ON THE SOLVABILITY OF CERTAIN EQUATIONS IN SEQUENCES OF POSITIVE UPPER LOGARITHMIC DENSITY

Dedicated to Professor L. J. Mordell on his 80th birthday

P. ERDŐS, A. SÁRKÖZI AND E. SZEMERÉDI

Let $a_1 < a_2 < \ldots$ be a sequence of integers, to be denoted by A, satisfying

$$\lim_{x\to\infty} \sup \frac{1}{\log x} \sum_{a_i < x} \frac{1}{a_i} = \alpha > 0. \tag{1}$$

A sequence satisfying (1) is said to have positive upper logarithmic density. Davenport and Erdős [2] proved that every sequence satisfying (1) contains an infinite division chain, in other words an infinite subsequence a_{i_j}, j = 1, 2, . .., satisfying $a_{i_j} | a_{i_{j+1}}$. P. Erdős, A. Sárközi and E. Szemerédi [3] proved that if (1) is satisfied then there are infinitely many distinct quadruplets of distinct integers a_i, a_j, *a,, a,* satisfying

$$(a_i, a_j) = a,, \quad [a_i, a_j] = a,.$$

In fact this is deduced in [3] from a weaker hypothesis than (1). In [3] we used an ingenious combinatorial theorem of Kleitman [4]. By the same method as used in [3] we could obtain the following result: For every k there is an η such that if the sequence A satisfies for infinitely many x

$$\sum_{a_i < x} \frac{1}{a_i} > \frac{x}{(\log \log x)^\eta}$$

then there is a k-tuple a;,, ... a_{i_k}, of which no a_{i_r} divides any other, such that all the integers

$$(a_{i_{r_1}}, a_{i_{r_2}}), \quad [a_{i_{r_1}}, a_{i_{r_2}}], \quad 1 \leqslant r_1 < r_2 \leqslant k$$

are in A.

This result suggests the following conjecture (which in fact was stated in [3]): If A is a sequence satisfying (1) then there exists an infinite subsequence $a_{i_j} \in A$, of which no *a,,* divides any other, such that all the integers

$$(a_{i_{j_1}}, a_{i_{j_2}}) \quad \text{and} \quad [a_{i_{j_1}}, a_{i_{j_2}}], \quad 1 \leqslant j_1 < j_2$$

occur in *A*.

In this note we prove this conjecture and in fact we prove considerably more, In fact we establish the following result, which seems to be definitive:

THEOREM 1. *Let A satisfy (1). Then there is an infinite subsequence $a_{i_j} \in A$, $1 \leqslant j < \infty$ such that both the greatest common divisor and the least common multiple of any set of a_{i_j}'s is in A and the least common multiples of any two distinct sets of a_{i_j}'s are distinct.*

Received 2 October, 1967.

[J. LONDON MATH. SOC., 43 (1968), 71–78]

Theorem 1 implies that no two a_{i_j}'s can divide each other. For if $a_{i_{j_1}} | a_{i_{j_2}}$ then $a_{i_{j_2}} = [a_{i_{j_2}}, a_{i_{j_1}}]$ which is impossible.

Our proof of Theorem 1 will not use the results of Kleitman [4].

Theorem 1 will follow fairly easily from the following:

THEOREM 2. *Let $p(q)$ denote the least prime factor of q. Let A satisfy* (1). *Then there are integers $a_u \in A$, $a_v \in A$, $a_u | a_v$ and a sequence $q_1 < q_2 < \dots$ of positive upper logarithmic density satisfying*

$$p(q_r) > a_v, \quad a_u q_r \in A, \quad a_v q_r \in A, \; r = 1, 2, \dots . \tag{2}$$

The proof of Theorem 2 will be our main difficulty. Assuming that Theorem 2 has already been proved we deduce Theorem 1 as follows. (The proof may seem complicated because of the many indices but is really almost obvious.)

Put $a_1 = a_{i_1}$, in other words a_1 is the first term of our sequence a_{i_j}, $j = 1, 2, \dots$. It will be convenient to put $q_r = a_r^{(1)}$, $r = 1, 2, \dots$ and to denote the sequence $a_r^{(1)}$, $r = 1, 2, \dots$ by A_1. A_1 has positive upper logarithmic density. By our construction and (2) we evidently have

$$(a_{i_1}, a_u a_r^{(1)}) = a_u \in A, \quad [a_{i_1}, a_u a_r^{(1)}] = a_v a_r^{(1)} = a_{i_1} a_r^{(1)} \in A, r = 1, 2, \dots . \tag{3}$$

All further members of the sequence a_{i_j}, $j \geq 2$ will be selected from the integers $a_u a_r^{(1)}$, $r = 1, 2, \dots$.

We now apply Theorem 2 to A_1. Thus there are integers $a_u^{(1)} \in A_1$, $a_v^{(1)} \in A_1$, $a_u^{(1)} | a_v^{(1)}$ and a sequence $q_1^{(1)} < q_2^{(1)} \dots$ of positive upper logarithmic density satisfying

$$p(q_r^{(1)}) > a_v^{(1)}, \quad a_u^{(1)} q_r^{(1)} \in A_1, \quad a_v^{(1)} q_r^{(1)} \in A_1, \quad r = 1, 2, \dots . \tag{4}$$

Put

$$a_{i_2} = a_u a_v^{(1)} \qquad (p(a_v^{(1)}) > a_v > a_u)$$

and $q_r^{(1)} = a_r^{(2)}$, $r = 1, 2, \dots$. The sequence of $a_r^{(2)}$ is denoted by A_2. A_2 has positive upper logarithmic density. All further members of the sequence a_{i_j}, $j \geq 3$ will be selected from the integers $a_u a_u^{(1)} a_r^{(2)}$, $r = 1, 2, \dots$. It is easy to see that all four integers

$$a_u a_u^{(1)}, \quad a_u a_v^{(1)}, \quad a_v a_u^{(1)}, \quad a_v a_v^{(1)}$$

are in A.

Our construction can clearly be carried on indefinitely and we obtain an infinite set of sequences of positive upper logarithmic density: A_j, $j = 0, 1, \dots$ $(A_0 = A)$. The elements of A_j are $a_r^{(j)}$, $r = 1, 2, \dots$. Further we have for every j two integers in A_j, $a_u^{(j)}$, $a_v^{(j)}$ satisfying

$$a_u^{(j)} | a_v^{(j)}, \quad a_u^{(j)} a_r^{(j+1)} \in A_j, \quad a_v^{(j)} a_r^{(j+1)} \in A_j, \quad p(a_r^{(j+1)}) > a_v^{(j)}. \tag{5}$$

Put

$$a_{i_{j+1}} = \prod_{s=0}^{j-1} a_u^{(s)} a_v^{(j)} \qquad (a_u^{(0)} = a_u). \tag{6}$$

By our construction it is easy to see that all the 2^{j+1} products

$$\prod_{s=0}^{j} a_{\lambda_s}^{(s)}, \quad \lambda_s = u \text{ or } \lambda_s = v \tag{7}$$

are in A for every j. Also, for $l > j$, a_{i_l} will be selected from the integers

$$\prod_{s=0}^{j-1} a_u^{(s)} a_r^{(j)}, \quad r = 1, 2, \ldots .$$

Finally it easily follows from (5), (6) and (7) that

$$(a_{i_{j_1}}, \ldots, a_{i_{j_l}}) = \prod_{s=0}^{j_1} a_u^{(s)} \in A$$

and

$$[a_{i_{j_1}}, \ldots, a_{i_{j_l}}] = \prod_{t=1}^{l} a_v^{(j_t)} \Pi' a_u^{(s)} \in A, \tag{8}$$

where, in Π', $0 \leqslant s \leqslant j_l$, $s \neq j_t$, $1 \leqslant t \leqslant l$.

From (5) we have

$$p(a_u^{(s)}) > a_v^{(s-1)}, \quad p(a_v^{(s)}) > a_v^{(s-1)}.$$

Thus the expressions (7) are distinct for distinct sequences $j_1 < \ldots < j_l$. Hence the proof of Theorem 1 is complete.

Thus we only have to prove Theorem 2. First we have to introduce some notations. Denote by $A(a_i, x, y)$ the set of integers $q < y/a_i$, $p(q) > x$ for which $a_i q \in A$. A set $A' \subset A$ is said to have property $P(x, y, \varepsilon)$ if for every $a_i \in A'$, $a_i < x$ we have

$$\sum_{q \in A(a_i, x, y)} \frac{1}{q} > \varepsilon \log y / \log x. \tag{9}$$

LEMMA 1. *Let A satisfy (1). Then there are arbitrarily large values of x and an infinite sequence $y_1 < y_2 < \ldots$ (depending on x) such that*

$$\sum\nolimits_{(y_j, x)} \frac{1}{a_i} > \frac{\alpha^2}{100} \log x, \tag{10}$$

where in $\sum_{(y_j,x)}$ the summation is extended over the a_i having property $P\left(x, y_j, \frac{\alpha^2}{100}\right)$.

$\frac{\alpha^2}{100}$ is not best possible, but any positive number depending only on α would serve our purpose equally well.

The proof of Lemma 1 is the most difficult step of our proof. Put $\varepsilon = \frac{\alpha^2}{100}$ and assume that our lemma is false. Then to every x there is an $f(x)$ so that for every $y > f(x)$

$$\sum\nolimits_{(y,x)} \frac{1}{a_i} \leqslant \varepsilon \log x \qquad (a_i < x \text{ and } a_i \text{ satisfies } P(x, y, \varepsilon)). \tag{11}$$

From (1) and (11) it easily follows that there is an infinite sequence $x_1 < x_2 < \ldots$ satisfying

$$\sum_{a_i < x_j} \frac{1}{a_i} > (\alpha - \varepsilon) \log x_j \tag{12}$$

and

$$\sum\nolimits_{(x_r, x_j)} \frac{1}{a_i} \leqslant \varepsilon \log x_j \tag{13}$$

for every $r > j$.

To prove (12) and (13) it suffices to observe that by (1) there are arbitrarily large values of x satisfying (12), and (13) follows from (11) if we choose $x_{j+1} > f(x_j)$.

Now we show that (12) and (13) lead to a contradiction, and this will complete the proof of Lemma 1.

Let $l = [4\alpha^{-1}]+2$ and let $x_1 < \dots < x_l$ satisfy (12) and (13), where we further assume that x_1 is sufficiently large and x_{r+1} is a sufficiently large number satisfying

$$x_{r+1} > \max\left(f(x_r), e^{x_r}\right), \quad 2 \leq r \leq l-1. \tag{14}$$

Denote by $a_i^{(1)}$, $i = 1, 2, \dots$ the $a \in A$ not exceeding x_1 and by $a_i^{(r)}$, $2 \leq r \leq l$, the integers $a \in A$ in (x_{r-1}, x_r) which cannot be written in the form

$$a_i q, \quad a_i < x_j, \quad p(q) > x_j, \quad 1 \leq j \leq r-1.$$

To complete the proof of Lemma 1 we first need two further Lemmas.

LEMMA 2. *The integers*

$$a_i^{(r)} q, \quad p(q) > x_r, \quad 1 \leq r \leq l$$

are all distinct.

Assume

$$a_{i_1}^{(r_1)} q_1 = a_{i_2}^{(r_2)} q_2, \quad p(q_1) > x_{r_1}, \quad p(q_2) > x_{r_2}, \quad r_2 > r_1. \tag{15}$$

From $p(q_2) > x_{r_2} > x_{r_1} > a_{i_1}^{(r_1)}$ we have $(q_2, a_{i_1}^{(r_1)}) = 1$, thus by (15) $q_2 | q_1$ or

$$a_{i_1}^{(r_1)} \frac{q_1}{q_2} = a_{i_2}^{(r_2)},$$

which contradicts the definition of the $a_i^{(r)}$'s. Hence (15) leads to a contradiction, which proves Lemma 2.

LEMMA 3. *Let $1 \leq r \leq l$. Then*

$$\sum_i \frac{1}{a_i^{(r)}} > \frac{\alpha}{2} \log x_r.$$

We evidently have

$$\sum_i \frac{1}{a_i^{(r)}} \geq \sum_{a_i < x_r} \frac{1}{a_i} - \sum_{a_i < x_{r-1}} \frac{1}{a_i} - \sum_{j=1}^{r-1} \sum_l^{(j)} \frac{1}{a_l}, \tag{16}$$

where in $\sum^{(j)}$ a_l runs through the a's not exceeding x_r of the form

$$a_l = a_i q, \quad a_i < x_j, \quad p(q) > x_j. \tag{17}$$

Now we estimate

$$\sum_l^{(j)} \frac{1}{a_l}.$$

Put

$$\sum_l^{(j)} \frac{1}{a_l} = \sum_1^{(j)} \frac{1}{a_l} + \sum_2^{(j)} \frac{1}{a_l}, \tag{18}$$

where in $\sum_1^{(j)}$ the summation is extended over the a_l of the form (17) where a_i has property $P(x_j, x_r, \varepsilon)$ and in $\sum_2^{(j)}$ a_i does not have property $P(x_j, x_r, \varepsilon)$. Clearly

$$\sum_1^{(j)} \frac{1}{a_l} \leq \sum_{(x_r, x_j)} \frac{1}{a_i} \sum_q \frac{1}{q}, \tag{19}$$

where in

$$\sum_q \frac{1}{q}, \quad p(q) > x_j, \quad q < x_r/a_i.$$

(19) becomes obvious if we observe that we replaced the integers $a_i q \in A$ by all the integers $a_i q$ ($p(q) > x_j$, $q < x_r/a_i$).

By the sieve of Eratosthenes we easily obtain from (14) and a classical result of Mertens

$$\sum_q \frac{1}{q} = (1+o(1)) \prod_{p<x_j} \left(1-\frac{1}{p}\right) \sum_{t=1}^{x_r} \frac{1}{t} < \frac{2 \log x_r}{\log x_j}. \tag{20}$$

From (19), (20) and (11) we obtain

$$\sum\nolimits_1^{(j)} \frac{1}{a_l} < 2\varepsilon \log x_j. \tag{21}$$

Note that (11) can be applied here because $x_r \geq x_{j+1} \geq f(x_j)$.

Now we estimate $\sum_2^{(j)} \frac{1}{a_l}$. We evidently have

$$\sum\nolimits_2^{(j)} \frac{1}{a_l} = \sum\nolimits' \frac{1}{a_i} \sum \frac{1}{q}, \tag{22}$$

where in $\sum' \frac{1}{a_i}$ a_i runs through the $a_i < x_j$ which do not have property $P(x_j, x_r, \varepsilon)$ and in $\sum \frac{1}{q}$ q satisfies

$$p(q) > x_j, \quad p < \frac{x_r}{a_i}, \quad a_i q \in A. \tag{23}$$

Since a_i does not have property $P(x_j, x_r, \varepsilon)$ we have by (23) and (9)

$$\sum \frac{1}{q} \leqslant \frac{\varepsilon \log x_r}{\log x_j}. \tag{24}$$

Further clearly

$$\sum_i{}' \frac{1}{a_i} \leqslant \sum_{t=1}^{x_j} \frac{1}{t} < 2 \log x_j. \tag{25}$$

Thus from (22), (24) and (25)

$$\sum\nolimits_2^{(j)} \frac{1}{a_l} < 2\varepsilon \log x_r. \tag{26}$$

(18), (21) and (26) imply

$$\sum_l^{(j)} \frac{1}{a_l} < 4\varepsilon \log x_r. \tag{27}$$

Thus finally, from (16), (27) and (14) we obtain for sufficiently large

$$x_r \left(r \leqslant l, \quad l = [4\alpha^{-1}]+2, \quad \varepsilon = \frac{\alpha^2}{100}\right)$$

$$\sum_i \frac{1}{a_i^{(r)}} > (\alpha-\varepsilon) \log x_r - \log\log x_r - 4r\varepsilon \log x_r > \frac{\alpha}{2} \log x_r,$$

which completes the proof of Lemma 3.

Now we are in a position to complete the proof of Lemma 1. Let y be large compared to x and consider the integers $\leqslant y$ of the form

$$a_i^{(r)}q, \quad i = 1, 2, \ldots, \quad 1 \leqslant r \leqslant l, \quad p(q) > x_r. \tag{28}$$

By Lemma 2 these integers are all distinct. It is easy to see that by Lemma 3 this leads to a contradiction.

We obtain as in (21) by the sieve of Eratosthenes (noting that $p(q) > x_r$, $q < \frac{y}{a_i^{(r)}}$ for sufficiently large y)

$$\sum_q \frac{1}{a_i^{(r)}q} = (1+o(1))\frac{1}{a_i^{(r)}}\prod_{p<x_r}\left(1-\frac{1}{p}\right)\sum_{t<(y/a_i^{(r)})}\frac{1}{t} > \frac{\log y}{2a_i^{(r)}\log x_r}. \tag{29}$$

From (29) and Lemma 3 we have

$$\sum_{i,q}\frac{1}{a_i^{(r)}q} > \frac{\alpha}{4}\log y. \tag{30}$$

Thus finally from (30) and Lemma 2

$$\sum_{t=1}^{y}\frac{1}{t} \geqslant \sum_{r=1}^{l}\sum_{i,q}\frac{1}{a_i^{(r)}q} > \tfrac{1}{4}l\alpha\log y,$$

which is false for $l = [4\alpha^{-1}]+2$. Thus the proof of Lemma 1 is completed.

Now we can prove Theorem 2. Let x; $y_1 < y_2 < \ldots$ be the numbers whose existence is guaranteed by Lemma 1. By Lemma 1 we can assume that x is sufficiently large. In other words (10) holds. Since there are infinitely many y's and only a finite number of subsets of the $a_i \leqslant x$, there is an infinite subsequence of the y's which we will again denote by $y_1 < \ldots$ for which the set $I\left(x, y_i, \frac{\alpha^2}{100}\right)$ is independent of i. Denote this set of a's by $a_1 < \ldots < a_s < x$. By (10) we have

$$\sum_{i=1}^{s}\frac{1}{a_i} > \frac{\alpha^2}{100}\log x. \tag{31}$$

A well-known theorem of Behrend [1] states that if $b_1 < \ldots < b_l \leqslant x$ is a sequence of integers no one of which divides any other then

$$\sum_{i=1}^{m}\frac{1}{b_i} < \frac{c_1 \log x}{(\log\log x)^{\frac{1}{2}}}, \tag{32}$$

where c_1 (and later $c_2, \ldots$) is an absolute constant. Thus from (31) and (32) we obtain by a simple argument that. there is a subsequence $a_{i_1} < \ldots < a_{i_t}$ of $a_1 < \ldots < a_s$ satisfying $a_{i_j} | a_{i_{j+1}}$, $1 \leqslant j \leqslant t-1$ and

$$t \geqslant \frac{\alpha^2}{100c_1}(\log\log x)^{\frac{1}{2}}. \tag{33}$$

To see this it suffices to consider a maximal subsequence $a_1^{(1)} < a_2^{(1)} < \ldots$ of $a_1 < \ldots < a_s$ where no a_j is a proper divisor of any $a_i^{(1)}$. Then omit $a_i^{(1)}$, $i = 1, 2, \ldots$ and repeat the same procedure, thus we obtain $a_i^{(2)}$, $i = 1, 2, \ldots$. Continuing we obtain the sequences $a_i^{(j)}$, $i = 1, 2, \ldots$ where $a_{i_1}^{(j)} \nmid a_{i_2}^{(j)}$ but each $a_i^{(j)}$ is a multiple of at least one $a_i^{(j-1)}$. (31) and (32) imply that j takes on at least

$$\frac{\alpha^2}{100} \log x \left(\frac{c_1 \log x}{(\log \log x)^{\frac{1}{2}}} \right)^{-1} = \frac{\alpha^2}{100 c_1} (\log \log x)^{\frac{1}{2}}$$

values, hence our assertion immediately follows.

By our construction there clearly corresponds to each y_s, $s = 1, 2, \ldots$ and a_{i_j} a set $\theta(s, j)$ of integers q_r, $r = 1, 2, \ldots$ satisfying

$$q_r < \frac{y_s}{a_{i_j}}, \quad p(q_r) > x, \quad a_{i_j} q_r \in A, \quad \sum_{q_r \in \theta(s,j)} \frac{1}{q_r} > \frac{\alpha^2}{100} \frac{\log y_s}{\log x}. \tag{34}$$

The last inequality of (34) follows from (9).

Put now $L = [400\alpha^{-2}]$. For sufficiently large x we have $l > L$ by (33). We evidently have

$$\sum_{j=1}^{L} \sum_{q_r \in \theta(s,j)} \frac{1}{q_r} \leqslant \sum_{\substack{q \leqslant y_s \\ p(q) > x}} \frac{1}{q} + \sum_{1 \leqslant j_1 < j_2 \leqslant L} \sum\nolimits^{(j_1, j_2)} \frac{1}{q_r}, \tag{35}$$

where in $\sum^{(j_1, j_2)}$ the summation is extended over the $q_r \in \theta(s, j_1) \cap \theta(s, j_2)$. As in (21) we have

$$\sum_{\substack{q < y_s \\ p(q) > x}} \frac{1}{q} < \frac{2 \log y_s}{\log x}. \tag{34}$$

From (35) and (36) and the last inequality of (34) we have

$$\sum_{1 \leqslant j_1 < j_2 \leqslant L} \sum\nolimits^{(j_1, j_2)} \frac{1}{q_r} > \frac{\log y_s}{\log x}. \tag{37}$$

From (37) there clearly are two values $1 \leqslant j_1 < j_2 \leqslant L$ for which

$$\sum\nolimits^{(j_1, j_2)} \frac{1}{q_r} > \frac{1}{\binom{L}{2}} \frac{\log y_s}{\log x} > \left(\frac{\alpha}{20} \right)^4 \frac{\log y_s}{\log x}. \tag{38}$$

The values of j, and j_2 which satisfy (38) depend on s, but since there are infintely many choices of s there are two values $1 \leqslant j_1 < j_2 \leqslant L$ which satisfy (38) for infinitely many values of s. In other words if 0(j) denotes the union of the sequences $\theta(s, j)$ then by (38) the sequence $\theta(j_1) \cap \theta(j_2)$ has positive upper logarithmic density (in fact it is greater than $\left(\frac{\alpha}{20} \right)^4 / \log x$). Denote now by $q_1 < q_2 < \ldots$ the sequence $\theta(j_1) \cap \theta(j_2)$. By (34)

$$a_{i_{j_1}} q_r \in A, \quad a_{i_{j_2}} q_r \in A,$$

which completes the proof of Theorem 2; and hence Theorem 1 is also proved.

References

1. F. Behrend, "On sequences of numbers not divisible one by another", *J. London Math. Soc.*, 10 (1935), 42-44.
2. H. Davenport and P. Erdős, "On sequences of positive integers", *Acta Arithmetica*, 2 (1937), 141-145; *see* also *J. Indian Math. Soc.* 15 (1951), 19–24.
3. P. Erdős, A. Sárközi and E. Szemerédi, "On the solvability of the equation $[a_i, a_j] = a_r$ and $(a_i', a_j') = a_r'$ in sequences of positive density", *J. Math. Anal. Appl.*, 15 (1966), 60–64.
4. D. Kleitman, "On a combinatorial problem of Erdős", *Proc. Amer. Math. Soc.*, 17 (1966), 139-141.

Mathematical Institute,
Hungarian Academy of Sciences
and
E.L.T.E. Mathematical Institute.

Printed by C. F. Hodgson & Son, Ltd., 23 Pakenham Street, London, W.C.1