Studies in Pure Mathematics
To the Memory of Paul Turán

On sums and products of integers

by

P. ERDŐS and E. SZEMERÉDI (Budapest)

Let $1 \leqq a_1 < \ldots < a_n$ be a sequence of integers Consider the integers of the form

$$(1) \qquad a_i + a_j, \quad a_i a_j, \quad 1 \leqq i \leqq j \leqq n.$$

It is tempting to conjecture that for every $\varepsilon > 0$ there is an n_0 so that for every $n > n_0$ there are more than $n^{2-\varepsilon}$ distinct integers of the form (1). We are very far from being able to prove this, but we prove the following weaker

Theorem 1. *Denote by $f(n)$ the largest integer so that for every $\{a_1, a_2, \ldots, a_n\}$ there are at least $f(n)$ distinct integers of the form* (1). *Then*

$$(2) \qquad n^{1+c_1} < f(n) < n^2 \exp(-c_2 \log n / \log\log n).$$

We expect that the upper bound in (2) may be close to the "truth".

More generally we conjecture that for every k and $n > n_0(k)$ there are more than $n^{k-\varepsilon}$ distinct integers of the form

$$a_{i_1} + \ldots + a_{i_k}, \quad \prod_{j=1}^{k} a_{i_j}$$

At the moment we do not see how to attack this plausible conjecture.

Denote now by $g(n)$ the largest integer so that for every $\{a_1, \ldots, a_n\}$ there are at least $g(n)$ distinct integers of the form

$$(3) \qquad \sum_{i=1}^{n} \varepsilon_i a_i, \quad \prod_{i=1}^{n} a_i^{\varepsilon_i} \quad (\varepsilon_i = 0 \text{ or } 1)$$

We conjecture that for $n > n_0(k)$, $g(n) > n^k$. Unfortunately we have not been able to prove this and perhaps we overlook a simple idea. We prove

Theorem 2.

$$g(n) < \exp(c_3 \log^2 n / \log\log n).$$

Again we believe (without too much evidence) that Theorem 2 may be close to the final truth. Perhaps our conjectures remain true if the a's are real or complex numbers.

Some more conjectures: Let $\mathcal{G}(n, k)$ be a graph of n vertices $x_1, x_2, \ldots, x_n$ and k edges. Make correspond a_i to x_i. Consider the set of $2k$ integers.

$$\{a_i + a_j, \quad a_i a_j\} \tag{4}$$

where x_i is joined to x_j. We conjecture that for every $\varepsilon > 0$ and $0 < \alpha \leqq 1$ if $k > n^{1+\alpha}$ then there are more than $n^{1+\alpha-\varepsilon}$ distinct integers of the form (4). Our proof of Theorem 1 does not seem to apply here. The conjecture very likely remains true if the a's can be real numbers. P. ERDŐS once thought that the conjecture may hold even if we only assume $k > cn$, but A. RUBIN showed that this is not true if the a's can be real numbers and it perhaps fails even if the a's are restricted to be positive integers.

Finally we state a few related problems. Let $a_i b_i = T$ $i = 1, 2, \ldots, n$. Consider the sums

$$a_{i_1} + a_{i_2}, \quad b_{i_1} + b_{i_2}, \quad a_{i_1} + b_{i_2} \quad 1 \leqq i_1 \leqq i_2 \leqq n .$$

Is it true that all but one of three sets have more than n^{1+c} distinct elements?

Consider the sets $\{k(n-k),\ 1 \leqq k < n\}$ and $\{l(m-l),\ 1 \leqq l < m\}$. Can one estimate the number of integers which are common to both sets?

Let $a_1, \ldots, a_n$ be such that there are only cn distinct sums of the form $a_i + a_j$, $1 \leqq i \leqq j \leqq n$. Then there certainly must be more than $n^{2-\varepsilon}$ distinct products of the form $a_i a_j$, $1 \leqq i \leqq j \leqq n$. Perhaps there are more than $n^2/(\log n)^c$ products of the form $a_i a_j$, $1 \leqq i \leqq j \leqq n$. The deep results of FREIMAN can possibly be used here [1].

Finally a problem of different kind. Let $2n - 1 \leqq t \leqq \dfrac{n^2+n}{2}$. It is easy to see that one can find a sequence of integers $a_1 < \ldots < a_n$ so that there should be exactly t distinct integers in the sequence $a_i + a_j$, $1 \leqq i \leqq j \leqq n$. We do not know for which t is it possible to find a sequence $a_1 < \ldots < a_n$ so that there should be exactly t distinct integers of the form

$$\sum_{i=1}^{n} \varepsilon_i a_i, \quad \varepsilon_i = 0 \text{ or } 1 .$$

It is probably even more difficult to find out for which $t > f(n)$ is there a sequence $a_1 \ldots < a_n$ so that there are exactly t distinct integers of the form (1).

First we prove Theorem 2 which will not be difficult. Let x be large. The a's are the integers of the form

$$\Pi p_i^{\alpha_i}, \quad p_i < (\log x)^{2/3}, \quad 0 \leqq \alpha_i \leqq (\log x)^{1/3} .$$

Put

$$[(\log x)^{1/3}] = t, \quad \pi([(\log x)^{2/3}]) = (1 + o(1)) \frac{3(\log x)^{2/3}}{2 \log\log x} = l . \tag{5}$$

The number of a's is

$$n=(t+1)^l=\exp\left(\frac{1}{2}(\log x)^{2/3}\right). \tag{6}$$

All the a's are less than x, thus the number of the distinct sums is less than x^2.

Next we have to estimate the number of the distinct product of the form $\prod_{i=1}^{n} a_i^{\varepsilon_i}$, $\varepsilon_i=0$ or 1. These integers are all composed of the first l primes. The highest exponent of a prime p which can occur in $\prod_{i=1}^{n} a_i^{\varepsilon_i}$ is at most $tn<(t+1)^{l-1}=(t+1)n$. Thus the number of the integers of the form $\prod_{i=1}^{n} a_i^{\varepsilon_i}$ $\varepsilon_i=0$ or 1, is less than

$$((t+1)n)^l=(t+1)^{l^2+l}. \tag{7}$$

To complete the proof of Theorem 2 we only have to show by (5) and (6) that

$$n^{c\log n/\log\log n}>(t+1)^{l^2+l}+x^2. \tag{8}$$

(8) immediately follows from (5) and (6), which completes the proof of Theorem 2.

Now we prove Theorem 1. First we prove the right side of (2). This will be a standard and comparatively simple estimation. We do not try to obtain the largest possible value of c_2 since we are not at all sure that the term $n^2\exp\left(-\frac{c_2\log n}{\log\log n}\right)$ is the final truth.

To prove the right side of (2) let $2j$ be the largest even integer not exceeding $\frac{\log x}{3\log\log x}$, $s=\pi((\log x)^3)$. The a_i are the integers of the form

$$\prod_{i=1}^{2j} p_i^{\varepsilon_i}, \quad p_i<(\log x)^3, \qquad \varepsilon_i=0 \text{ or } 1. \tag{9}$$

These integers are clearly all less than x. Their number clearly equals

$$t_x=\binom{s}{2j}=x^{2/3+o(1)}. \tag{10}$$

The number of distinct integers of the form a_i+a_j is by (10) and $a_i<x$ less than $2x<t_x^{3/2+o(1)}$ and thus can be neglected. Next we have to estimate the number of distinct integers of the form a_ia_k. We split these integers into two classes. In the first class are the a_ia_k for which ($v(n)$ denotes the number of distinct prime factors of n)

$$v((a_i,a_k))>j.$$

The number of these integers is by a simple computation less than

$$t_x \log x \binom{2j}{j}\binom{s}{j} < t_x \log x\, 2^{2j}(\log x)^{(2+o(1))j} < t_x x^{1/3+o(1)}.$$

Thus the numbers of the first class can be also neglected.

Now if $a_i a_k$ is in the second class we can write

$$a_i a_k = Q^2 L$$

where $Q=(a_i, a_k)$ is squarefree and L is the product of two relatively prime squarefree integers having $2j-\nu(Q)$ prime factors, where $\nu(Q)<j=\dfrac{\log x}{6\log\log x}$. But then clearly Q^2L can be written in at least $\binom{2j}{j}$ ways as the product of two numbers a_i, a_k, $\nu(a_i, a_k)=Q$. Thus the number of integers in the second class is less than

$$t_x^2\, 2^{-\frac{\log x}{3\log\log x}}.$$

which proves the right side of (2).

To complete the proof of Theorem 1 we now have to prove the left side of (2), and this in fact is the main novelty and difficulty of our paper. We make no attempt to get a large value for c_1 as stated in the introduction $c_1>1-\varepsilon$ for every $\varepsilon>0$ and our method cannot even give $c_1=\dfrac{1}{2}$.

First a few remarks. If $a_n<n^k$ our Theorem follows trivially with $c_1>1-\varepsilon$, thus the only difficulty is if some of the a's are very large. First we prove that we can assume without loss of generality that all the a_i are in some interval $u\leqq a_i\leqq 2u$.

Denote by S_i the set of a_j's satisfying $2^i<a_j\leqq 2^{i+1}$. First observe that we can assume without loss of generality that

$$|S_i|=0 \quad\text{or}\quad |S_i|\geqq n^{1/4}. \tag{11}$$

Assume that (11) does not hold. Let $S_{i_1}, \ldots, S_{i_k}$ satisfy

$$0<|S_{i_j}|<n^{1/4}, \quad 1\leqq j\leqq k. \tag{12}$$

If $\bigcup_{j=1}^{k}|S_{i_j}|<\dfrac{n}{2}$ we simply omit all the a's satisfying (12) and we only work with the remaining a's and since their number is greater than $\dfrac{n}{2}$ this clearly can be done. If $\bigcup_{j=1}^{k}|S_{i_j}|\geqq\dfrac{n}{2}$ then by (12) clearly $k\geqq n^{3/4}/2$. Let a_{i_j} be an arbitrary element of

S_{i_j}, $j=1, 2, \dots k$, $k \geqq n^{3/4}/2$. Clearly $a_{i_{2j+2}} > 2a_{i_{2j}}$ and thus the sums

$$a_{i_{2j_1}} + a_{i_{2j_2}}, \quad 2 \leqq j_1 < j_2 \leqq k$$

are all distinct, so there are at least $\dfrac{n^{\frac{3}{2}}}{\delta}$ distinct sums of the form $a_u + a_v$, $1 \leqq u < v \leqq n$, which proves Theorem 1 if (11) does not hold.

Thus we can now assume that (11) holds.

Now we state the crucial

Lemma. *Let $m < b_1 < \dots < b_t \leqq 2m$. Then the number of distinct integers of the form*

$$b_i + b_j, \quad b_i b_j, \quad 1 \leqq i < j \leqq t$$

is greater than $\varepsilon t^{1+\alpha}$ for some $\alpha > 0$ and $\varepsilon > 0$.

Suppose that our Lemma has already been proved. Then by (11) and our Lemma the number of distinct integers of the form $a_i + a_j$, $a_i a_j$ is at least

$$\sum{}' \varepsilon |S_i|^{1+\alpha} > cn^{1+\alpha/4} \tag{13}$$

(where the dash indicates that the summation is extended over the i satisfying $|S_i| \geqq n^{1/4}$) (13) of course gives the left side of (2) and hence proves Theorem 1.

Thus we only have to prove our Lemma. Put $[t^{1/8}] = s$. Denote by B_i the set of b's $\{b_{(i-1)s+1}, \dots, b_{is}\}$. In other words we divided the index set of the b's into $[t^{7/8}]$ sets of size $[t^{1/8}]$. Denote by $B = B_r$ the B_j of smallest diameter (i.e. $b_{(j-1)s+1} - b_{js}$ is minimal). Observe now that if $u - v \geqq 10$ and $(u \neq r,\ v \neq r)$ $b_1 \in B$, $b_2 \in B$, $b_3 \in B_u$, $b_4 \in B_v$ then $b_1 + b_3 \neq b_2 + b_4$ and $b_1 b_3 \neq b_2 b_4$. This is obvious for the sum and nearly obvious for the product. Put $b_2 = b_1 + x$, $b_4 = b_3 - y$. Then if $b_1 b_3 = b_2 b_4$ we would have $b_1 b_3 = (b_1 + x)(b_3 - y)$ or $xy = b_3 x - b_1 y$ and this easily leads to a contradiction since $y > 10x$ by the minimality property of $B = B_r$ and $u - v \geqq 10$. Further $1/2 < b_3/b_1 < 2$. Thus $b_3 x - b_1 y < 0 < xy$ which is impossible.

Consider now the $s^7/10$ B_j's, $j \equiv 1 \pmod{10}$. We divide the indices j into two classes. In the first class are the indices j for which the number of distinct integers of the form

$$b_i + b_l, \quad b_i b_l, \quad b_i \in B, \quad b_l \in B_j$$

is greater than $s^{1+8\alpha}$. If at least half of the indices belong to the first class then our Lemma immediately follows since the number of distinct integers of the form $b_i + b_j$, $b_i b_j$ is greater than $\dfrac{1}{10} \cdot \dfrac{1}{2} \cdot s^7 s^{1+8\alpha} = \dfrac{1}{20} t^{1+\alpha}$ which proves the Lemma in this case.

Let now j be an index of the second class. We remind the reader that in this case the number of distinct integers of the form $b_u + b_v$, $b_u b_v$, $b_u \in B$, $b_v \in B_j$ is less than $s^{1+8\alpha}$.

We want to find six integers $b_1, b_2, b_3, b_4, b_5, b_6$, $b_i \in B_j$ $(i=1, 2)$, $b_i \in B$ $(3 \leqq i \leqq 6)$, satisfying

$$(14) \qquad b_1 + b_3 = b_2 + b_4 \quad \text{and} \quad b_1 b_5 = b_2 b_6 .$$

Consider the s^2 products $b_u b_v$, $b_u \in B$, $b_v \in B_j$. Since B_j is in the second class there are fewer than $s^{1+8\alpha}$ distinct integers of this form. Therefore there is a T so that $T = b_u b_v$ has at least $s^{1-8\alpha}$ solutions. Put

$$T = b_{u_r} b_{v_r}, \quad b_{u_r} \in B \qquad b_{v_r} \in B_j, \quad 1 \leqq r \leqq s^{1-8\alpha} \quad .$$

Consider now the $s^{2-16\alpha}$ sums of the form $b_{u_w} + b_{v_j}$. For sufficiently small α these sums clearly cannot all be different.

Thus there are indices u_w, v_l, u_p, v_q so that $b_{u_w} + b_{v_l} = b_{u_p} + b_{v_q}$. But $b_{u_l} b_{v_l} = b_{u_q} b_{v_q}$. Thus b_{u_w}, b_{u_p}, b_{u_l}, $b_{u_q} \in B$, b_{v_l}, $b_{v_q} \in B_j$ are our required six integers. Observe that if b_3, b_4, b_5, b_6 are fixed there is at most one b_1, b_2 pair which solves (14).

We have at least $\frac{1}{2} \cdot \frac{1}{10} \cdot s^7$ B_j's in the second class and the number of different b_3, b_4, b_5, b_6 quadrouples is at most s^4. This contradicts our observation, and this contradiction completes the proof of Theorem 1.

Reference

[1] FREIMAN, G. A., *Foundations of a structural theory of set addition.* Translations of Math. Monographs, Amer. Math. Soc., Vol. 37. Providence R.I. 1973.

MATHEMATICAL INSTITUTE OF THE
HUNGARIAN ACADEMY OF SCIENCES
H-1053 BUDAPEST, REÁLTANODA U. 13–15.
HUNGARY