ACTA ARITHMETICA
XCV.3 (2000)

On the maximal density of sum-free sets

by

TOMASZ ŁUCZAK (Poznań) and TOMASZ SCHOEN (Kiel and Poznań)

1. Introduction. For a set $A \subseteq \mathbb{N}$, let $A(n) = |A \cap \{1, \ldots, n\}|$ and

$$\mathcal{P}(A) = \Big\{ \sum_{a \in B} a : B \subseteq A,\ 1 \le |B| < \infty \Big\}.$$

It is well known that if for some $\varepsilon > 0$ for all sufficiently large n we have $A(n) \ge n^{1/2+\varepsilon}$, then the set $\mathcal{P}(A)$ contains an infinite arithmetic progression, i.e. the following holds.

THEOREM 1. *Let $\varepsilon > 0$ and suppose that for a set $A \subseteq \mathbb{N}$ we have $A(n) \ge n^{1/2+\varepsilon}$ whenever n is large enough. Then there exist b and d such that $\mathcal{P}(A)$ contains all terms of the infinite arithmetic progression $b, b+d, b+2d, b+3d, \ldots$* ∎

Theorem 1 is due to Folkman [4], who also asked whether its assertion remains true if $\varepsilon > 0$ is replaced by a function which tends to 0 as $n \to \infty$. Theorem 2 below states that this is indeed the case and, furthermore, for every set A dense enough, one can take $b = 0$. It should be mentioned that recently a similar result has been independently proved by Hegyvári [5], who showed that the assertion of Theorem 1 holds for all $A \subseteq \mathbb{N}$ with $A(n) > 300\sqrt{n \log n}$ for n large enough.

THEOREM 2. *Let A be a set of natural numbers such that $A(n) > 402\sqrt{n \log n}$ for n large enough. Then there exists d' such that*

$$\{d', 2d', 3d', \ldots\} \subseteq \mathcal{P}(A).$$

We use Theorem 2 to estimate the maximal density of sum-free sets of natural numbers. Recall that a set $A \subseteq \mathbb{N}$ is *sum-free* if $A \cap \mathcal{P}'(A) = \emptyset$, where

$$\mathcal{P}'(A) = \Big\{ \sum_{a \in B} a : B \subseteq A,\ 2 \le |B| < \infty \Big\}.$$

2000 *Mathematics Subject Classification*: Primary 11B05.

Erdős [3] (see also Deshoulliers, Erdős and Melfi [1]) proved that the density of every sum-free set A is zero, and that for such a set A we have

$$\liminf_{n\to\infty} \frac{A(n)}{n^c} = 0$$

provided $c > (\sqrt{5}-1)/2$. As an immediate consequence of Theorem 2 we obtain the following strengthening of this result.

THEOREM 3. *If $A \subseteq \mathbb{N}$ is sum-free, then for each n_0 there exists $n \ge n_0$ such that $A(n) \le 403\sqrt{n\log n}$.*

In the last part of the note for every $\varepsilon > 0$ we construct a sum-free set A^ε such that $A^\varepsilon(n) \ge n^{1/2}\log^{-1/2-\varepsilon} n$ for all n large enough. Thus, the upper bound for the upper density of a sum-free set given by Theorem 3 is close to best possible.

2. Proofs of Theorems 2 and 3. Throughout the note by the (d,k,m)-*set* we mean the set of terms of the arithmetic progression $\{kd,(k+1)d,\ldots \ldots,(k+m)d\}$. Our argument relies on the following remarkable result of Sárközy [6, 7], which states that if a finite set A is dense enough, then $\mathcal{P}(A)$ contains large (d,k,m)-sets.

THEOREM 4. *Let $n \ge 2500$ and let A be a subset of $\{1,\ldots,n\}$ with $|A| > 200\sqrt{n\log n}$ elements. Then $\mathcal{P}(A)$ contains a (d,k,m)-set, where $1 \le d \le 10000n/|A|$, $k \le n$ and $m \ge 7^{-1}10^{-4}|A|^2 - n$.* ■

We shall also need the following simple observation.

FACT 5. *For $i = 1,2$, let A_i be a (d_i,k_i,m_i)-set, and let $m_i/2 \ge d_2 \ge d_1$. Then there exists an integer k_3 such that the set $A_1 + A_2$ contains a (d_1,k_3,m_3)-set A_3 with $m_3 \ge m_1 + m_2 - 2d_1$.*

Proof. Note that A_2 contains a (d_1d_2,k_4,m_4)-set B with $k_4 = \lceil k_2/d_1\rceil$ and $m_4 \ge (m_2-2d_1)/d_1$. Hence, A_1+B contains a $(d_1,k_1+k_4d_2,m_1+m_4d_2)$-set. ■

LEMMA 6. *Let A be a set of natural numbers such that for each n large enough, $A(n) > 201\sqrt{n\log n}$. Then there exists d such that for each $m \in \mathbb{N}$ the set $\mathcal{P}(A)$ contains a (d,k,m)-set for some k.*

Proof. For $i \ge 1$ set $n_1 = 2$ and $n_{i+1} = n_i^2 = 2^{2^i}$ and let

$$I_i = \{n \in \mathbb{N} : n_{i-1} < n \le n_i\}.$$

Since for large enough n we have $A(n) > 201\sqrt{n\log n}$, there exists $i_0 \ge 30$ such that for $i \ge i_0$ the set $A \cap I_i$ has more than $200\sqrt{n_i\log n_i}$ elements. Hence, by Theorem 4, for $i \ge i_0$, the set $\mathcal{P}(A_i)$ contains a (d_i,k_i,m_i)-set B_i with $1 \le d_i \le 50\sqrt{n_i/\log n_i}$ and $m_i \ge 10^{-3}n_i\log n_i$. Let i' be the value of index which minimizes d_i for all $i \ge i_0$ (note that $d_{i'} \le \sqrt{n_{i_0}}$). We shall show that the set $\mathcal{P}(A_{i'}) + \mathcal{P}(A_{i'+1}) + \ldots + \mathcal{P}(A_l)$ contains a $(d_{i'}, k'_l, m'_l)$-set for some k'_l and $m'_l \geq 0.001 n_l$.

We use induction on l. For $l = i'$ we have

$$m'_{i'} = m_{i'} \geq 10^{-3} n_{i'} \log n_{i'} \geq 10^{-3} n_{i'}.$$

Thus, assume that the assertion holds for $l_0 \geq i'$. By the choice of i' we have $d_{l_0+1} \geq d_{i'}$, and

$$d_{l_0+1} \leq 50\sqrt{\frac{n_{l_0+1}}{\log n_{l_0+1}}} = \frac{50 n_{l_0}}{\sqrt{\log 2^{2^{l_0}}}} < \frac{n_{l_0}}{2000} = \frac{m'_{l_0}}{2}.$$

Hence, from Fact 5 and the induction hypothesis we infer that $\mathcal{P}(A_{i'}) + \mathcal{P}(A_{i'+1}) + \ldots + \mathcal{P}(A_{l_0+1})$ contains a $(d_{i'}, k'_{l_0+1}, m'_{l_0+1})$-set for some k'_{l_0+1} and

$$\begin{aligned} m'_{l_0+1} &\geq m_{l_0+1} + 0.001 n_{l_0} - 2d_{i'} \\ &\geq 0.001 n_{l_0+1} \log n_{l_0+1} - 2\sqrt{n_{i_0}} \geq 0.001 n_{l_0+1}. \ \blacksquare \end{aligned}$$

In the proof of Theorem 2 we shall also need the following fact (see, for instance, Folkman [4]).

FACT 7. *For every natural d there exists a constant C such that for every set A of natural numbers with $A(n) \geq C\sqrt{n}$ for n large enough, there exist $r = r(d, A)$ and $k_0 = k_0(d, A)$ such that for each $k \geq k_0$,*

$$\{kd, (k+1)d, \ldots, (k+r)d\} \cap \mathcal{P}(A) \neq \emptyset. \ \blacksquare$$

Proof of Theorem 2. Let $A = \{a_1 < a_2 < \ldots\}$ and $A_1 = \{a_{2n-1} : n \in \mathbb{N}\}$, $A_2 = A \setminus A_1$. Then, for n large enough, we have $A_1(n) \geq 201\sqrt{n \log n}$. Hence, by Lemma 6, there exists d such that $\mathcal{P}(A_1)$ contains (d, k, m)-sets with arbitrarily large m. Furthermore, Fact 7 applied to A_2 implies that on the set of multiplicities of d, the set $\mathcal{P}(A_2)$ has only bounded gaps. Consequently, $\mathcal{P}(A_1) + \mathcal{P}(A_2)$ contains an infinite arithmetic progression of the form $\{k'd, (k'+1)d, \ldots\}$ and thus the assertion holds with $d' = k'd$. $\blacksquare$

Proof of Theorem 3. Let A be a set of natural numbers such that for some n_0 we have $A(n) > 403\sqrt{n \log n}$ for $n \geq n_0$. We shall show that A is not sum-free. Indeed, choose an infinite subset $A_1 \subseteq A$ such that for the set $A_2 = A \setminus A_1$ we have $A_2(n) > 402\sqrt{n \log n}$ whenever $n \geq n_0$. Theorem 3 implies that for some d and k we have

$$\{d, 2d, 3d, \ldots\} \subseteq \mathcal{P}(A_2).$$

Let $a_1, a_2 \in A_1$ be such that $a_1 \geq a_2 + d$ and $a_1 \equiv a_2 \pmod{d}$. Then $a_2 \in \{a_1 + d, a_1 + 2d, a_1 + 3d, \ldots\} \subseteq \mathcal{P}'(A)$. $\blacksquare$

3. Dense sum-free sets. We conclude the note with an example of a sum-free set A such that for each n large enough we have $A(n) \geq$

$n^{1/2}\log^{-1/2-\varepsilon} n$, where $\varepsilon > 0$ can be chosen arbitrarily small. In our construction we use a method of Deshouilliers, Erdős and Melfi [1] who showed that one can slightly "perturb" the set of all cubes to get a sum-free set. We remark that the fact that this approach can be used to build dense sum-free sets has been independently observed by Ruzsa (private communication).

Let α be an irrational number such that all terms of its continued fraction expansion are bounded, e.g. let

$$\alpha = \frac{\sqrt{5}-1}{2} = [0;1,1,1,\ldots],$$

and let $\{\alpha n\} = \alpha n - \lfloor \alpha n \rfloor$. Then the set $\{\{\alpha n\} : 1 \le n \le M\}$ is uniformly distributed in the interval $(0,1)$, i.e. the following holds (see, for instance, [2], Corollary 1.65).

Theorem 8. *For some absolute constant C and all M*

$$\sup_{0<x<y<1} \big|\,|\{\{\alpha n\} : 1 \le n \le M\} \cap (x,y)| - M(y-x)\big| \le C \log M. \ \blacksquare$$

Now let $\varepsilon > 0$ and $n_i = i^3$ for $i \ge 1$. Furthermore, set

$$A_i = \left\{ n_i \le n < n_{i+1} : \{\alpha n\} \in \left(\frac{1}{2i^{3/2}\log^{1/2+\varepsilon} i}, \frac{1}{i^{3/2}\log^{1/2+\varepsilon} i} \right) \right\},$$

and

$$A = \bigcup_{i \ge i_0} A_i,$$

where i_0 is a large natural number which will be chosen later. Using Theorem 8 we infer that for i large enough

$$|A_i| = \frac{3i^{1/2}}{2\log^{1/2+\varepsilon} i} + O(\log i),$$

and thus, for large m,

$$\sum_{i=i_0}^{m} |A_i| = \frac{3}{2}\sum_{i=i_0}^{m} \frac{i^{1/2}}{\log^{1/2+\varepsilon} i} + O(m\log m) = \frac{m^{3/2}}{\log^{1/2+\varepsilon} m} + O(m \log m).$$

Let $n_m \le n < n_{m+1}$. Then $n^{1/3} - 1 < m \le n^{1/3}$ and

$$\sum_{i=i_0}^{m-1} |A_i| \le A(n) \le \sum_{i=i_0}^{m} |A_i|.$$

Hence,

$$A(n) = \frac{n^{1/2}}{\log^{1/2+\varepsilon} n^{1/3}} + O(n^{1/3}\log n).$$

Now suppose that for some $a_1, \ldots, a_l, b \in A$ we have

$$b = a_1 + \ldots + a_l. \tag{$*$}$$

Then also

$$\{\alpha b\} \equiv \{\alpha a_1\} + \ldots + \{\alpha a_l\} \pmod 1.$$

But for i_0 large enough we have

$$\sum_{i=1}^{l} \{\alpha a_i\} \le \sum_{n\in A} \{\alpha n\} \le \sum_{i=i_0}^{\infty} \left(\frac{3i^{1/2}}{2\log^{1/2+\varepsilon} i} + O(\log i) \right) \frac{1}{i^{3/2}\log^{1/2+\varepsilon} i}$$
$$\le \sum_{i=i_0}^{\infty} \frac{2}{i\log^{1+2\varepsilon} i} < 1,$$

so that

$$\{\alpha b\} = \{\alpha a_1\} + \ldots + \{\alpha a_l\}.$$

But this is impossible, since b is larger than any of $a_1, \ldots, a_l$, and, consequently, from the definition of A,

$$\{\alpha b\} < \{\alpha a_1\} + \{\alpha a_2\}.$$

Hence the equation $(*)$ has no solutions in A, i.e. A is sum-free.

References

[1] J.-M. Deshoulliers, P. Erdős and G. Melfi, *On a question about sum-free sequences*, Discrete Math. 200 (1999), 49–54.

[2] M. Drmota and R. F. Tichy, *Sequences*, *Discrepancies and Applications*, Lecture Notes in Math. 1651, Springer, Berlin, 1997.

[3] P. Erdős, *Remarks on number theory*, *III*, Math. Lapok 13 (1962), 28–38 (in Hungarian).

[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] N. Hegyvári, *On the representation of integers as sums of distinct terms of a fixed set*, Acta Arith. 92 (2000), 99–104.

[6] A. Sárközy, *Finite addition theorems I*, J. Number Theory 32 (1989), 114–130.

[7] —, *Finite addition theorems II*, ibid. 48 (1994), 197–218.

Department of Discrete Mathematics
Adam Mickiewicz University
Matejki 48/49
60-769 Poznań, Poland
E-mail: tomasz@amu.edu.pl

Mathematisches Seminar
Universität zu Kiel
Ludewig-Meyn-Str. 4
24098 Kiel, Germany
E-mail: tos@numerik.uni-kiel.de

Received on 6.7.1999
and in revised form 19.4.2000 (3645)