ILLINOIS JOURNAL OF MATHEMATICS
Volume 38, Number 1, Spring 1994

ON THE AVERAGE VALUE FOR THE NUMBER OF DIVISORS OF SUMS $a + b$

A. SÁRKÖZY[1] AND C.L. STEWART[2]

1. Introduction

For any set X we shall denote its cardinality by $|X|$. Let N be a positive integer and let A and B be subsets of $\{1, \dots, N\}$. In recent years several authors have investigated, subject to various assumptions on the cardinalities of A and B, the arithmetical character of the sums $a + b$ with a from A and b from B, see for instance [1], [3], [5], [6] and [8]. If A and B are sufficiently dense subsets of $\{1, \dots, N\}$ then many of the arithmetical properties of the sumset $A + B$ are similar to those of the set of consecutive integers $\{1, \dots, 2N\}$. In [3], Erdös, Maier and Sárközy developed this analogy by proving that if A and B are sufficiently dense then the sums $a + b$ with a from A and b from B satisfy a theorem of Erdös-Kac type. This work was refined later by Elliott and Sárközy [2] and by Tenenbaum [9]. For any positive integer n let $\omega(n)$ denote the number of distinct prime factors of n. In particular, it follows from [2] that if A and B are subsets of $\{1, \dots, N\}$ with

$$(|A|\,|B|)^{1/2} = N/\exp\Big(o\big((\log\log N)^{1/2}\log\log\log N\big)\Big) \tag{1}$$

then

$$\frac{1}{|A|\,|B|}\sum_{a\in A,\, b\in B}\omega(a+b) \sim \log\log N. \tag{2}$$

The asymptotic result (2) need not hold if (1) is replaced by the less stringent condition

$$(|A|\,|B|)^{1/2} > N/\exp(\delta\log\log N\log\log\log N),$$

Received November 19, 1991.
1991 Mathematics Subject Classification. Primary 11 N56; Secondary 11 B75.

[1]The research of the first author was partially supported by the Hungarian National Foundation for Scientific Research, Grant No. 1901.

[2]The research of the second author was supported in part by a Killam Research Fellowship and by a grant from the Natural Sciences and Engineering Research Council of Canada.

where δ is any positive real number, see [8]. Nevertheless Sárközy and Stewart [8] proved that, for each $\varepsilon > 0$,

$$\frac{1}{|A|\,|B|} \sum_{a\in A,\, b\in B} \omega(a+b) > (1-\varepsilon)\log\log N, \tag{3}$$

for N sufficiently large as A and B run over subsets of $\{1,\dots,N\}$ with

$$(|A|\,|B|)^{1/2} = N\exp\left(-(\log N)^{o(1)}\right). \tag{4}$$

For any positive integer n we denote the number of positive divisors of n by $\tau(n)$. In this article we shall investigate the average value of $\tau(a+b)$ as a and b run over the elements of A and B respectively where A and B are sufficiently dense subsets of $\{1,\dots,N\}$. In this context the τ function is more difficult to treat than the ω function for the following reasons. First the average of $\tau(a+b)$ over a and b grows exponentially more quickly than the average of $\omega(a+b)$ over a and b. Secondly the main contribution to the average

$$\frac{1}{|A|\,|B|} \sum_{a\in A,\, b\in B} \tau(a+b),$$

comes from a sparse set of pairs (a,b) for which $\tau(a+b)$ is large. This phenomenon also holds for the set of consecutive integers. By Theorem 319 of [7],

$$\frac{1}{n}\sum_{j=1}^{n} \tau(j) \sim \log n,$$

whereas it can be shown that, for each positive real number ε, the set of positive integers n for which

$$\tau(n) > (\log n)^{\log 2+\varepsilon},$$

is a set of positive upper density zero.

Since $\tau(n) \geq 2^{\omega(n)}$ for all positive integers n, we have from (3) and the arithmetic-geometric mean inequality that for each positive real number ε,

$$\frac{1}{|A|\,|B|} \sum_{a\in A,\, b\in B} \tau(a+b) > (\log N)^{\log 2-\varepsilon},$$

provided that N is sufficiently large and that A and B run over subsets of $\{1,\dots,N\}$ for which (4) holds. Our principal result is the following.

THEOREM 1. *Let ε be a positive real number, N be a positive integer and A and B be subsets of $\{1, \ldots, N\}$ with*

$$\min(|A|, |B|) > \varepsilon N. \tag{5}$$

There exist effectively computable positive constants C_0, C_1 and C_2 such that if N exceeds C_0 and

$$\exp\left(-C_1(\log N)^{1/2}\right) < \varepsilon < \frac{1}{8}, \tag{6}$$

then

$$\frac{1}{|A|\,|B|} \sum_{a \in A,\, b \in B} \tau(a + b) > \frac{C_2 \log N}{\left(\log\left(\frac{1}{\varepsilon}\right)\right)^5 \log\log\left(\frac{1}{\varepsilon}\right)}. \tag{7}$$

In particular is $|A| \gg N$ and $|B| \gg N$ then the average of $\tau(a + b)$ is $\gg \log N$, which is best possible as can be seen on taking $A = B = \{1, \ldots, N\}$. Moreover whenever ε tends to zero as N tends to infinity there exists a sequence of sets A and B satisfying (5) for which the average of $\tau(a + b)$ is $o(\log N)$, as our next result shows.

THEOREM 2. *There exist effectively computable positive constants C_3, C_4 and C_5 such that if N is a positive integer which exceeds C_3 and ε is a real number satisfying*

$$\exp(-\log N/\log\log N) < \varepsilon < C_4, \tag{8}$$

then there is a subset A of $\{1, \ldots, N\}$ with $|A| > \varepsilon N$ for which

$$\frac{1}{|A|^2} \sum_{a, a' \in A} \tau(a + a') < \frac{C_5 \log N}{\log\log\left(\frac{1}{\varepsilon}\right)}. \tag{9}$$

We suspect that the upper bound given by (9) is closer to the truth than the lower bound given by (7).

Our final result shows that if ε tends to zero as N tends to infinity there exists a sequence of sets A with $|A| > \varepsilon N$ for which

$$\frac{1}{|A|^2 \log N} \sum_{a, a' \in A} \tau(a + a') \to \infty.$$

THEOREM 3. *For each real number δ with $\delta > 0$ there are positive numbers C_6 and C_7, which are effectively computable in terms of δ, such that if N*

exceeds C_6 *and* ε *is a real number with*

$$N^{-1/8} < \varepsilon < C_7, \tag{10}$$

then there is a subset A *of* $\{1, \ldots, N\}$ *with*

$$|A| > \varepsilon N, \tag{11}$$

for which

$$\frac{1}{|A|^2} \sum_{a, a' \in A} \tau(a + a') > \left(\exp\left((1 - \delta)\log 2 \log\left(\frac{1}{\varepsilon}\right)/\log\log\left(\frac{1}{\varepsilon}\right)\right)\right)\log N. \tag{12}$$

While we have not worked out an upper bound for the average of $\tau(a + b)$ subject to (5) we suspect that (12) cannot be improved on substantially. In particular we conjecture that one cannot replace $-\delta$ in (12) by $+\delta$.

Finally we remark that since $\tau(n) \geq 2^{\omega(n)}$, estimates from below for the quantity

$$\max_{a \in A,\, b \in B} \tau(a + b)$$

may be deduced from lower estimates for the maximum of $\omega(a + b)$ as a and b run over A and B respectively. Such estimates have been obtained in two recent papers [4], [8]. The first paper [4] treats the case when $(|A|\,|B|)^{1/2} \gg N$ whereas the second [8] applies to much thinner sets.

2. Preliminary lemmas

LEMMA 1. *Let* u, v *and* k *be integers with* v *and* k *positive. There exists an effectively computable positive constant* C_8 *such that if*

$$v > C_8 e^{3k}, \tag{13}$$

and H *is a subset of* $\{u + 1, \ldots, u + v\}$ *with*

$$|H| > \left(1 - \frac{1}{4}\prod_{p \leq 2k}\left(1 - \frac{1}{p}\right)\right)v, \tag{14}$$

then there exist integers $d_1, d_2, \ldots, d_k$ *with* $d_i \in H$ *for* $i = 1, \ldots, k$ *for which* $(d_i, d_j) = 1$ *whenever* $i \neq j$.

Proof. We take

$$C_8 = \max_{k \geq 1} \left\{ e^{-3k} 12k \prod_{p \leq 2k} p \left(1 - \frac{1}{p}\right)^{-1} \right\} \tag{15}$$

and suppose that (13) and (14) hold. That C_8 is well defined follows from the prime number theorem and Mertens' theorem. Put

$$P = \prod_{p \leq 2k} p,$$

and let $H(h)$ denote the set of the terms of H which are congruent to h modulo P. We shall now show that there exists an integer h_0 which is coprime with P with

$$|H(h_0)| > \frac{2}{3} \frac{v}{P}. \tag{16}$$

This is so since otherwise

$$\begin{aligned} |H| &= \sum_{1 \leq h \leq P} |H(h)| = \sum_{\substack{1 \leq h \leq P \\ (h,P)>1}} |H(h)| + \sum_{\substack{1 \leq h \leq P \\ (h,P)=1}} |H(h)| \\ &= \sum_{\substack{1 \leq h \leq P \\ (h,P)>1}} \sum_{\substack{u < n \leq u+v \\ n \equiv h (\mathrm{mod}\, P)}} 1 + \sum_{\substack{1 \leq h \leq P \\ (h,P)=1}} \frac{2}{3} \frac{v}{P} \\ &< \sum_{\substack{1 \leq h \leq P \\ (h,P)>1}} \left(\frac{v}{P} + 1\right) + \sum_{\substack{1 \leq h \leq P \\ (h,P)=1}} \frac{v}{P} - \frac{1}{3} \frac{v}{P} \sum_{\substack{1 \leq h \leq P \\ (h,P)=1}} 1 \\ &\leq v + P - \frac{1}{3} \frac{v}{P} \left(P \prod_{p | P} \left(1 - \frac{1}{p}\right)\right) \end{aligned}$$

By (13) and (15) we conclude that

$$|H| \leq \left(1 - \frac{1}{4} \prod_{p \leq 2k} \left(1 - \frac{1}{p}\right)\right) v$$

which contradicts (14). Thus there is an integer h_0 satisfying (16) and coprime with P. Define m to be that integer which satisfies $m \equiv h_0 (\mathrm{mod}\, P)$

and $m \le u < m + P$. Then

$$H(h_0) = \bigcup_{l=1}^{[v/2kP]+1} \left(H \cap \{n : m + 2(l-1)kP < n \le m + 2lkP,\ n \equiv h_0 (\bmod P)\}\right)$$

and so, by (13), (15) and (16), there exists an integer l_0 such that

$$\begin{aligned}&|H \cap \{n : m + 2(l_0 - 1)kP < n \le m + 2l_0kP,\ n \equiv h_0 (\bmod P)\}| \\ &\quad > \frac{2}{3}\frac{v}{P}\left(\left[\frac{v}{2kP}\right] + 1\right)^{-1} > \frac{2}{3}\frac{v}{P}\left(\frac{4}{3}\frac{v}{2kP}\right)^{-1} = k.\end{aligned}$$

Thus there exist integers $d_1, d_2, \dots, d_k$ from H with

$$m + 2(l_0 - 1)kP < d_i \le m + 2l_0kP, \tag{17}$$

and

$$d_i \equiv h_0 (\bmod P), \tag{18}$$

for $i = 1, \dots, k$. If $1 \le i < j \le k$ then by (17) and (18), $d_i - d_j = yP$ where $0 < y < 2k$. Since h_0 is coprime with P so also are d_i and d_j. But all the prime divisors of $d_i - d_j$ are less than $2k$ and thus $(d_i, d_j) = 1$.

LEMMA 2. *Let δ and η be positive real numbers. Let k be a positive integer and let $d_1, \dots, d_{2k}$ be positive integers with $(d_i, d_j) = 1$ for $i \ne j$. Put $D = d_1 \dots d_{2k}$. Let R be a subset of $\{1, \dots, D\}$ and, for any integer j and for $i = 1, \dots, 2k$, let $R_i(j)$ denote the terms of R which are congruent to j modulo d_i. If there are k integers d_i with $1 \le i \le 2k$ for which there are at least δd_i integers j from $\{1, \dots, d_i\}$ with $|R_i(j)| < \eta D/d_i$ then*

$$|R| \le \left((1 - \delta)^k + \eta k\right)D.$$

Proof. We shall suppose, without loss of generality, that the k integers d_i with $1 \le i \le 2k$ for which there are at least δd_i integers j with $|R_i(j)| < \eta D/d_i$ are $d_1, \dots, d_k$. We write R as $R_1 \cup R_2$ where R_1 consists of those terms of R which are not congruent to any of the integers j with $|R_i(j)| < \eta D/d_i$ modulo d_i for $i = 1, \dots, k$ and R_2 is the balance of R. Then, by the Chinese Remainder Theorem, $|R_1| \le (1 - \delta)^k D$. Plainly

$$|R_2| < \sum_{i=1}^{k} d_i \eta \frac{D}{d_i} = \eta k D,$$

and the result follows.

LEMMA 3. *For each positive integer n, we have*

$$\sum_{d|n} \frac{\mu(d)\log d}{d} = -\prod_{p|n}\left(1 - \frac{1}{p}\right)\sum_{p|n}\frac{\log p}{p-1}. \tag{19}$$

Proof. For every complex number s we have

$$-\sum_{d|n}\frac{\mu(d)}{d^s} = -\prod_{p|n}\left(1 - \frac{1}{p^s}\right).$$

Differentiating we obtain

$$\sum_{d|n}\frac{\mu(d)\log d}{d^s} = -\prod_{p|n}\left(1 - \frac{1}{p^s}\right)\sum_{p|n}\frac{\left(1 - \frac{1}{p^s}\right)'}{1 - \frac{1}{p^s}}$$

$$= -\prod_{p|n}\left(1 - \frac{1}{p^s}\right)\sum_{p|n}\frac{\log p}{p^s - 1}.$$

Substituting $s = 1$, we obtain (19).

3. Proof of Theorem 1

We have

$$\sum_{a\in A,\, b\in B}\tau(a+b) = \sum_{d\le 2N}\sum_{\substack{a\in A,\, b\in B\\ d|(a+b)}} 1$$

$$= \sum_{x=0}^{[(\log N)/\log 2]+1}\ \sum_{2^x\le d<2^{x+1}}\ \sum_{\substack{a\in A,\, b\in B\\ d|(a+b)}} 1. \tag{20}$$

Take

$$k = \left[\frac{\log(C_{12}\varepsilon/\log(1/\varepsilon))}{\log(2/3)}\right] + 3$$

where the constant C_{12} will be defined by (31) and (32). By (20),

$$\sum_{a\in A,\, b\in B}\tau(a+b) > \sum_{x=[(\log N)/6k\log 2]}^{[(\log N)/3k\log 2]}\ \sum_{2^x\le d<2^{x+1}}\ \sum_{\substack{a\in A,\, b\in B\\ d|(a+b)}} 1. \tag{21}$$

Note that for $N \geq 2^{18k}$,

$$\left[\frac{\log N}{3k \log 2}\right] - \left[\frac{\log N}{6k \log 2}\right] > \frac{\log N}{7k}. \tag{22}$$

Put

$$\kappa = \frac{1}{4} \prod_{p \leq 4k} \left(1 - \frac{1}{p}\right). \tag{23}$$

For each integer x with

$$\left[\frac{\log N}{6k \log 2}\right] \leq x \leq \left[\frac{\log N}{3k \log 2}\right], \tag{24}$$

we shall prove that for at least $\kappa 2^{2x}$ integers d with $2^x \leq d < 2^{x+1}$,

$$\sum_{\substack{a \in A,\, b \in B \\ d \mid (a+b)}} 1 > C_{14}(\log(1/\varepsilon))^{-4} \frac{|A|\,|B|}{d} \tag{25}$$

where C_{14} will be defined by (34). It then follows from (21), (22), (23), (24) and (25) that

$$\sum_{a \in A,\, b \in B} \tau(a+b) > \frac{C_{14}}{14}(\log(1/\varepsilon))^{-4} \frac{\kappa}{k}|A|\,|B| \log N,$$

and employing Mertens' theorem we deduce our result.

Accordingly, suppose that x is an integer satisfying (24) for which there are less than $\kappa 2^x$ integers d with $2^x \leq d < 2^{x+1}$ satisfying (25). Let H_x be the set of integers d with $2^x \leq d < 2^{x+1}$ for which (25) fails. Then $|H_x| > (1 - \kappa)2^x$. There exist effectively computable positive constants C_0 and C_1 such that if N exceeds C_0 and (6) holds then

$$2^x \geq 2^{[(\log N)/(6k \log 2)]} > N^{1/7k} > C_8 e^{6k}.$$

Thus we may apply Lemma 1 with $u = 2^x - 1$, $v = 2^x$ to deduce that there are $2k$ integers $d_1, \ldots, d_{2k}$ in H_x with $(d_i, d_j) = 1$ whenever $i \neq j$. Put $D = d_1 \ldots d_{2k}$ and let $F(n)$ and $G(n)$ denote the number of integers a in A with $a \equiv n (\mathrm{mod}\ D)$, and the number of integers b in B with $b \equiv n (\mathrm{mod}\ D)$, respectively. Thus $F(n) \leq N/D + 1 \leq 2N/D$ and similarly, $G(n) \leq 2N/D$. Write

$$\mathscr{R}(A, t) = \{n \colon 1 \leq n \leq D,\, F(n) \geq t\}$$

and

$$\mathscr{R}(B,t) = \{n: 1 \le n \le D, G(n) \ge t\}.$$

We obtain, by partial summation, that

$$\begin{aligned} |A| &= \sum_{1 \le n \le D} F(n) \le \sum_{\substack{1 \le n \le D \\ F(n) \le |A|/2D}} \frac{|A|}{2D} + \sum_{\substack{1 \le n \le D \\ F(n) > |A|/2D}} F(n) \\ &\le D \cdot \frac{|A|}{2D} + \sum_{|A|/2D < t \le 2N/D} t(|\mathscr{R}(A,t)| - |\mathscr{R}(A,t+1)|) \\ &= \frac{|A|}{2} + \sum_{|A|/2D + 1 < t \le 2N/D} |\mathscr{R}(A,t)| + ([|A|/2D] + 1) \\ &\quad |\mathscr{R}(A,[|A|/2D] + 1)|. \end{aligned}$$

We now put

$$M_A = \max_{|A|/2D < t \le 2N/D} t|\mathscr{R}(A,t)|. \tag{26}$$

Thus we have

$$\frac{|A|}{2} < \sum_{|A|/2D + 1 < t \le 2N/D} \frac{M_A}{t} + M_A.$$

$C_9, C_{10}, \ldots$ will denote effectively computable positive constants. Then, by (5),

$$\begin{aligned} \frac{|A|}{2} &< M_A\left(\log\left(\frac{2N/D}{|A|/2D}\right) + C_9\right) = M_A(\log(4N/|A|) + C_9) \\ &< M_A(\log(1/\varepsilon) + C_{10}) \end{aligned}$$

whence

$$M_A > C_{11}|A|(\log(1/\varepsilon))^{-1}.$$

Similarly, writing

$$M_B = \max_{|B|/2D < t \le 2N/D} t|\mathscr{R}(B,t)|, \tag{27}$$

we have

$$M_B > C_{11}|B|(\log(1/\varepsilon))^{-1}.$$

Let t_A, respectively t_B, denote an integer t for which the maximum in (26), respectively (27), is attained so that

$$|A|/2D < t_A \le 2N/D, \quad |B|/2D < t_B \le 2N/D, \tag{28}$$

$$t_A|\mathscr{R}(A, t_A)| = M_A > C_{11}|A|(\log(1/\varepsilon))^{-1} \tag{29}$$

and

$$t_B|\mathscr{R}(B, t_B)| = M_B > C_{11}|B|(\log(1/\varepsilon))^{-1}. \tag{30}$$

Then

$$\begin{aligned} |\mathscr{R}(A, t_A)| &> t_A^{-1}C_{11}|A|(\log(1/\varepsilon))^{-1} \\ &\ge \frac{D}{2N}C_{11}|A|(\log(1/\varepsilon))^{-1} > C_{12}\varepsilon(\log(1/\varepsilon))^{-1}D \end{aligned} \tag{31}$$

and similarly,

$$|\mathscr{R}(B, t_B)| > C_{12}\varepsilon(\log(1/\varepsilon))^{-1}D. \tag{32}$$

We now apply Lemma 2 with $\delta = 1/3$ and $\eta = \eta_A = |\mathscr{R}(A, t_A)|/2kD$. Note that, in view of (31),

$$\begin{aligned} ((1-\delta)^k + \eta k)D &= \left(\left(\frac{2}{3}\right)^k + \frac{|\mathscr{R}(A, t_A)|}{2D}\right)D \\ &< C_{12}\varepsilon(\log(1/\varepsilon))^{-1}\left(\frac{2}{3}\right)^2 D + \frac{|\mathscr{R}(A, t_A)|}{2} \\ &< |\mathscr{R}(A, t_A)|. \end{aligned}$$

Thus by Lemma 2, we conclude that there are at most $k-1$ integers d_i with $1 \le i \le 2k$ for which there are at least $\frac{1}{3}d_i$ integers j from $\{1, \ldots, d_i\}$ with

$$|\{n : n \in \mathscr{R}(A, t_A), n \equiv j (\text{mod } d_i)\}| < \eta_A \frac{D}{d_i}.$$

Put $\eta_B = |\mathscr{R}(B, t_B)|/2kD$. A similar result holds on replacing $\mathscr{R}(A, t_A)$ and η_A by $\mathscr{R}(B, t_B)$ and η_B respectively. Thus there exists an integer d_i from $\{d_1, \ldots, d_{2k}\}$, d_1 say, for which there are at most $\frac{1}{3}d_i$ integers j from

$\{1, \ldots, d_i\}$ with

$$\left|\{n\colon n \in \mathscr{R}(A, t_A),\, n \equiv j \pmod{d_i}\}\right| < \eta_A \frac{D}{d_i}$$

and at most $\frac{1}{3}d_i$ integers j from $\{1, \ldots, d_i\}$ with

$$\left|\{n\colon n \in \mathscr{R}(B, t_B),\, n \equiv j \pmod{d_i}\}\right| < \eta_B \frac{D}{d_i}.$$

But then

$$\begin{aligned}
\sum_{\substack{a\in A,\, b\in B\\ d_1|(a+b)}} 1 &\geq \sum_{n=1}^{d_1}\Bigg(\sum_{\substack{u\in\mathscr{R}(A,t_A)\\ u\equiv n \pmod{d_1}}}\ \sum_{\substack{a\in A\\ a\equiv u \pmod{D}}} 1\Bigg)\Bigg(\sum_{\substack{v\in\mathscr{R}(B,t_B)\\ v\equiv -n \pmod{d_1}}}\ \sum_{\substack{b\in B\\ b\equiv v \pmod{D}}} 1\Bigg)\\
&\geq \sum_{n=1}^{d_1}\Bigg(\sum_{\substack{u\in\mathscr{R}(A,t_A)\\ u\equiv n \pmod{d_1}}} t_A\Bigg)\Bigg(\sum_{\substack{v\in\mathscr{R}(B,t_B)\\ v\equiv -n \pmod{d_1}}} t_B\Bigg)\\
&= \sum_{n=1}^{d_1} t_A t_B\Bigg(\sum_{\substack{u\in\mathscr{R}(A,t_A)\\ u\equiv n \pmod{d_1}}} 1\Bigg)\Bigg(\sum_{\substack{v\in\mathscr{R}(B,t_B)\\ v\equiv -n \pmod{d_1}}} 1\Bigg). \qquad (33)
\end{aligned}$$

For at least $\frac{2}{3}$ of the residue classes n from $1, \ldots, d_1$,

$$\sum_{\substack{u\in\mathscr{R}(A,t_A)\\ u\equiv n \pmod{d_1}}} 1 \geq \eta_A \frac{D}{d_1}$$

and for at least $\frac{2}{3}$ of the residue classes $-n$ from $1, \ldots, d_1$,

$$\sum_{\substack{v\in\mathscr{R}(B,t_B)\\ v\equiv -n \pmod{d_1}}} 1 \geq \eta_B \frac{D}{d_1}.$$

Thus, by (29) and (30),

$$\sum_{n=1}^{d_1} t_A t_B \left(\sum_{\substack{u \in \mathscr{R}(A, t_A) \\ u \equiv n (\text{mod } d_1)}} 1 \right) \left(\sum_{\substack{v \in \mathscr{R}(B, t_B) \\ v \equiv -n (\text{mod } d_1)}} 1 \right)$$

$$\geq \frac{1}{3} \eta_A \eta_B t_A t_B \frac{D^2}{d_1} = \frac{1}{12} \frac{1}{k^2 d_1} t_A |\mathscr{R}(A, t_A)| t_B |\mathscr{R}(B, t_B)| \cdot$$

$$> C_{13}(\log(1/\varepsilon))^{-2} \frac{|A|\,|B|}{k^2 d_1} > C_{14}(\log(1/\varepsilon))^{-4} \frac{|A|\,|B|}{d_1} \tag{34}$$

By (33) and (34), $d_1 \in H_x$ contrary to our assumption. Our result now follows.

4. Proof of Theorem 2

$C_{15}, C_{16}, \ldots$ will denote positive effectively computable constants. Also denote the ith prime by p_i, so $p_1 = 2$, $p_2 = 3, \ldots,$ and for $n = 1, 2, \ldots,$ and $i = 1, 2, \ldots,$ define the integer $r_i(n)$ by

$$r_i(n) \equiv n (\text{mod } p_i), \qquad 0 \leq r_i(n) < p_i.$$

Let $t = [\frac{1}{4} \log(1/\varepsilon)]$ and $P = \prod_{i=2}^{t} p_i$. Then, by the prime number theorem and (8),

$$P < 3^{t \log t} < \exp\left(\frac{1}{2} \log(1/\varepsilon) \log\log(1/\varepsilon) \right) < \sqrt{N}, \tag{35}$$

for $\varepsilon < C_{15}$. Define A by

$$A = \left\{ a \colon 1 \leq a \leq N, 0 < r_i(a) < \frac{p_i}{2} \text{ for } i = 2, \ldots, t \right\}.$$

Then, for $N > C_{16}$,

$$|A| > \frac{1}{2} N \prod_{i=2}^{t} \frac{p_i - 1}{2 p_i} = 2^{-t} N \prod_{i=2}^{t} \left(1 - \frac{1}{p_i} \right),$$

by the Chinese Remainder Theorem. Thus

$$|A| > 3^{-t}N > \exp(-\log(1/\varepsilon))N = \varepsilon N,$$

for $N > C_{16}$. Moreover we have

$$\begin{aligned}\sum_{a,a'\in A} \tau(a+a') &= \sum_{a,a'\in A}\ \sum_{d|(a+a')} 1 \\ &\le \sum_{a,a'\in A} 2 \sum_{\substack{d|(a+a')\\ d\le\sqrt{N}}} 1 \le 2 \sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}}\ \sum_{\substack{a,a'\in A\\ d|(a+a')}} 1 \\ &= 2\sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}}\ \sum_{j=1}^{d}\Bigg(\sum_{\substack{a'\in A\\ a'\equiv j(\mathrm{mod}\ d)}} 1\Bigg)\Bigg(\sum_{\substack{a'\in A\\ a'\equiv -j(\mathrm{mod}\ d)}} 1\Bigg), \qquad (36)\end{aligned}$$

since by the construction of the set A, if a and a' are from A and d divides $a+a'$ then d and P are coprime. By (35) and the Chinese Remainder Theorem, for each positive integer d up to $\sqrt{N}$ which is coprime with P and each integer j,

$$\begin{aligned}\sum_{\substack{a\in A\\ a\equiv j(\mathrm{mod}\ d)}} 1 \\ &= |\{a\colon 1\le a\le N,\, 0 < r_i(a) < p_i/2 \text{ for } i=2,\dots,t,\, a\equiv j(\mathrm{mod}\ d)\}| \\ &\le 2\frac{1}{d}|\{a\colon 1\le a\le N,\, 0< r_i(a) < p_i/2 \text{ for } i=2,\dots,t\}| \\ &= 2\frac{|A|}{d}.\end{aligned}$$

Thus it follows from (36) that

$$\begin{aligned}\sum_{a,a'\in A}\tau(a+a') &\le 2\sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}}\ \sum_{j=1}^{d}\left(2\frac{|A|}{d}\right)^2 \\ &= 8|A|^2\sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}}\frac{1}{d}. \qquad (37)\end{aligned}$$

Observe that

$$\begin{aligned}
\sum_{\substack{d \le \sqrt{N} \\ (d,P)=1}} \frac{1}{d} &= \sum_{d \le \sqrt{N}} \Big(\sum_{D | (d,P)} \mu(D) \Big) \frac{1}{d} \\
&= \sum_{D|P} \mu(D) \sum_{k \le \sqrt{N}/D} \frac{1}{Dk} = \sum_{D|P} \frac{\mu(D)}{D} \sum_{k \le \sqrt{N}/D} \frac{1}{k} \\
&\le \sum_{D|P} \frac{\mu(D)}{D} \log\big(\sqrt{N}/D\big) \\
&\quad + \sum_{D|P} \frac{|\mu(D)|}{D} \Bigg| \sum_{k \le \sqrt{N}/D} \frac{1}{k} - \log\big(\sqrt{N}/D\big) \Bigg| \\
&\le \frac{1}{2} \log N \sum_{D|P} \frac{\mu(D)}{D} - \sum_{D|P} \frac{\mu(D) \log D}{D} + C_{17} \sum_{D|P} \frac{|\mu(D)|}{D}.
\end{aligned}$$

Thus, by Lemma 3,

$$\begin{aligned}
\sum_{\substack{d \le \sqrt{N} \\ (d,P)=1}} \frac{1}{d} &\le \frac{1}{2} \log N \prod_{i=2}^{t} \left(1 - \frac{1}{p_i}\right) + \prod_{i=2}^{t} \left(1 - \frac{1}{p_i}\right) \sum_{i=2}^{t} \frac{\log p_i}{p_i - 1} \\
&\quad + C_{17} \prod_{i=2}^{t} \left(1 + \frac{1}{p_i}\right).
\end{aligned}$$

By Mertens' theorem and the prime number theorem,

$$\begin{aligned}
\sum_{\substack{d \le \sqrt{N} \\ (d,P)=1}} \frac{1}{d} &< C_{18} \left(\prod_{i=2}^{t} \left(1 - \frac{1}{p_i}\right) \left(\log N + \sum_{i=2}^{t} \frac{1}{i} \right) + \prod_{i=2}^{t} \left(1 - \frac{1}{p_i}\right)^{-1} \right) \\
&< C_{19} \big((\log t)^{-1} (\log N + \log t) + \log t \big) \\
&< C_{20} \big(\log\log(1/\varepsilon) \big)^{-1} \log N. \qquad (38)
\end{aligned}$$

We obtain (9) from (37) and (38). This completes the proof of Theorem 2.

5. Proof of Theorem 3

As before for each positive integer i let p_i denote the i-th prime number. Let δ be a positive real number. $C_{21}, C_{22}, \ldots$ will denote positive numbers which are effectively computable in terms of δ. Suppose that ε is a real number satisfying (10) and define the positive integer k by the inequalities

$$p_1 \dots p_k \le \frac{1}{2\varepsilon} < p_1 \dots p_{k+1}. \tag{39}$$

Put

$$P = p_1 \dots p_k,$$

and define

$$A = \{n \colon 1 \le n \le N, P|n\}.$$

By (10), (30) and the prime number theorem

$$P < N^{1/8}, \tag{40}$$

and

$$|A| = (1+o(1))\frac{N}{P} > \frac{N}{2P} \ge \varepsilon N, \tag{41}$$

provided that N exceeds C_{21}. Thus (11) holds. It remains to verify (12).

Plainly

$$\sum_{a, a' \in A} \tau(a+a') = \sum_{u, v \le N/P} \tau(P(u+v)).$$

We shall restrict our attention to those pairs (u, v) of positive integers less than or equal to N/P for which d_1, the greatest common divisor of $u+v$ and P^2, is square-free. For such a pair (u, v) there is a unique integer t such that

$$u + v \equiv d_1 t \pmod{P^2},\ 1 \le t \le P^2/d_1 \text{ and } \left(t, P^2/d_1\right) = 1.$$

Thus

$$\sum_{a, a' \in A} \tau(a+a') \ge \sum_{d_1|P} \prod_{\substack{1 \le t \le P^2/d_1 \\ (t, P^2/d_1)=1}} \sum_{\substack{u, v \le N/P \\ u+v \equiv d_1 t \pmod{P^2}}} \tau(P(u+v)). \tag{42}$$

Observe that since $d_1|P$ and $(t, P^2/d_1) = 1$ then

$$\begin{aligned}\sum_{\substack{u,v\le N/P\\ u+v\equiv d_1t \pmod{P^2}}} \tau(P(u+v)) &\ge \sum_{\substack{N/2P<m\le N/P\\ m\equiv d_1t \pmod{P^2}}} \tau(Pm) \sum_{\substack{u,v\le N/P\\ u+v=m}} 1\\ &\ge \frac{N}{2P} \sum_{\substack{N/2P<m\le N/P\\ m\equiv d_1t \pmod{P^2}}} \tau(d_1^2)\tau\left(\frac{P}{d_1}\right)\tau\left(\frac{m}{d_1}\right)\\ &= \frac{N}{2P} 3^{\omega(d_1)}2^{k-\omega(d_1)} \sum_{\substack{N/2P<m\le N/P\\ m\equiv d_1t \pmod{P^2}}} \sum_{\substack{d|m\\ (d,P)=1}} 1\\ &= \frac{N}{2P}\left(\frac{3}{2}\right)^{\omega(d_1)} 2^k \sum_{\substack{d\le N/P\\ (d,P)=1}} \sum_{\substack{N/2Pd<z\le N/Pd\\ dz\equiv d_1t \pmod{P^2}}} 1.\end{aligned}$$

Thus, by (40),

$$\begin{aligned}\sum_{\substack{u,v\le N/P\\ u+v\equiv d_1t \pmod{P^2}}} \tau(P(u+v)) &\ge \frac{N}{2P}\left(\frac{3}{2}\right)^{\omega(d_1)} 2^k \sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}} \sum_{\substack{N/2Pd<z\le N/Pd\\ dz\equiv d_1t \pmod{P^2}}} 1\\ &\ge \frac{N}{2P}\left(\frac{3}{2}\right)^{\omega(d_1)} 2^k \sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}} \left(\frac{N}{2P^3d} - 1\right),\\ &\ge \frac{N^2}{8P^4}\left(\frac{3}{2}\right)^{\omega(d_1)} 2^k \sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}} \frac{1}{d}, \end{aligned} \tag{43}$$

whenever N exceeds C_{22}. As in the proof of Theorem 2 we deduce that

$$\begin{aligned}\sum_{\substack{d\le\sqrt{N}\\ (d,P)=1}} \frac{1}{d} \ge \frac{1}{2}(\log N)\prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right) + \prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)\sum_{i=1}^{k}\frac{\log p_i}{p_i-1}\\ - C_{23}\prod_{i=1}^{k}\left(1+\frac{1}{p_i}\right).\end{aligned}$$

Thus, by Mertens' theorem, the prime number theorem, (10) and (39),

$$\sum_{\substack{d \le \sqrt{N} \\ (d,P)=1}} \frac{1}{d} \ge \frac{1}{4}(\log N) \prod_{p|P}\left(1 - \frac{1}{p}\right), \tag{44}$$

whenever N exceeds C_{24}.

Therefore, by (42), (43) and (44),

$$\begin{aligned}\sum_{a,a' \in A} \tau(a+a') &\ge \sum_{d_1|P} \sum_{\substack{1 \le t \le P^2/d_1 \\ (t,P^2/d_1)=1}} \frac{N^2 \log N}{32P^4}\left(\frac{3}{2}\right)^{\omega(d_1)} 2^k \prod_{p|P}\left(1 - \frac{1}{p}\right) \\ &= \frac{N^2 \log N}{32P^2} 2^k \prod_{p|P}\left(1 - \frac{1}{p}\right)^2 \sum_{d_1|P} \frac{1}{d_1}\left(\frac{3}{2}\right)^{\omega(d_1)}, \end{aligned} \tag{45}$$

for N greater than C_{25}. Note that

$$\begin{aligned}\sum_{d_1|P} \frac{1}{d_1}\left(\frac{3}{2}\right)^{\omega(d_1)} &= \prod_{p|P}\left(1 + \frac{3}{2p}\right) \ge C_{26} \prod_{p|P}\left(1 + \frac{1}{p}\right)^{3/2} \\ &\ge C_{27} \prod_{p|P}\left(1 - \frac{1}{p}\right)^{-3/2}. \end{aligned} \tag{46}$$

It now follows from (41), (45), (46), Mertens' theorem and the prime number theorem that

$$\sum_{a,a' \in A} \tau(a+a') \ge C_{28}|A|^2(\log N)2^k(\log k)^{-1/2}, \tag{47}$$

for N greater than C_{25}. By (39) and the prime number theorem

$$k > \left(1 - \frac{\delta}{2}\right)\log(1/\varepsilon)/\log\log(1/\varepsilon),$$

provided that $\varepsilon < C_{29}$, and so (12) follows from (47).

References

1. A. Balog and A. Sárközy, *On sums of sequences of integers*, I, Acta Arith. **44** (1984), 73–86.
2. P.D.T.A. Elliott and A. Sárközy, *The distribution of the number of prime divisors of sums $a+b$*, J. Number Theory **29** (1988), 94–99.
3. P. Erdös, H. Maier and A. Sárközy, *On the distribution of the number of prime factors of sums $a+b$*, Trans. Amer. Math. Soc. **302** (1987), 269–280.

4. P. ERDÖS, C. POMERANCE, A. SÁRKÖZY and C.L. STEWART, *On elements of sumsets with many prime factors*, to appear.
5. P. ERDÖS, C.L. STEWART and R. TIJDEMAN, *Some diophantine equations with many solutions*, Compositio Math. **66** (1988), 37–56.
6. K. GYÖRY, C.L. STEWART and R. TIJDEMAN, *On prime factors of sums of integers, I*, Compositio Math. **59** (1986), 81–88.
7. G.H. HARDY and E.M. WRIGHT, *An introduction to the theory of numbers*, 5-th ed., Oxford University Press, Oxford, 1979.
8. A. SÁRKÖZY and C.L. STEWART, *On divisors of sums of integers, V*, to appear.
9. G. TENENBAUM, *Facteurs premiers de sommes d'entiers*, Proc. Amer. Math. Soc. **106** (1989), 287–296.

THE UNIVERSITY OF WATERLOO
WATERLOO, ONTARIO, CANADA

MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCE
BUDAPEST, HUNGARY