SOME RESULTS ON ADDITIVE NUMBER THEORY[1]

PAUL ERDÖS

Let $0<a_1<a_2< \cdots$ be any infinite sequence of integers. Denote by $N(a_i, n)$ the number of $a_i \leqq n$. I conjectured that to every sequence a_i there corresponds a sequence b_j of density 0 (i.e., such that $\lim_{n=\infty} (1/n)N(b_j, n)=0$) so that every sufficiently large integer is of the form a_i+b_j. Lorentz[2] in a recent paper proved this conjecture; in fact, he showed that there exists a sequence b_j with the required property satisfying for every n

$$N(b_j, n) < c_1 \sum_{k=1}^{n} \frac{\log N(a_i, k)}{N(a_i, k)} \tag{1}$$

where c_1 and later $c_2, c_3, \cdots$ denote suitable absolute constants.

Lorentz communicated this result to me and also asked: If the sequence a_i consists of the primes, what is the thinnest sequence (i.e. $N(b_j, n)$ should be minimal) $b_1<b_2< \cdots$ for which $p+b_j$ will represent every sufficiently large integer.[3]

From the fact that $\pi(x)<c_2x/\log x$ it immediately follows that $N(b_j, n)$ must be greater than $c_3 \log n$. Lorentz's formula (1) gives that there exists such a sequence $b_1<b_2< \cdots$ so that $N(b_j, n) <c_4(\log n).^3$

After I received Lorentz's letter, I observed that a method which I applied in a recent paper[3] applies here. Since my method is more complicated than that of Lorentz and in general does not give better results than (1), it is not worthwhile to work it out for general sequences, but in the case of the primes it gives better results than (1). In fact, we shall prove the following

THEOREM 1. *There exists a sequence* $\{b_j\}$ *satisfying* $N(b_j, n) <c_5(\log n)^2$ *for all* n *so that all sufficiently large integers are of the form* $p+b_j$.

If the sequence a_i has positive lower density (i.e. there exists an $\alpha>0$ so that for all large n, $N(a_i, n)>\alpha n$) then (1) gives that there

Received by the editors March 16, 1954.

[1] This paper is written with partial support from O.N.R.

[2] This paper of Lorentz appears in Proc. Amer. Math. Soc. vol. 5 (1954) pp. 838–841.

[3] The paper in question will appear in Acta Univ. Szeged. I prove there the existence of a sequence $\{a_i\}$ so that every large integer is of the form a_i+a_j, but the number of solutions of $n=a_i+a_j$ is less than $c \log n$.

exists a sequence b_j satisfying for all n the inequality $(Nb_j, n) < c_6(\log n)^2$ so that all sufficiently large integers are of the form a_i+b_j.

We shall prove that this result is best possible by proving

THEOREM 2. *There exists a sequence $\{a_i\}$, so that for all large n, $N(a_i, n) > \alpha n$, and if $\{b_j\}$ is such that all sufficiently large integers are of the form a_i+b_j, then for all n, $N(b_j, n) > c_7(\log n)^2$.*

Buck and Volkman[4] call a sequence $A=\{a_i\}$ pseudorational if to every ϵ there exist two sequences S_1 and S_2 both of which are finite unions of arithmetic progressions (such sequences are called by Volkman rational sequences) such that for large n

$$S_1 \subset A \subset S_2, \qquad N(S_2 - S_1, n) < \epsilon n$$

(i.e. the number of of integers contained in S_2 and not in S_1 is $<\epsilon n$).

It was conjectured[5] that there exist two pseudorational sequences $\{a_i\}$ and $\{b_j\}$ whose sum (that is, the set of all integers of the form $\{a_i+b_j\}$) is not pseudorational. We are going to prove this conjecture and also make some remarks about a few similar questions.

PROOF OF THEOREM 1. We first prove the following:

LEMMA. *One can find x integers $d_1<d_2<\cdots<d_x$ satisfying $x=c_8[\log n]^2$, $n^{5/8}/2<d_1<d_2<\cdots<d_x<n^{5/8}$, and so that every integer $n^{5/8}<u\leqq n$ is of the form $p+d_i$.*

Put $T=n^{5/8}/2$. The number of ways one can pick the d's clearly equals $C_{T,x}$. We shall show that for almost all (i.e. except for $o(C_{T,x})$) choices of the d's every integer $n^{5/8}<u\leqq n$ is of the form $p+d_i$.

First we estimate in how many ways one can pick

$$n^{5/8}/2 < d_1 < d_2 < \cdots < d_x < n^{5/8}, \qquad x = c_8 (\log n)^2$$

so that $u\neq p+d_i$, where $n^{5/8}<u\leqq n$ is any given number.

Consider the interval $(u-n^{5/8}, u-(1/2)n^{5/8})$. By a theorem of Hoheisel-Ingham[6] the number of primes $q_1, q_2, \cdots, q_y$ in this interval satisfies $y>c_9n^{5/8}/\log n$. If $u\neq p+d_i$, then clearly none of the integers

$$u - q_i, \qquad 1 \leqq i \leqq y,$$

can be a d. Thus the number of possible choices of the d's so that $u\neq d_i+p$ is not greater than

[4] R. C. Buck, Amer. J. Math. vol. 68 (1946) pp. 560–580; see also E. F. Buck and R. C. Buck, ibid. vol. 69, p. 413–420.

[5] Volkman, ibid.

[6] Ingham, Quarterly Journal of Mathematics vol. 8 (1937) pp. 255–266.

$$C_{T-y,x} < C_{T,x}\left(1-\frac{y}{T}\right)^{x} < C_{T,x}\left(1-\frac{2c_9}{\log n}\right)^{c_8(\log n)^2} < \frac{1}{n^2}C_{T,x}$$

for sufficiently large c_8. Thus the number of possible choices of the d's so that at least one of the u's, $n^{5/8}<u\leqq n$, should not be of the form $p+d_i$ is less than $(1/n)C_{T,x}=o(C_{T,x})$.

Thus for at least one choice of $n^{5/8}/2<d_1<\cdots<d_x\leqq n$ all integers $n^{5/8}<u\leqq n$ are of the form $p+d_i$.

Now to complete our proof let n_1 be sufficiently large. Put $n_k=n_{k-1}^{8/5}$ and let

$$n_{k-1}^{5/8}/2 < d_1^{(k-1)} < \cdots < d_{x_{k-1}}^{(k-1)} < n_{k-1}^{5/8}, \qquad x_{k-1} = c_8\,(\log n_k)^2$$

be such that every integer $n_{k-1}^{5/8}<u\leqq n_k$ is of the form $p+d_j^{(k-1)}$. Such a sequence $d_j^{(k-1)}$ exists by our lemma. Now if we form a sequence $b_1<b_2<\cdots$ of the $d_j^{(k)}$, $1\leqq j\leqq x_k$, $k=1, 2, \cdots$, then a simple computation gives $N(b_j, x)<c_5(\log x)^2$, and every integer greater than $n_1^{5/8}$ is of the form $p+b_i$; thus the proof of our theorem is complete.

It would be interesting to know if our result is best possible. Lorentz also asked the question: Does there exist a sequence b_j satisfying $N(b_j, x)<c_{10}x^{1/2}$ so that every large integer is of the form k^2+b_j? Using (1) or the method which I just developed one obtains $N(b_j, x)<c_{11}x^{1/2}\log x$.

Let $a_1, a_2, \cdots, a_x$ be x distinct residues mod n. Both Lorentz and I proved that there always exist residues $b_1, b_2, \cdots, b_y$ so that

$$y < c_{12}\frac{n}{x}\log x \tag{2}$$

and every residue (mod n) can be written in the form a_i+b_j.

Proof of Theorem 2. Let $0<t<1$ be any real number. Put

$$t = \sum_{l=1}^{\infty}\frac{\epsilon_l(t)}{2^l}, \qquad \epsilon_l(t) \text{ is } 0 \text{ or } 1.$$

We define the sequence A_t as follows: A_t consists of all integers l satisfying $\epsilon_l(t)=+1$, $8^k<l\leqq 2\cdot 8^k$, $k=1, 2, \cdots$. Roughly we can define the sequences A_t as follows: In the intervals $(8^k, 2\cdot 8^k)$, $k=1, 2, \cdots$, we choose each integer with probability $1/2$ to belong to A_t.

We shall show that for almost all t, A_t satisfies the conditions of Theorem 2. First of all it is easily seen by standard probability theorems that for almost all t,

$$\liminf_{n\to\infty} N(A_t, n)/n = 1/14 \tag{3}$$

(since asymptotically half the integers of the interval $(8^k, 2\cdot 8^k)$ belong to A_t). Thus for almost all t, A_t has positive lower density.

Next we show that for almost all t, A_t has the following property: Let $b_1, b_2, \cdots$ be a sequence B such that the sum of A_t and B contains all sufficiently large integers; then for all $n > n_0$

$$N(B, n) > c_7 (\log n)^2. \tag{4}$$

(3) and (4) clearly imply Theorem 2.

To prove (4) we shall show that for almost all t, A_t has the property that if the sum of A_t and B contains all sufficiently large integers, then

$$N(B, 8^{k+1}) - N(B, 2\cdot 8^k) \geqq k \tag{5}$$

holds for all but a finite number of values of k.

(5) clearly implies (4), thus it will suffice to prove (5). Because of the Borel-Cantelli lemma it will suffice to show that for $k > k_0$ the measure of the values of t for which (5) does not hold is less than $1/2^k$.

Every integer of the interval $(4\cdot 8^k, 8^{k+1})$ must be of the form $a_i + b_j$. Since all the a_i less than 8^{k+1} are less than $2\cdot 8^k$, only the b's in the interval $(2\cdot 8^k, 8^{k+1})$ give (a_i+b_j)'s in $(4\cdot 8^k, 8^{k+1})$. Assume that the number of these b's is less than k. The total number of possible choices of the b's is thus less than $C_{6\cdot 8^k, k}$.

Let $2\cdot 8^k \leqq b_1 < b_2 < \cdots < b_r < 8^{k+1}$ be an arbitrary set of r b's. We shall estimate the measure of the set $\{t\}$ so that the sum of A_t and the b's should contain all integers of the interval $(4\cdot 8^k, 8^{k+1})$. Let $u_1, u_2, \cdots, u_x$ be a maximal set of integers in the interval $(4\cdot 8^k, 8^{k+1})$ with the property that the integers

$$u_s - b_j, \qquad 1 \leqq s \leqq x;\ 1 \leqq j \leqq r,$$

are all distinct. The condition of maximality means that every integer of the interval $(4\cdot 8^k, 8^{k+1})$ is of the form

$$u_s + b_{j_1} - b_{j_2}, \quad 1 \leqq s \leqq x;\ 1 \leqq j_1, j_2 \leqq r. \tag{6}$$

For if not then there is a number say u_{x+1} in the interval $(4\cdot 8^k, 8^{k+1})$ which is not of the form (6). But then all the integers $u_s - b_j$, $1 \leqq s \leqq x+1$; $1 \leqq j \leqq r$, are distinct, which contradicts the maximality of $u_1, u_2, \cdots, u_x$.

The number of integers of the form (6) equals $xr^2 < xk^2$, and since every integer in $(4\cdot 8^k, 8^{k+1})$ is of the form (6) we have

$$x > 4\cdot 8^k/k^2. \tag{7}$$

Now since each of the u's must be of the form $a_i + b_j$, at least one of the integers $u_s - b_j$, $1 \leqq j \leqq r$, must be an a and this must hold for each $1 \leqq s \leqq x$. The measure of the set in t so that at least one of the integers $u_s - b_j$ is an a is less than or equal to $1 - 1/2^r < 1 - 1/2^k$ and since for different values of s the sets $u_s - b_j$, $1 \leqq j \leqq r$, are disjoint, the x events: "$u_s - b_j$, $1 \leqq j \leqq r$, contains at least one a", $s = 1, 2, \cdots, x$, are independent. Thus the measure in t of all these events happening simultaneously is less than (for $k > k_0$)

$$\left(1 - \frac{1}{2^k}\right)^x < \left(1 - \frac{1}{2^k}\right)^{4 \cdot 8k/k2} < \left(1 - \frac{1}{2^k}\right)^{4k} < e^{-2k}. \tag{8}$$

Since the number of choices for the b's is less than $C_{6 \cdot 8^k, k} < 8^{k(k+1)}$, we obtain from (8) that the measure of the set in t so that (5) should hold is less than

$$e^{-2k} 8^{k(k+1)} < 1/2^k$$

for $k > k_0$, which proves Theorem 2.

By the same method we can prove that under the assumption $N(a_i, n) > n^{1-\epsilon}$, ϵ small enough, (1) gives the best possible result under fairly general conditions. Similarly, if $x > n^{1-\epsilon}$, (2) gives the best possible results. We do not formulate these results precisely since they are somewhat complicated and no doubt very far from being best possible. (2) in fact may be best possible for every x.

One might be tempted to define the sequence A_t as follows: Let $t = \sum_{l=1}^{\infty} \epsilon_l(t)/2^l$, l belongs to A_t if and only if $\epsilon_l(t) = +1$. Unfortunately (4) fails to be true here. In fact, it is not difficult to show that if the sequence B consists of the integers 2^k, $2^k + 1$, $k = 1, 2, \cdots$, then for almost all t all large integers are of the form $a + b$, $a \in A_t$, $b \in B$.

It is immediate that the Schnirelmann sum of two rational sequences is again rational. Next we show that the Schnirelmann sum of two pseudorational sequences does not have to be pseudorational.

It is obvious that the squares form a pseudorational set of density 0. (This follows immediately from the fact that there are $(p-1)/2$ non-residues mod p.) If the integers of the form $x^2 + y^2$ are not pseudorational we already have an example of a pseudorational set S so that $S + S$ is not pseudorational. Thus we can assume that the set of integers S_1 of the form $x^2 + y^2$ is pseudorational. As a matter of fact, Ruchte proved this. His proof follows easily from the well-known characterization of the integers which are of the form $x^2 + y^2$ and can be left to the reader. Also it is well known that the density of S_1 is 0 and that every integer is of the form $S_1 + S_1$ (i.e. every integer is the sum of 4 squares). Now to complete our proof we show that if A is any pseudorational set of density 0 so that all integers are of the form $A+A$, then there exists a subset B of A so that $B+B$ has upper density 1 and lower density 0, thus $B+B$ can not be pseudorational. (B is pseudorational since it is a subset of a pseudorational set of density 0.)

We construct B as follows: Let $n_1<n_2< \cdots <n_k< \cdots$ tend to infinity so fast that

$$(9) \qquad n_{k+1}/n_k \to \infty \quad \text{and} \quad n_k \cdot N(A, n_{k+1})/n_{k+1} \to 0.$$

Since the density of A is 0, (9) can be satisfied. The sequence B now consists of the integers of the sequence A in the intervals (n_{2k-1}, n_{2k}). Clearly $N(B+B, n_{2k-1}) \leqq 2n_k$ (since if b is in B and $b \leqq n_{2k+1}$, then $b \leqq n_{2k}$). Thus from (9)

$$(10) \qquad \lim_{k=\infty} N(B+B, n_{2k+1})/n_{2k+1} = 0.$$

Further since all integers are of the form $A+A$

$$(11) \qquad N(B+B, n_{2k}) \geqq n_{2k} - N(B, n_{2k-1}) \cdot N(B, n_{2k})$$

(i.e. we have to omit the integers of the form a_i+a_j, $a_i \in A$, $a_j \in A$, where one of the summands is $\leqq n_{2k-1}$, thus does not have to belong to B). Thus from (11) and (9)

$$(12) \qquad N(B+B, n_{2k}) > n_{2k} - n_{2k-1} \cdot N(B, n_{2k}) = n_{2k} - o(n_{2k}).$$

(10) and (12) show that the lower density of $B+B$ is 0 and its upper density is 1, thus the construction of our counter example is completed.

Define B as the sequence of integers in $(2^{2^{2k}}, 2^{2^{2k+1}})$, $k=1, 2, \cdots$, which are of the form x^2+y^2. A simple computation will show that $B+B$ is not pseudorational, in fact has upper density 1 and lower density 0. We leave the simple proof to the reader.

Now we state a characterization of those sequences S of density 0 for which there exists a pseudorational set B so that $S+B$ is not pseudorational. First of all it easily follows from (1) that if S is any sequence of density 0 there exists a sequence B of density 0 so that $S+B$ has lower density 0 and upper density 1. In general of course B will not be pseudorational.

Denote by

$$(13) \qquad u_1^{(k)}, u_2^{(k)}, \cdots, u_{s_k}^{(k)}, \qquad k = 2, 3, \cdots,$$

those residues (mod $k!$) which contain at least one element of S. It is easy to see that S is pseudorational if and only if

$$\lim_{k=\infty} S_k/k! = 0.$$

The system of residues (13) is said to contain a branching subsystem if there exists an infinite sequence $k_1 < k_2 < \cdots < k_r < \cdots$ of integers and 2^r residues $u_i^{(k_r)} \pmod{k_r!}$, $1 \leqq i \leqq 2^r$, $r = 1, 2 \cdots$, so that each $u_i^{(k_r)}$ is congruent to two $u_i^{(k_{r+1})} \pmod{k_r!}$.

If the sequence S is such that the system (13) belonging to it contains a branching system, then there always exists a pseudorational sequence B of density 0 so that the Schnirelmann sum $S+B$ has upper density 1 and lower density 0.

The proof uses (2), is somewhat lengthy but straightforward, and will be omitted. Incidentally it is easy to see that if S is not pseudorational, the system (13) always contains a branching subsystem.

If S is such that the system (13) does not contain a branching subsystem, then S is easily seen to be pseudorational and if B is any pseudorational sequence, then $S+B$ is also pseudorational. We omit the proof, which is not difficult.

To give examples: for the pseudorational sequence 2^k, $k = 1, 2, \cdots$, the system (13) contains a branching subsystem; for $k!$ it does not.

One final remark. The rational sequences S have the property that if B is any sequence, then $S+B$ has a density (in fact, is again rational). As remarked before, to every sequence S of density 0 there exists a sequence B so that $S+B$ has no density. One can ask: Are there any nonrational sequences S so that, for any B, $S+B$ has a density.[3] A simple probability argument shows that almost all sequences have this property. To every sequence $a_1, a_2, \cdots$ corresponds the real number $\sum_{k=1}^{\infty} 1/2^{a_k} = t$ and our statement means that for almost all t the corresponding sequence has the required property. In fact, if B has infinitely many elements, the density of $S+B$ is 1; if B has k elements, the density of $S+B$ is $1 - 1/2^{k+1}$.

After finishing this paper I noticed that it is indeed easy to show that there exists a sequence $b_1 < b_2 < \cdots$ satisfying $N(b_j, x) < c_{10}x^{1/2}$ so that every large integer is of the form $l^2 + b_j$. It suffices to take as the b's the integers of the intervals

$$2^k < b < 2^k + 4 \cdot 2^{k/2}, \qquad k = 1, 2, \cdots .$$

An analogous example shows that there is a sequence $b_1 < b_2 < \cdots$ satisfying $N(b_j, x) < c_k x^{1-1/k}$ so that every sufficiently large integer is of the form $l^k + b_j$. The following question seems more difficult: Does there exist a sequence $b_1 < b_2 < \cdots$ satisfying $N(b_j, x) < c'_{10}x/\log x$ so that every sufficiently large integer is of the form $2^l + b_j$?

UNIVERSITY OF NOTRE DAME