ACTA ARITHMETICA
XCII.2 (2000)

On the representation of integers as sums of distinct terms from a fixed set

by

NORBERT HEGYVÁRI (Budapest)

Introduction. Let A be a strictly increasing sequence of positive integers. The set of all the subset sums of A will be denoted by $P(A)$, i.e. $P(A) = \{\sum \epsilon_i a_i : a_i \in A;\ \epsilon_i = 0 \text{ or } 1\}$. A is said to be *subcomplete* if $P(A)$ contains an infinite arithmetic progression. A natural question of P. Erdős asked how dense a sequence A which is subcomplete has to be. He conjectured that $a_{n+1}/a_n \to 1$ implies the subcompleteness. But in 1960 J. W. S. Cassels (cf. [1]) showed that for every $\varepsilon > 0$ there exists a sequence A for which $a_{n+1} - a_n = o(a_n^{1/2+\varepsilon})$ and A is not subcomplete. In 1962 Erdős [2] proved that if $A(n) > Cn^{(\sqrt{5}-1)/2}$ $(C > 0)$ then A is subcomplete, where $A(n)$ is the counting function of A, i.e. $A(n) = \sum_{a_i \le n} 1$. In 1966 J. Folkman [4] improved this result showing that $A(n) > n^{1/2+\varepsilon}$ $(\varepsilon > 0)$ implies the subcompleteness.

In this note we improve this result. In Section 3 we prove

THEOREM 1. *Let $A = \{0 < a_1 < a_2 < \ldots\}$ be an infinite sequence of integers. Assume that $A(n) > 300\sqrt{n \log n}$ for $n > n_0$. Then A is subcomplete.*

We mention here that $300\sqrt{n \log n}$ cannot be replaced by $\sqrt{2n}$; it is easy to construct a sequence A for which $A(n) > \sqrt{2n}$ and A is not subcomplete.

The main tool for the proof of Theorem 1 is a remarkable theorem of G. Freiman and A. Sárközy (they proved it independently, see [5] and [7]). We are going to use it as Lemma 3.

We use the following notations. The cardinality of the finite set S is denoted by $|S|$. The set of positive integers is denoted by $\mathbb{N}$. $A + B$ denotes

2000 *Mathematics Subject Classification*: 11B75, 11A67.
Key words and phrases: subcomplete sequence, additive representations.
Research partially supported by Hungarian National Foundation for Scientific Research, Grant No. T025617, Grant No. T029759 and by DIMACS (Center for Discrete Mathematics and Theoretical Computer Science) NSF-STC-91-19999.

the set of integers that can be represented in the form $a+b$ with $a \in A$, $b \in B$. We write $X_1 + \ldots + X_n = (X_1 + \ldots + X_{n-1}) + X_n$, $n = 3, 4, \ldots$

Acknowledgements. I would like to express my thanks to Prof. G. Freiman for his helpful comments and suggestions.

1. Preliminaries. First we prove

PROPOSITION. *Let $A = \{0 < a_1 < a_2 < \ldots\}$ be an infinite sequence of integers. Assume that $A(n) > 2\sqrt{n \log n}$ for $n > n_0$. Then for every d there exists an $L > 0$ and an infinite sequence $\{y_1 < y_2 < \ldots\}$ in $P(A)$ for which $d \mid y_i$ and $y_{i+1} - y_i < L, i = 1, 2, \ldots$*

Proof. $A(n) > 2\sqrt{n \log n}$ implies

$$a_n < \frac{n^2}{\log n}. \tag{1.1}$$

Let $U_i = \{a_{(i-1)d+1} < \ldots < a_{id}\}$. We need some lemmas.

LEMMA 1. *If $d \in \mathbb{N}$ and $u_1, \ldots, u_d$ are integers, then there is a sum of the form*

$$u_{i_1} + \ldots + u_{i_t} \qquad (1 \le i_1 < \ldots < i_t \le d)$$

such that $d \mid u_{i_1} + \ldots + u_{i_t}$.

Proof. Either there is a k, $1 \le k \le d$, such that $d \mid u_1 + \ldots + u_k$ or there are k, m with $k < m$ and $u_1 + \ldots + u_k \equiv u_1 + \ldots + u_m \pmod d$ so that $d \mid u_{k+1} + \ldots + u_m$.

By Lemma 1, for every i there exists y_i such that $d \mid y_i = a_{i_1} + \ldots + a_{i_t}$, $a_{i_1} < \ldots < a_{i_t}$ and $\{a_{i_1}, \ldots, a_{i_t}\} \subseteq U_i$. Furthermore by (1.1) we get

$$y_i < d a_{id} < d \frac{(id)^2}{\log i} = d^3 \frac{i^2}{\log i}$$

or equivalently

$$Y(n) > \frac{\sqrt{n \log n}}{d^3}, \quad \text{where} \quad Y = \{y_1, y_2, \ldots\}.$$

Now if $y_m = a_{i_1} + \ldots + a_{i_t} = a_{j_1} + \ldots + a_{j_u}$, $\{a_{i_1}, \ldots, a_{i_t}\} \subseteq U_r$, $\{a_{j_1}, \ldots, a_{j_u}\} \subseteq U_s$ for some m and $r < s$ then clearly $u < t \le d$. This implies that if we renumber the elements $y_1, y_2, \ldots$ so that $y_1 \le y_2 \le \ldots$ and $y_i = y_{i+v}$ for some i then $v \le d$. Thus we conclude that there is a sequence $Y^* = \{y_1 < y_2 < \ldots\}$ in $P(A)$ for which $d \mid y_i$ and $Y^*(n) \ge Y(n)/d \ge \sqrt{n \log n}/d^4$ or $y_i < d^9 i^2 / \log i$ $(i = 1, 2, \ldots)$.

LEMMA 2. *Let $Y = \{y_1 < y_2 < \ldots\}$ be a sequence of positive integers and let $P(Y) = \{s_1 < s_2 < \ldots\}$. Assume that there exists n^* such that for*

$n > n^*$ we have

$$y_{n+1} \le \sum_{i=1}^{n} y_i.$$

Then there is $L > 0$ such that $s_{i+1} - s_i < L$ for every i.

We omit the easy proof (see [6]).

By Lemma 2 the proof of the Proposition will be complete if we check that the sequence Y^* defined in Lemma 1 satisfies the condition $y_{n+1} \le \sum_{i=1}^{n} y_i$ for large n.

Assume contrary to the assertion that there are infinitely many n for which $y_{n+1} > \sum_{i=1}^{n} y_i$. Then

$$d^9 \frac{(n+1)^2}{\log(n+1)} > y_{n+1} > \sum_{i=1}^{n} y_i \ge \sum_{i=1}^{n} i > \frac{n^2}{2},$$

which is impossible if n is large enough. This proves the Proposition.

2. Arithmetic progressions

DEFINITION. Let $A(d,l) = \{a + kd : 0 \le k \le l\}$ be an arithmetic progression.

In this section we prove

THEOREM 2. *Let A be an infinite sequence of positive integers. Assume that $A(n) > 200\sqrt{n \log n}$ for $n > n_0$. Then there exists a $\Delta > 0$ such that for every $l \in \mathbb{N}$ there is an arithmetic progression $A(d,l) = \{u + kd : 0 \le k \le l\} \subset P(A)$ and $d < \Delta$.*

To prove Theorem 2 we shall use the following important lemma:

LEMMA 3. *Let $0 < a_1 < \ldots < a_k \le n$ be an increasing sequence of integers. Assume that $n > 2500$ and $k > 100\sqrt{n \log n}$. Then there exist integers d, b, z such that $1 \le d \le 100\sqrt{n/\log n}$, $z > \frac{1}{7} n \log n$, $b < 7z/\log n$ and*

$$\{sd : b \le s \le z\} \subseteq P(\{a_1, \ldots, a_k\}).$$

Lemma 3 is a special case of Theorem 4 in [7].

Now we prove the following

LEMMA 4. *Let $A_i := A(D_i, H_i) = \{a_i + tD_i : 0 \le t \le H_i\}$ $(i = 1, 2, \ldots)$ be an infinite sequence of arithmetic progressions. Assume that $\lim_{i\to\infty} H_i = \infty$ and*

$$H_i > D_1 + D_{i+1} \tag{2.1}$$

for every $i \ge 1$. Then for every T there is an n for which $A_1 + \ldots + A_n$ contains an arithmetic progression $A(d,h)$ with $d \le D_1$ and $h > T$.

Thus we are led to construct a long arithmetic progression with bounded difference.

Proof. We shall prove that for every n, $A_1+\ldots+A_n$ contains an $A(d,h)$, where

$$d \le D_1, \qquad h \ge H_n - D_1. \tag{2.2}$$

By the condition $\lim_{i\to\infty} H_i = \infty$, (2.2) completes the proof.

We show (2.2) by induction on n. For $n=1$, (2.2) is trivial. Assume now that $n \ge 2$ and the assertion holds with $1,\ldots,n-1$ in place of n.

By the inductive hypothesis there exists $A(d',h') \subseteq A_1+\ldots+A_{n-1}$ with $d' \le D_1, h' \ge H_{n-1} - D_1$. Since

$$A_1+\ldots+A_n = (A_1+\ldots+A_{n-1}) + A_n \supseteq A(d',h') + A_n$$

it is enough to show that there exists $A(d,h)$ with

$$A(d,h) \subseteq A(d',h') + A_n \quad \text{and} \quad d \le D_1, h \ge H_n - D_1.$$

Let $d = (d', D_n)$ and $u = d'/d, w = D_n/d$. Now $(u,w) = 1$. Then

$$\begin{aligned} A(d',h') + A_n &= \{a + td' : 0 \le t \le h'\} + \{a_n + sD_n : 0 \le s \le H_n\} \\ &= \{a + a_n + d(tu + sw) : 0 \le t \le h', 0 \le s \le H_n\}. \end{aligned}$$

It follows from a result of Frobenius (cf. [3]) that if $(u,w) = 1$ and if $t \ge w$ then every integer in the interval $[(u-1)(w-1)+1, H_n w]$ can be represented in the form

$$tu + sw, \qquad 0 \le t \le w,\ 0 \le s \le H_n.$$

By (2.1) we infer $h' \ge H_{n-1} > D_n + D_1 \ge D_n/d = w$. Thus by Frobenius' result we get

$$A(d',h') + A_n \supset A(d,h) := \{(a + a_n + duw) + rd : 0 \le r \le H_n w - uw\},$$

where $h = H_n w - uw = (H_n - u)w \ge H_n - u \ge H_n - d'/d \ge H_n - D_1$ and $d \le d' \le D_1$.

This completes the proof of the lemma.

Now define the infinite sequence of integers $[e^{20}] + 1 = n_0 < n_1 < \ldots$ where

$$n_i = n_{i-1}^2, \qquad i = 1, 2, \ldots$$

Let $B_i := (n_{i-1}, n_i] \cap A$. Now $|B_i| = A(n_i) - A(n_{i-1}) > 200\sqrt{n_i \log n_i} - n_{i-1} > 200\sqrt{n_i \log n_i} - \sqrt{n_i} > 100\sqrt{n_i \log n_i}$ since $n_i \ge n_0 = [e^{20}] + 1$. By Lemma 2 there are arithmetic progressions

$$A(D_i, H_i) = \{a_i + kD_i : 0 \le k \le H_i\} \subseteq P(B_i),$$

where

$$D_i \mid a_i, \qquad D_i \le 100\sqrt{\frac{n_i}{\log n_i}}, \qquad \frac{1}{8} n_i \log n_i < H_i \tag{2.3}$$

if n_i is large enough. Since $B_i \cap B_j = \emptyset$, for $i \neq j$ we get $A(D_1, H_1) + \ldots + A(D_n, H_n) \subset P(A)$ for every $n \in \mathbb{N}$.

Proof of Theorem 2. In view of Lemma 4 taking the arithmetic progressions $A(D_1, H_1), A(D_2, H_2), \ldots$ given above we have to show that for $i = 1, 2, \ldots,$

$$H_i > D_1 + D_{i+1}.$$

By (2.3),

$$H_i > \frac{1}{8} n_i \log n_i \geq 20e^{10} + 100 \frac{n_i}{\sqrt{\log n_i}} \geq D_1 + D_{i+1}.$$

Thus for every l there is an arithmetic progression $A(D_n, H_n) \subset P(A)$ where $H_n > l$ and $D_n < D_1$.

Theorem 2 is proved.

3. Proof of Theorem 1. Let $B = \{a_{2n-1} : n = 1, 2, \ldots\} \subset A$, $C = A \setminus B$. Now if $n > n_0$ then

$$B(n) \geq 300 \sqrt{\frac{n}{2} \log \frac{n}{2}} \geq 200 \sqrt{n \log n} \quad \text{and} \quad C(n) \geq 200 \sqrt{n \log n}.$$

By Theorem 2 there is a Δ such that for every l there is an arithmetic progression $A(d, l) = \{u + kd : 0 \leq k \leq l\} \subseteq P(B)$ and $d \leq \Delta$. Let $D = \text{l.c.m.}[1, 2, \ldots, [\Delta]]$. By the Proposition there are an L and an infinite sequence $\{x_1 < x_2 < \ldots\}$ in $P(C)$ for which $D \mid x_i$ and $x_{i+1} - x_i < L$ ($i = 1, 2, \ldots$). Now choose an arithmetic progression $A(d, l)$ contained in $P(B)$, $l > L$. Here $d < \Delta$, thus $d \mid D$ and $d \mid x_i$, $i \in \mathbb{N}$, as well.

We claim $\{kd : (x_1 + u)/d \leq k\} \subset P(A)$. Indeed, let $pd \in [x_j, x_{j+1})$, $x_j > x_1 + u$. This yields that there exists an $i \leq j$ for which $x_1 + u < pd - x_i < u + Ld$.

Now $d \mid x_i$ so $pd - x_i = u + td$, $t < L$. This means $pd = x_i + u + td \in P(A)$.

Theorem 1 is proved.

Addendum (December 8, 1999). I have learned that T. Łuczak and T. Schoen proved a theorem essentially equivalent to my Theorem 1. They obtained their result independently and later.

References

[1] J. W. S. Cassels, *On the representation of integers as sums of distinct summands taken from a fixed set*, Acta Sci. Math. (Szeged) 21 (1960), 111–124.

[2] P. Erdős, *On the representation of large integers as sums of distinct summands taken from a fixed set*, Acta Arith. 7 (1962), 345–354.

[3] P. Erdős and R. L. Graham, *On a linear diophantine problem of Frobenius*, ibid. 21 (1972), 399–408.

[4] J. Folkman, *On the representation of integers as sums of distinct terms from a fixed sequence*, Canad. J. Math. 18 (1966), 643–655.
[5] G. Freiman, *New analytical results in subset-sum problem*, Discrete Math. 114 (1993), 205–218.
[6] R. L. Graham, *Complete sequences of polynomial values*, Duke Math. J. 31 (1964), 275–286.
[7] A. Sárközy, *Finite addition theorems II*, J. Number Theory 48 (1994), 197–218.

ELTE TFK
Eötvös University
Markó u. 29
H-1055 Budapest, Hungary
E-mail: Norb@ludens.elte.hu

Received on 2.4.1996
and in revised form on 21.5.1999 (2956)