left(64m^4 + 128m^2\right)x + 32m^2, \qquad (6) \\
z_2 &= 2m^{10}x^8 + \left(8m^{10} + 16m^8\right)x^7 + \left(20m^{10} + 80m^6\right)x^6 \\
&\quad - \left(8m^{10} - 32m^8 - 64m^6 + 64m^4\right)x^5 \\
&\quad - \left(14m^{10} + 96m^6 + 224m^2\right)x^4 \\
&\quad + \left(16m^8 - 64m^6 - 128m^4 + 128m^2\right)x^3
\end{aligned}$$

$$+(80m^6+320m^2)x^2-(64m^4+128m^2)x+32m^2, \tag{7}$$

$$\begin{aligned} z_3 = &-(2m^{10}-4m^8)x^8\\ &+(4m^{12}-4m^{10}+8m^8-16m^6-32m^4)x^6\\ &-(2m^{10}+60m^8+32m^6-320m^4+32m^2-64)x^4\\ &+(16m^8-16m^6+32m^4-64m^2-128)x^2-32m^2+64, \end{aligned} \tag{8}$$

$$\begin{aligned} z_4 = &\,2m^{10}x^8-(8m^{10}-16m^8)x^7-(12m^{10}+48m^6)x^6\\ &+(8m^{10}+32m^8-64m^6-64m^4)x^5\\ &+(18m^{10}+160m^6+288m^2)x^4\\ &+(16m^8+64m^6-128m^4-128m^2)x^3\\ &-(48m^6+192m^2)x^2-(64m^4-128m^2)x+32m^2, \end{aligned} \tag{9}$$

$$\begin{aligned} z_5 = &\,2m^{10}x^8+(8m^{10}-16m^8)x^7-(12m^{10}+48m^6)x^6\\ &-(8m^{10}+32m^8-64m^6-64m^4)x^5\\ &+(18m^{10}+160m^6+288m^2)x^4\\ &-(16m^8+64m^6-128m^4-128m^2)x^3\\ &-(48m^6+192m^2)x^2+(64m^4-128m^2)x+32m^2, \end{aligned} \tag{10}$$

$$\begin{aligned} z_6 = &\,(m^{12}-2m^{10})x^8\\ &-(2m^{12}+4m^{10}-8m^8+16m^6-64m^4)x^6\\ &+(m^{12}-2m^{10}+80m^8-32m^6-240m^4-32m^2)x^4\\ &-(8m^8+16m^6-32m^4+64m^2-256)x^2\\ &+16m^4-32m^2. \end{aligned} \tag{11}$$

The failure of the computer search suggests that, perhaps, one can obtain only finitely many 6-tuples with property Q in this way. We shall show that this is so for any fixed value for m.

THEOREM 5. *For any rational number m there are only finitely many rational numbers x such that $a_1, a_2, a_3, a_4, a_5, a_6$ in* (6)–(11) *are* distinct *integers which form a set with property Q, and similarly, for any rational number x there are only finitely many rational numbers m with this property.*

The computer searches described above and Theorem 5 seem to suggest that $g(N)$ is bounded or, equivalently, that the answer to the question of Erdős and Moser is negative. While we are not able to establish such a result we are able to show that $g(N)$ must grow slowly as a function of N.

THEOREM 6. *There is an effectively computable real number N_5 such that if N*

exceeds N_5 *then*

$$g(N) < 37 \log N . \tag{12}$$

An interesting feature of this result is that the proof is based on a sieve result (Gallagher's larger sieve). It is unusual to obtain so small an upper bound by using a sieve method.

3. A lemma

The argument in this section is essentially based on [15].
The proof of Theorem 1 will be based on the following lemma.

LEMMA 2. *For* $k \in \mathbb{N}$, $k \geq 2$, $h \in \mathbb{Z}$ *and* $\alpha \in \mathbb{R}$ *we have*

$$\left| \sum_{\substack{n \leq N \\ \Omega(n) \equiv h \bmod k}} e(n\alpha) - \frac{1}{k} \sum_{n=1}^{N} e(n\alpha) \right| \ll \frac{N}{(\log N)^{E(k)}} \tag{13}$$

with $E(k) = \min(1, 2\sin^2(\frac{\pi}{k})) \geq \frac{4}{k^2} > 0$.

The proof of the lemma will require several steps. First we will show that Lemma 2 can be deduced from the following result.

LEMMA 3. *Let* $N \in \mathbb{N}$, $N \geq 2$, $z \in \mathbb{C}$, $|z| = 1$, *and* $\alpha \in \mathbb{R}$. *We have*

$$\left| \sum_{n \leq N} e(n\alpha) z^{\Omega(n)} \right| \ll N (\log N)^{\max(-1, \Re(z)-1)}$$

Indeed, using

$$\frac{1}{k} \sum_{l=0}^{k-1} e\left(n \frac{l}{k}\right) = \begin{cases} 1 & \text{if } n \equiv 0 \bmod k \\ 0 & \text{otherwise} \end{cases}$$

we can write

$$\begin{aligned} \sum_{\substack{n \leq N \\ \Omega(n) \equiv h \bmod k}} e(n\alpha) &= \sum_{n \leq N} e(n\alpha) \frac{1}{k} \sum_{l=0}^{k-1} e\left((\Omega(n) - h)\frac{l}{k}\right) \\ &= \frac{1}{k} \sum_{l=0}^{k-1} e\left(-h\frac{l}{k}\right) \sum_{n \leq N} e(n\alpha) e\left(\Omega(n)\frac{l}{k}\right) \\ &= \frac{1}{k} \sum_{n \leq N} e(n\alpha) \\ &\quad + \frac{1}{k} \sum_{l=1}^{k-1} e\left(-h\frac{l}{k}\right) \sum_{n \leq N} e(n\alpha) \left(e\left(\frac{l}{k}\right)\right)^{\Omega(n)} \end{aligned}$$

Hence

$$\left|\sum_{\substack{n\le N\\ \Omega(n)\equiv h \bmod k}} e(n\alpha) - \frac{1}{k}\sum_{n\le N} e(n\alpha)\right|$$
$$\le \max_{1\le l\le k-1}\left|\sum_{n\le N} e(n\alpha)\left(e\left(\frac{l}{k}\right)\right)^{\Omega(n)}\right|$$
$$\le \sup_{\substack{z\in\mathbb{C},|z|=1\\ \Re(z)\le\cos(2\pi/k)}}\left|\sum_{n\le N} e(n\alpha)z^{\Omega(n)}\right|$$

and $\cos\left(\frac{2\pi}{k}\right) - 1 = -2\sin^2\left(\frac{\pi}{k}\right)$ so Lemma 2 follows from Lemma 3.

It remains to prove Lemma 3.

Let $Q \ge 1$. By Dirichlet's theorem there exist coprime integers a and q with $1 \le q \le Q$ and $|\alpha - \frac{a}{q}| \le \frac{1}{qQ}$. We will treat the "large" and "small" values of q separately.

We now recall Corollary 1 of Montgomery and Vaughan [8].

LEMMA 4. *Let A be an arbitrary constant with $A \ge 1$, and let f be a multiplicative function such that $|f(p)| \le A$ for all primes p and $\sum_{n=1}^{N} |f(n)|^2 \le A^2 N$ for all natural numbers N.*

Suppose that $|\alpha - a/q| \le q^{-2}$, $(a,q) = 1$ and $2 \le R \le q \le N/R$. Then

$$\sum_{n=1}^{N} f(n)e(n\alpha) \ll_A \frac{N}{\log N} + NR^{-1/2}(\log R)^{3/2}.$$

We now choose $Q = N(\log N)^{-3}$. We can use Lemma 4 with $R = (\log N)^3$, and for $(\log N)^3 \le q \le Q$, get

$$\sup_{\substack{z\in\mathbb{C}\\ |z|=1}}\left|\sum_{n\le N} e(n\alpha)z^{\Omega(n)}\right| \ll \frac{N}{\log N}$$

In order to complete the proof of Lemma 3 it is sufficient to show that for $|\alpha - a/q| \le (qQ)^{-1}$, $(a,q) = 1$, $1 \le q \le (\log N)^3$, $z \in \mathbb{C}$, $|z| = 1$ we have

$$\left|\sum_{n\le N} e(n\alpha)z^{\Omega(n)}\right| \ll N(\log N)^{\Re(z)-1}.$$

The key argument is the following generalized Selberg formula, which is a weak version of Lemma 4 of Dupain, Hall, and Tenenbaum [2] (see also Lemma 1 of Tenenbaum [15]).

LEMMA 5. *Let $z \in \mathbb{C}$, $|z| = 1$, $N \in \mathbb{N}$, $N \geq 3$, $(a, q) = 1$, $1 \leq q \leq (\log N)^3$. We have*

$$\sum_{n \leq N} e(n\frac{a}{q}) z^{\Omega(n)} = N(\log N)^{z-1} \left(\tau_1(z,q) + \frac{\tau_2(z,q)}{\log N} + \frac{\tau_3(z,q)}{(\log N)^2} \right) + O(N(\log N)^{\Re(z)-4} (\log\log 3q)^7)$$

with some $\tau_j(z,q)$, $j = 1, 2, 3$, such that $\tau_j(z,q) \ll (\log\log 3q)^{j+3}$. .

Furthermore, writing

$$F(z) = \frac{1}{\Gamma(z+1)} \prod_p \left(1 - \frac{z}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^z$$

we have $\tau_1(z,q) = F(z) z^{\Omega(q)-\omega(q)+1} (z-1)^{\omega(q)} (\varphi(q))^{-1} \ll 1$.

Using this lemma we will be able to conclude our proof by partial summation.

PROPOSITION 1. *Let f, G_1, G_2 be arithmetic functions, and $a, b \in \mathbb{N}_0$. We then have*

$$\left| \sum_{a<n<b} f(n)(G_1(n) - G_1(n-1)) - \sum_{a<n<b} f(n)(G_2(n) - G_2(n-1)) \right| \leq \left((b-a) \max_{a<n<b} |f(n+1) - f(n)| + 2 \max_{a \leq n \leq b} |f(n)| \right) \max_{a \leq n \leq b} |(G_1 - G_2)(n)|$$

This proposition is an immediate consequence of the following partial summation formula with $G = G_1 - G_2$:

$$\sum_{a<n<b} f(n)(G(n) - G(n-1)) = \sum_{a<n<b} (f(n) - f(n+1))G(n) - f(a+1)G(a) + f(b)G(b-1)$$

Now we will use this proposition with a replaced by 0, b replaced by $N+1$ and

$$f(n) = e\left(n(\alpha - \frac{a}{q})\right),$$

$$G_1(n) = \sum_{1 \leq m \leq n} e\left(m\frac{a}{q}\right) z^{\Omega(m)},$$

$$G_2(x) = x(\log x)^{z-1} \left(\tau_1(z,q) + \frac{\tau_2(z,q)}{\log x} + \frac{\tau_3(z,q)}{(\log x)^2} \right).$$

Using the estimates

$$\begin{aligned} G_2(n) - G_2(n-1) &\ll (\log N)^{\Re(z)-1}, \\ |f(n+1) - f(n)| &\le 2\pi|\alpha - \frac{a}{q}| \le \frac{2\pi}{qQ}, \\ (G_2 - G_1)(n) &\ll N(\log N)^{\Re(z)-4}(\log\log 3q)^7, \end{aligned}$$

we finally obtain

$$\begin{aligned} \sum_{n\le N} e(n\alpha)z^{\Omega(n)} &\ll N(\log N)^{\Re(z)-1} \\ &\quad + \left(N\frac{2\pi}{qQ} + 1\right) N(\log N)^{\Re(z)-4}(\log\log 3q)^7 \\ &\ll N(\log N)^{\Re(z)-1} \\ &\quad + \frac{(\log N)^3}{q} N(\log N)^{\Re(z)-4}(\log\log 3q)^7 \\ &\ll N(\log N)^{\Re(z)-1}, \end{aligned}$$

which completes the proof of Lemma 3.

4. Completion of the proof of Theorem 1

Write

$$F(\alpha) = \sum_{a\in\mathcal{A}} e(a\alpha), \qquad G(\alpha) = \sum_{b\in\mathcal{B}} e(b\alpha)$$

and

$$H(\alpha) = \sum_{\substack{1\le n\le 2N \\ \Omega(n)\equiv h \bmod k}} e(n\alpha)$$

so that, by Lemma 2,

$$\left| H(\alpha) - \frac{1}{k}\sum_{n=1}^{2N} e(n\alpha) \right| \ll N(\log N)^{-E} \tag{14}$$

for all α. Set

$$\mathcal{J} = \int_0^1 F(\alpha)G(\alpha)H(-\alpha)d\alpha.$$

Clearly,

$$\begin{aligned} \mathcal{J} &= \int_0^1 \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} \sum_{\substack{1\le n\le 2N \\ \Omega(n)\equiv h \bmod k}} e((a+b-n)\alpha)d\alpha \qquad (15) \\ &= |\{(a,b) : a\in\mathcal{A}, b\in\mathcal{B}, \Omega(a+b)\equiv h \bmod k\}|. \end{aligned}$$

On the other hand, we have

$$\begin{aligned}\mathcal{J} &= \frac{1}{k}\int_0^1 F(\alpha)G(\alpha)\left(\sum_{n=1}^{2N} e(-n\alpha)\right)d\alpha \\ &\quad + \int_0^1 F(\alpha)G(\alpha)\left(H(-\alpha) - \frac{1}{k}\sum_{n=1}^{2N} e(-n\alpha)\right)d\alpha \qquad (16)\\ &= \mathcal{J}_1 + \mathcal{J}_2,\end{aligned}$$

say. Clearly we have

$$\begin{aligned}\mathcal{J}_1 &= \frac{1}{k}\int_0^1 \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}}\sum_{n=1}^{2N} e((a+b-n)\alpha)d\alpha \qquad (17)\\ &= \frac{1}{k}\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} 1 = \frac{1}{k}|\mathcal{A}||\mathcal{B}|.\end{aligned}$$

Moreover, by (14), Cauchy's inequality and Parseval's formula we have

$$\begin{aligned}|\mathcal{J}_2| &\le \int_0^1 |F(\alpha)||G(\alpha)|\left|H(-\alpha) - \frac{1}{k}\sum_{n=1}^{2N} e(-n\alpha)\right|d\alpha\\ &\ll N(\log N)^{-E}\int_0^1 |F(\alpha)||G(\alpha)|d\alpha\\ &\le N(\log N)^{-E}\left(\left(\int_0^1 |F(\alpha)|^2 d\alpha\right)\left(\int_0^1 |G(\alpha)|^2 d\alpha\right)\right)^{1/2}\\ &= (|\mathcal{A}||\mathcal{B}|)^{1/2}N(\log N)^{-E}. \qquad (18)\end{aligned}$$

(1) follows from (15),(16), (17),(18) and this completes the proof of Theorem 1.

5. Proof of Theorem 4, upper bound

First we shall prove the upper bound in (4). Fix N and assume that $\mathcal{A}$, $\mathcal{B}$ satisfy (2) and (3), and $\mathcal{A}$ is maximal, so that

$$|\mathcal{A}| = f = f(N). \qquad (19)$$

Write $\mathcal{A} = \{a_1, a_2, \ldots, a_f\}$, $\mathcal{B} = \{b_1, b_2, \ldots\}$ (where $b_1 < b_2 < \ldots$). Then by (2) and (3), for each of $i = 1, 2, \ldots, f$, there exist positive integers x_i, y_i such that

$$a_i + b_2 = x_i^2 \qquad (20)$$

and

$$a_i + b_1 = y_i^2 \qquad (21)$$

so that

$$x_i^2 - y_i^2 = b_2 - b_1.$$

Write

$$x_i - y_i = d_i \ (> 0), x_i + y_i = d_i' \tag{22}$$

so that

$$d_i d_i' = b_2 - b_1 \text{ for } i = 1, 2, \ldots, f. \tag{23}$$

Clearly, $b_2 - b_1$ can be factorized in at most $\tau(b_2 - b_1)$ ways in form (23), and by (22), d_i and d_i' determine x_i and y_i uniquely. Finally, by (20), x_i determines a_i uniquely. Thus $|\mathcal{A}| = f$, the number of a_i's, is at most $\tau(b_2 - b_1)$.

By (2) and Wigert's Theorem [16], it follows that

$$|\mathcal{A}| \leq \max_{n \leq N} \tau(n) < \exp\left((\log 2 + \varepsilon)\frac{\log N}{\log\log N}\right) \tag{24}$$

if N is large enough in terms of ε. The upper bound in (4) follows from (19) and (24).

6. Proof of Theorem 4, lower bound

The proof will be based on the following lemma.

LEMMA 6. *For all $\delta > 0$, $A > 0$ there are numbers $n_0 = n_0(\delta, A)$, $K = K(\delta, A)$ such that if $n > n_0$ and for all $p^\alpha || n$ we have*

$$p^\alpha < (\log n)^A \tag{25}$$

then for all but $\delta\tau(n)$ divisors d of n we have

$$\left|\frac{1}{2}\log n - \log d\right| < K(\log n \log\log n)^{1/2}. \tag{26}$$

(Note that a much sharper version of this result will be published soon in a joint paper by H. Maier and A. Sárközy.)

Proof of Lemma 6. Write $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$. Consider the independent random variables $\xi_1, \xi_2, \ldots, \xi_r$ defined in the following way: for $i = 1, 2, \ldots, r$, let ξ_i assume the values $0, \log p_i, 2\log p_i, \ldots, \alpha_i \log p_i$, each with probability $\frac{1}{\alpha_i + 1}$. Then the expectation and standard deviation of ξ_i are

$$M(\xi_i) = \frac{\alpha_i}{2}\log p_i \tag{27}$$

and

$$D(\xi_i) = \left(\frac{\alpha_i(\alpha_i+2)}{12}\right)^{1/2} \log p_i \le \frac{\alpha_i}{2}\log p_i. \tag{28}$$

respectively. Write $\eta = \xi_1 + \xi_2 + \cdots + \xi_r$ so that η assumes the values $\log d$ with $d \mid n$, and it assumes each of these values with probability $1/\tau(n)$. Then by (25), (27) and (28) we have

$$M(\eta) = M(\xi_1 + \xi_2 + \cdots + \xi_r) = \frac{1}{2}\log n \tag{29}$$

and

$$\begin{aligned} D(\eta) = \left(\sum_{i=1}^{r} D^2(\xi_i)\right)^{1/2} &\le \frac{1}{2}\left(\sum_{i=1}^{r}(\alpha_i \log p_i)^2\right)^{1/2} \\ &< \frac{1}{2}\left(\log(\log n)^A \sum_{i=1}^{r}\alpha_i \log p_i\right)^{1/2} \\ &= \frac{A^{1/2}}{2}(\log\log n \log n)^{1/2}. \end{aligned} \tag{30}$$

By Chebyshev's inequality, it follows from (29) and (30) that there is a number $L = L(\delta)$ such that

$$\begin{aligned} \delta &> P\left(|\eta - M(\eta)| \ge LD(\eta)\right) \\ &= \frac{1}{\tau(n)}\left|\left\{d : d \mid n, \left|\log d - \frac{1}{2}\log n\right| \ge LD(\eta)\right\}\right| \\ &\ge \frac{1}{\tau(n)}\left|\left\{d : d \mid n, \left|\log d - \frac{1}{2}\log n\right| \ge \frac{LA^{1/2}}{2}(\log n \log\log n)^{1/2}\right\}\right| \end{aligned}$$

whence the statement of the lemma follows with $K = LA^{1/2}/2$.

Now we shall prove the lower bound in (4). Let p_i denote the i-th prime, define t by

$$p_2 p_3 \cdots p_t \le N^{1-\varepsilon/2} < p_2 p_3 \cdots p_{t+1} \tag{31}$$

and write

$$B = p_2 p_3 \cdots p_t. \tag{32}$$

Then by the prime number theorem we have

$$t = (1+o(1))\left(1 - \frac{\varepsilon}{2}\right)\frac{\log N}{\log\log N}$$

so that,

$$\tau(B) = 2^{t-1} = \exp\left((1+o(1))(\log 2)\left(1-\frac{\varepsilon}{2}\right)\frac{\log N}{\log\log N}\right). \tag{33}$$

Consider all the divisors $d_1 < d_2 < \cdots < d_v$ of B with

$$B^{1/2-\varepsilon/4} < d_i < B^{1/2} - 1 \text{ (for } 1 \le i \le v) . \tag{34}$$

By (33) and Lemma 6, their number is

$$v > \frac{1}{3}\tau(B) > \exp\left((\log 2 - \varepsilon)\frac{\log N}{\log\log N}\right) , \tag{35}$$

for N sufficiently large in terms of ε. Let $b_1 = 1, b_2 = B+1$, and $\mathcal{B} = \{b_1, b_2\}$. Then by (31) and (32), clearly we have $\mathcal{B} \subset \{1, 2, \ldots, N\}$. For each d_i with $1 \le i \le v$, define d_i' by

$$d_i d_i' = B = b_2 - b_1 \tag{36}$$

so that, by the right hand side of inequality (34),

$$d_i + 2 < d_i'. \tag{37}$$

Next, define x_i and y_i by (22) so that

$$x_i = \frac{d_i + d_i'}{2} \text{ and } y_i = \frac{d_i' - d_i}{2} . \tag{38}$$

Note that B is odd and thus x_i, y_i are integers. Finally define a_i by (21) for $i = 1, \ldots, v$ and put $\mathcal{A} = \{a_1, \ldots, a_v\}$. Observe that $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$, since by (21), (31), (32), (34), (36), and (38),

$$\begin{aligned} 0 < a_i &= y_i^2 - b < y_i^2 < (d_i')^2 = \left(\frac{B}{d_i}\right)^2 \\ &< (B^{1/2+\varepsilon/4})^2 = B^{1+\varepsilon/2} \le (N^{1-\varepsilon/2})^{1+\varepsilon/2} < N \end{aligned}$$

for $i = 1, 2, \ldots, v$. Since (20) and (21) hold for $i = 1, \ldots, v$, $\mathcal{A}$ and $\mathcal{B}$ have the property P. Thus by (35),

$$f(N) \ge v > \exp\left((\log 2 - \varepsilon)\frac{\log N}{\log\log N}\right) ,$$

which completes the proof of Theorem 4.

Proof of Theorem 5. It suffices to consider the equation $y^2 = z_1 + z_2$ or equivalently $y^2 = 4m^2 f(x, m)$ where

$$f(x, m) = m^8x^8 + (10m^8 + 40m^4)x^6 - (7m^8 + 48m^4 + 112)x^4 + (40m^4 + 160)x^2 + 16$$

in rationals y, x and m.

We shall treat first the case when m is fixed. Notice that if $m = 0$ then $z_1 = z_2$ and so property Q does not hold. For m a fixed non-zero rational number, $f(x, m)$ is a polynomial of degree 8 in x with discriminant.

$$2^{48}m^{24}(m^2 - 2m + 2)^8(m^2 + 2m + 2)^8 g_1(m)^2 g_2(m)^2$$

where

$$g_1(m) = 7m^8 + 40m^6 + 40m^4 + 160m^2 + 112$$

and

$$g_2(m) = 7m^8 - 40m^6 + 40m^4 - 160m^2 + 112\,,$$

as computed in MAPLE. The discriminant is non-zero for m rational and so $f(x, m)$ has no repeated roots. Thus the curve $y^2 = f(x, m)$ has genus 3 and so by Faltings' Theorem [5] there are only finitely many rational points (x, y) on the curve and the result follows.

Next we consider the cases when x is fixed. We may assume that x is not 0, 1 or -1 since otherwise $z_1 = z_2$, $z_2 = z_4$ or $z_1 = z_5$ respectively. Then $f(x, m)$ is a polynomial of degree 8 in m with discriminant

$$-2^{72}x^{28}(x - 1)^8(x + 1)^8(x^2 + 1)^8(x^4 + 10x^2 - 7)^3(7x^4 - 10x^2 - 1)^3\,.$$

The discriminant was computed with the help of MAPLE. It is non-zero for x rational and different from 0, 1 and -1. Thus the curve $y^2 = f(x, m)$, this time with x fixed, has genus 3 and again by Faltings' Theorem there are only finitely many rational points (m, y) on the curve. The result now follows.

In order to prove Theorem 6, we shall need two lemmas.

LEMMA 7. *If p is a prime number with $p > 2$, $\mathcal{B} \subset \mathbb{Z}$, and $b, b' \in \mathcal{B}, b \neq b'$ implies that*

$$b \not\equiv b' \bmod p \tag{39}$$

and

$$\text{either } b + b' \equiv 0 \bmod p \text{ or } \left(\frac{b + b'}{p}\right) = +1 \tag{40}$$

(where $\left(\frac{n}{p}\right)$ denotes the Legendre symbol), then we have

$$|\mathcal{B}| < 6p^{1/2}.$$

Proof of Lemma 7. Write

$$G(h, p) = \sum_{x=0}^{p-1} e\left(\frac{hx^2}{p}\right)$$

and

$$G_0 = G(1, p)$$

so that

$$|G_0| = p^{1/2} \tag{41}$$

and

$$G(h, p) = \begin{cases} G_0 & \text{for} \quad \left(\frac{h}{p}\right) = +1, \\ -G_0 & \text{for} \quad \left(\frac{h}{p}\right) = -1, \\ p & \text{for} \quad p \mid h. \end{cases} \tag{42}$$

(See [1].) Assume that $\mathcal{B}$ satifies the assumptions in Lemma 7, and write

$$S = \sum_{x=0}^{p-1} \left(\sum_{b \in \mathcal{B}} e\left(\frac{bx^2}{p}\right)\right)^2.$$

Then, by (39), (40), (41) and (42), we have

$$\begin{aligned} |S| &= \left| \sum_{x=0}^{p-1} \sum_{b \in \mathcal{B}} \sum_{b' \in \mathcal{B}} e\left(\frac{(b+b')x^2}{p}\right)\right| \\ &= \left| \sum_{b \in \mathcal{B}} \sum_{b' \in \mathcal{B}} G(b+b', p)\right| \\ &\geq \left| \sum_{b \in \mathcal{B}} \sum_{b' \in \mathcal{B}} G_0 \right| - \sum_{b \in \mathcal{B}} \sum_{b' \in \mathcal{B}} |G_0 - G(b+b', p)| \\ &\geq |\mathcal{B}|^2 |G_0| - \sum_{b \in \mathcal{B}} |G_0 - G(2b, p)| - \sum_{\substack{b, b' \in \mathcal{B} \\ p \mid (b+b')}} |G_0 - G(0, p)| \\ &\geq |\mathcal{B}|^2 p^{1/2} - \sum_{b \in \mathcal{B}} 2p - \sum_{\substack{b, b' \in \mathcal{B} \\ p \mid (b+b')}} 2p \\ &\geq |\mathcal{B}|^2 p^{1/2} - 2|\mathcal{B}|p - 2|\mathcal{B}|p = |\mathcal{B}|^2 p^{1/2} - 4|\mathcal{B}|p. \end{aligned} \tag{43}$$

On the other hand, we have

$$|S| \leq \sum_{x=0}^{p-1} \left| \sum_{b \in \mathcal{B}} e\left(\frac{bx^2}{p}\right)\right|^2.$$

If x runs over $0, 1, \dots, p-1$, then x^2 meets every residue class at most twice. Thus by (39), it follows that

$$\begin{aligned} |S| &\le 2\sum_{y=0}^{p-1}\left|\sum_{b\in\mathcal{B}} e\left(\frac{by}{p}\right)\right|^2 \\ &= 2\sum_{y=0}^{p-1}\sum_{b\in\mathcal{B}}\sum_{b'\in\mathcal{B}} e\left(\frac{(b-b')y}{p}\right) \\ &= 2\sum_{b\in\mathcal{B}}\sum_{b'\in\mathcal{B}}\sum_{y=0}^{p-1} e\left(\frac{(b-b')y}{p}\right) = 2\sum_{b\in\mathcal{B}}\sum_{y=0}^{p-1} 1 = 2|\mathcal{B}|p. \end{aligned} \tag{44}$$

By (43) and (44) we have

$$|\mathcal{B}|^2 p^{1/2} - 4|\mathcal{B}|p \le 2|\mathcal{B}|p$$

whence the statement of the lemma follows.

LEMMA 8 (Gallagher). *Let $\mathcal{A}$ be a set of integers in the interval $[M+1, M+N]$. For each prime p let $\nu(p)$ denote the number of residue classes modulo p that contain an element of $\mathcal{A}$. Then for any finite set of primes $\mathcal{P}$ we have*

$$|\mathcal{A}| \le \frac{\sum_{p\in\mathcal{P}} \log p - \log N}{\sum_{p\in\mathcal{P}} \frac{\log p}{\nu(p)} - \log N} \tag{45}$$

provided that the denominator is positive.

Proof of Lemma 8. This is Gallagher's "larger sieve" [6].

7. Proof of Theorem 6

Let $\mathcal{A} \subset \{1, 2, \dots, N\}$ be a set with property Q. Then for every $a, a' \in \mathcal{A}$, $a \ne a'$ there is an $x \in \mathbb{N}$ with

$$a + a' = x^2.$$

This clearly implies that for every prime p either $p \mid (a + a')$ or $(\frac{a+a'}{p}) = +1$. From each residue class modulo p that meets $\mathcal{A}$ pick an element of $\mathcal{A}$ and denote by $\mathcal{B}_p$ the set obtained in this way. Then clearly $\mathcal{B} = \mathcal{B}_p$ satisfies the conditions in Lemma 7 and thus we have

$$|\mathcal{B}_p| < 6p^{1/2}. \tag{46}$$

Write

$$\mathcal{P} = \{p : p \text{ prime}, p \le 36(\log N)^2\}$$

Defining $\nu(p)$ as in Lemma 8, by (46) for all $p \in \mathcal{P}$ we have

$$\nu(p) = |\mathcal{B}_p| < 6p^{1/2} \tag{47}$$

By Lemma 8, (45) holds. By (47) and the prime number theorem, for $N \to +\infty$ the denominator in (45) is

$$\begin{aligned}\sum_{p\in\mathcal{P}} \frac{\log p}{\nu(p)} - \log N &> \frac{1}{6} \sum_{p\le 36(\log N)^2} \frac{\log p}{p^{1/2}} - \log N \\ &= \left(\frac{1}{6} + o(1)\right) \int_2^{\frac{18(\log N)^2}{\log\log N}} \frac{\log u}{(u\log u)^{1/2}} du - \log N \\ &= (2 + o(1))\log N - \log N = (1 + o(1))\log N \end{aligned} \tag{48}$$

(which is positive) and the numerator is

$$\sum_{p\in\mathcal{P}} \log p - \log N = \sum_{p\le 36(\log N)^2} \log p - \log N = (36 + o(1))(\log N)^2. \tag{49}$$

(12) follows from (45), (48) and (49), and this completes the proof of Theorem 6.

References

1. H. Davenport, *Multiplicative number theory*, Markham Publishing Co., Chicago, first ed., 1967.
2. Y. Dupain, R. R. Hall, and G. Tenenbaum, *Sur l'équirépartition modulo 1 de certaines fonctions de diviseurs*, J. London Math. Soc. **26** (1982), 397–411.
3. P. Erdős, *Quelques problèmes de la théorie des nombres*, Monographie de l'enseignement mathématique, Genève, 1963, 81–135.
4. P. Erdős, C. L. Stewart, and R. Tijdeman, *Some Diophantine equations with many solutions*, Compositio Math. **66** (1988), 37–56.
5. G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), 349–366.
6. P. X. Gallagher, *A larger sieve*, Acta Arithmetica **18** (1971), 77–81.
7. J. Lagrange, *Six entiers dont les sommes deux à deux sont des carrés*, Acta Arithmetica **40** (1981), 91–96.
8. H. L. Montgomery and R. C. Vaughan, *Exponential sums with multiplicative coefficients*, Invent. Math. **43** (1977), 69–82.
9. J.-L. Nicolas, *Six nombres dont les sommes deux à deux sont des carrés, Calculateur en Math.* (1975, *Limoges*), Bulletin de la Société Mathématique de France, mémoire 49-50, (1977), 141–143.
10. C. Pomerance, A. Sárközy, and C. L. Stewart, *On divisors of sums of integers, III*, Pacific J. Math. **133** (1988), 363–379.
11. A. Sárközy, *On the number of prime factors of integers of the form $a_i + b_j$*, Studia Sci. Math. Hungar. **23** (1988), 161–168.
12. ———, "Hybrid problems in number theory" in *Number theory, New York* 1985–88, Lecture Notes in Mathematics, vol. 1383, Springer-Verlag, 1989.

13. A. Selberg, *Note on a paper by L. G. Sathe*, J. Indian Math. Soc. **18** (1954), 83–87.
14. W. Sierpiński, *A selection of problems in the theory of numbers*, Pergamon Press, Warszawa, first ed., 1964.
15. G. Tenenbaum, *Facteurs premiers de sommes d'entiers*, Proc. Amer. Math. Soc. **106** (1989), 287–296.
16. S. Wigert, *Sur l'ordre de grandeur du nombre de diviseurs d'un entier*, Ark. Mat. **3** (1906/1907), 1–9.

J. Rivat, Institut Girard Desargues, Université Lyon I, 43, bd du 11 novembre 1918, 69622 Villeurbanne, France
rivat@desargues.univ_lyon1.fr

C. L. Stewart, Department of Pure Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1
cstewart@watserv1.uwaterloo.ca

A. Sárközy, Department of Algebra and Number Theory, University Eötvös Loránd, Muzeum krt. 6-8, H-1088 Budapest, Hungary
sarkozy@cs.elte.hu