ACTA ARITHMETICA
147.3 (2011)

Arithmetic progressions in sums of subsets of sparse sets

by

TOMASZ SCHOEN (Poznań)

1. Introduction. Let $A = \{a_1, \ldots, a_k\}$ be a subset of integers. Denote by

$$\mathcal{S}(A) = \Big\{ \sum_{i=1}^{k} \varepsilon_i a_i : \varepsilon_i \in \{0,1\} \Big\}$$

the subsets sumset of A and let $\mathsf{L}(S)$ stand for the length of the longest arithmetic progression in S. The problem of finding large arithmetic structures in $\mathcal{S}(A)$ (or generally in sumsets) is one of the most fundamental in combinatorial number theory. It has been intensively studied, especially in the case of sufficiently dense sets $A \subseteq \{1, \ldots, n\}$ ($|A| \geq n^{\alpha}$ for some $\alpha > 1/2$; see for example [1], [3], [4], [5]). A complete solution of this problem for sets of polynomial size was given recently by Szemerédi and Vu in [6], [7] and [8]. They proved, among other things, that if $A \subseteq [n]$ and $|A| \gg_d n^{1/d}$, where $d \geq 2$ is a fixed integer, then

$$\mathsf{L}(\mathcal{S}(A)) \gg_d |A|^{1+1/(d-1)}. \tag{1.1}$$

However, much less is known in the case of sparse sets, with $|A| = n^{o(1)}$. The only paper dealing with this question is [2], where the authors showed that for every set $A \subseteq \{1, \ldots, n\}$ we have

$$\mathsf{L}(\mathcal{S}(A)) \gg |A|/\log^2 n.$$

Actually, they proved that one can find an arithmetic progression of the form $a, 2a, \ldots, \mathsf{L}a$ in $\mathcal{S}(A)$.

The aim of this note is to provide new estimates on $\mathsf{L}(\mathcal{S}(A))$ for sets $A \subseteq \{1, \ldots, n\}$ of size $n^{o(1)}$. First, we improve the bound of Erdős and Sárközy by a $\log n$ factor. Then, we establish a bound for sets with at least $\exp(\log^{1/2+o(1)} n)$ elements, which is very similar to (1.1). Finally, we also provide some examples to show that our estimates are close to best possible.

2010 *Mathematics Subject Classification*: Primary 11B25; Secondary 11P99.
Key words and phrases: arithmetic progressions, sumsets.

DOI: 10.4064/aa147-3-7

We will use the following notation. For subsets of integers A, B we put $A+B=\{a+b : a\in A,\, b\in B\}$. By $\log x$ we always mean $\log_e x$, where $e=2.71\ldots$ is the Euler number. Furthermore, throughout the paper, we assume that n is large enough if necessary.

2. Sparse sets. By a *d-cube* we mean any set of the form

$$\mathcal{C}=\mathcal{C}(x;x_1,\ldots,x_d)=\Big\{x+\sum_{i=1}^{d}\varepsilon_i x_i : \varepsilon_i\in\{0,1\}\Big\},$$

where $x,x_1,\ldots,x_d\in\mathbb{Z}$. Observe that $\mathcal{C}=x+\{0,x_1\}+\cdots+\{0,x_d\}$. Our approach is based on the following elementary lemma.

Lemma 2.1. *Let A, B be finite sets of integers such that $|A+B|\le K|A|$. If B contains a d-cube, then*

$$\mathsf{L}(A+B)\ge\lceil (K^{1/d}-1)^{-1}\rceil+1\ge\frac{d}{(e-1)\log K}+1.$$

Proof. Suppose that $\mathcal{C}(x;x_1,\ldots,x_d)\subseteq B$ and put $T_0=A+x$, $T_i=T_{i-1}+\{0,x_i\}$ for $i=1,\ldots,d$. We have

$$T_0\subseteq T_1\subseteq\cdots\subseteq T_d\subseteq A+B,$$

so that

$$|A|=|T_0|\le|T_1|\le\cdots\le|T_d|\le K|A|.$$

Hence, for some $0\le j\le d-1$, $|T_{j+1}|=|T_j+\{0,x_{j+1}\}|\le K^{1/d}|T_j|$, so that

$$|(T_j+x_{j+1})\cap T_j|\ge(2-K^{1/d})|T_j|.$$

It is easy to see that if $|(S+x)\cap S|\ge(1-\delta)|S|$, then there is an arithmetic progression in S of length $\lceil 1/\delta\rceil$ and common difference x. Thus, $\mathsf{L}(T_j)\ge\lceil (K^{1/d}-1)^{-1}\rceil$ and

$$\mathsf{L}(A+B)\ge\mathsf{L}(T_{j+1})\ge\lceil (K^{1/d}-1)^{-1}\rceil+1.$$

If $d\ge\log K$, then

$$K^{1/d}\le 1+\frac{\log K}{d}\sum_{i\ge1}\frac{1}{i!}=1+(e-1)\frac{\log K}{d},$$

and the assertion follows. ■

Now we can improve the bound of Erdős and Sárközy.

Corollary 2.2. *If $A\subseteq\{1,\ldots,n\}$, then*

$$\mathsf{L}(\mathcal{S}(A))\ge\frac{|A|}{4(e-1)\log n}.$$

Proof. Observe that $\mathcal{S}(A) \subseteq [n^2]$ and $\mathcal{C}(a_1; a_2, \ldots, a_k) \subseteq \mathcal{S}(A)$, where $A = \{a_1, \ldots, a_k\}$. Put

$$\mathcal{C}_1 = \{0\} \quad \text{and} \quad \mathcal{C}_2 = \mathcal{C}(a_1; a_2, \ldots, a_k).$$

Clearly, $\mathcal{C}_1 + \mathcal{C}_2 \subseteq \mathcal{S}(A)$, so that

$$|\mathcal{C}_1 + \mathcal{C}_2| = |\mathcal{C}_2|\,|\mathcal{C}_1| \le |\mathcal{S}(A)|\,|\mathcal{C}_1| \le n^2|\mathcal{C}_1|.$$

By Lemma 2.1 applied with $A = \mathcal{C}_1, B = \mathcal{C}_2$ and $K = n^2$ we have

$$\mathsf{L}(\mathcal{S}(A)) \ge \mathsf{L}(\mathcal{C}) \ge \frac{d(\mathcal{C}_2)}{2(e-1)\log n^2} + 1 \ge \frac{|A|}{4(e-1)\log n}. \ \blacksquare$$

The set of powers of 3 contained in $\{1, \ldots, n\}$ shows that in general one cannot improve the bound given by Corollary 2.2. However, in the next section, we show that this is possible for sufficiently dense sets.

3. Sets of size at least $e^{(\log n)^{1/2+o(1)}}$. We start with a version of Lemma 2.1 more suitable for our purpose.

LEMMA 3.1. *Suppose that $A \subseteq \{1, \ldots, n\}$ and $|A| \ge 100 \log n$. Then there exists a subset A_1 of A such that $|A_1| \ge |A|/2$ and for any $x_1, \ldots, x_h \in A_1$ and any integers $i_1, \ldots, i_h$ there is an arithmetic progression in $\mathcal{S}(A \setminus A_1)$ of length $|A|(10\log n)^{-1}(\sum_j |i_j|)^{-1}$ and common difference $\sum_j i_j x_j$.*

Proof. We define a set $B = \{a_1, \ldots, a_l\} \subseteq A$ recursively in the following way. Let $a_1 \in A$ be arbitrary. Suppose that $B_i = \{a_1, \ldots, a_i\}$ has already been chosen. If there is $a_{i+1} \in A \setminus \{a_1, \ldots, a_i\}$ such that

$$|\mathcal{S}(B_i \cup \{a_{i+1}\})| \ge (1 + 5|A|^{-1}\log n)|\mathcal{S}(B_i)|, \tag{3.1}$$

then we define $B_{i+1} = \{a_1, \ldots, a_{i+1}\}$. If such an element does not exist, we stop the algorithm putting $B = B_i$. We have to prove that this procedure terminates after at most $|A|/2$ steps. Clearly, for each i,

$$|\mathcal{S}(B_i)| \ge (1 + 5|A|^{-1}\log n)^{i-1}.$$

Thus, using the inequality $(1 + 1/t)^{t+1/2} > e$ for $t > 0$, we see that

$$|\mathcal{S}(B_i)| \ge e^{\frac{5(i-1)\log n}{|A| + 2\log n}}.$$

On the other hand $|\mathcal{S}(B_i)| \le |\mathcal{S}(A)| \le n^2$, so that $l \le |A|/2$. We show that our assertion holds for $A_1 = A \setminus B$. Suppose that $x_1, \ldots, x_h \in A_1$ and $i_1, \ldots, i_h \in \mathbb{Z}$. By (3.1), for every $1 \le j \le h$ we have

$$|\mathcal{S}(B) \cap (\mathcal{S}(B) + x_j)| \ge (1 - 10|A|^{-1}\log n)|\mathcal{S}(B)|,$$

hence

$$\Big|\mathcal{S}(B) \cap \Big(\mathcal{S}(B) + \sum_j i_j x_j\Big)\Big| \ge \Big(1 - 10|A|^{-1}\sum_j |i_j| \log n\Big)|\mathcal{S}(B)|,$$

so $\mathcal{S}(B)$ contains an arithmetic progression of length $|A|(8\sum_j |i_j|\log n)^{-1}$ and common difference $\sum_j i_j x_j$. ∎

Now we are ready to prove the main result of this section.

THEOREM 3.2. *Suppose that* $A \subseteq \{1,\ldots,n\}$ *and* $|A| \ge 8(n/\log n)^{1/d}$, *where* $0 < d \le \sqrt{\frac{\log n}{2\log\log n}}$ *is an integer. Then*

$$\mathsf{L}(\mathcal{S}(A)) \ge 2^{-10}(|A|/\log n)^{1+1/d}.$$

Proof. Let A_1 and A_2 be the sets given by Lemma 3.1 applied for A and A_1, respectively. We have $|A_1| \ge |A|/2$ and $|A_2| \ge |A|/4$.

Put $t = \lfloor (|A|/\log n)^{1/d}\rfloor$. By hypothesis it follows that $t \ge 2$. We show that the equation

$$x_0 + tx_1 + \cdots + t^d x_d = y_0 + ty_1 + \cdots + t^d y_d \tag{3.2}$$

can be nontrivially (this means $x_i \ne y_i$ for some i) solved in A_2. Indeed, all sums $x_1 + tx_2 + \cdots + t^d x_{d+1},\ x_i \in A_2$, are less than $2t^d n$. Since $|A_2|^{d+1} \ge 2t^d n$ two among them are equal, giving a nontrivial solution $x_0,\ldots,x_d,y_0,\ldots,y_d \in A_2$. Denote by i the largest index $0 \le j \le d$ such that $x_j \ne y_j$. Then

$$t^i(x_i - y_i) = (y_1 - x_1) + \cdots + t^{i-1}(y_{i-1} - x_{i-1}) \ne 0.$$

Put $b = x_i - y_i$ and $c = (y_0 - x_0) + \cdots + t^{i-1}(y_{i-1} - x_{i-1})$. By Lemma 3.1, $\mathcal{S}(A\setminus A_1)$ contains an arithmetic progression P_1 of length $|A_1|/(20\log n) \ge |A|/(40\log n)$ and common difference b, and $\mathcal{S}(A_1\setminus A_2)$ contains an arithmetic progression P_2 of length

$$|A_2|/(10(1+t+\cdots+t^{i-1})\log n) \ge |A|/(80t^{i-1}\log n)$$

and common difference c. As $tb = c$, $P_1 + P_2$ is an arithmetic progression of length at least $2^{-9}t|A|/\log n \ge 2^{-10}(|A|/\log n)^{1+1/d}$ and common difference b. The assertion follows from the inclusion

$$P_1 + P_2 \subseteq \mathcal{S}(A\setminus A_1) + \mathcal{S}(A_1\setminus A_2) \subseteq \mathcal{S}(A). \ \blacksquare$$

Using the same argument one can prove an analogous result for $\mathbb{Z}/p\mathbb{Z}$. An improvement relies on the fact that the equation (3.2) can be solved in $\mathbb{Z}/p\mathbb{Z}$ in sparser subsets.

THEOREM 3.3. *Let* $A \subseteq \mathbb{Z}/p\mathbb{Z}$ *and* $|A| \ge 8(p/\log p)^{1/d}$, *where* $0 < d \le \sqrt{\frac{\log p}{2\log\log p}}$ *is an integer. Then*

$$\mathsf{L}(\mathcal{S}(A)) \ge 2^{-10}(|A|/\log p)^{1+1/(d-1)}.$$

4. A construction of sets with small $\mathsf{L}(\mathcal{S}(A))$. We prove here that bounds given by Corollary 2.2 and Theorem 3.2 are near optimal. First, we observe that an example of sets given by Szemerédi and Vu can be applied for all $d \ll \sqrt{\log n}$. We recall briefly their construction. Let $n, d, m \in \mathbb{N}$ with $2m^d \le n^{1/d}$ and set

$$P = \Big\{ \sum_{i=0}^{d-1} x_i (2m^{d+1})^i : 1 \le x_i \le m \Big\}.$$

Thus,

$$\mathcal{S}(P) \subseteq \Big\{ \sum_{i=0}^{d-1} x_i (2m^{d+1})^i : 1 \le x_i \le m^{d+1} \Big\}.$$

Clearly $P \subseteq \{1, \dots, n\}$ is a d-dimensional, proper generalized arithmetic progression, hence $|P| = m^d$ and $\mathcal{S}(P)$ is 2-Freiman isomorphic (see [9]) to a subset of $[m^{d+1}]^d$, so that $\mathsf{L}(\mathcal{S}(P)) \le m^{d+1} = |P|^{1+1/d}$. In particular, for every k there is a subset of $\{1, \dots, 2^{k^2}\}$ of size 2^{k-1} containing an arithmetic progression of length at most 2^k in its subsets sumset. Now, we adapt this construction to sparser sets.

THEOREM 4.1. *For every positive integer n and $100 \log n \le t \le 2^{\sqrt{\log_2 n} - 1}$ there is a set $A \subseteq \{1, \dots, n\}$ such that $|A| = t$ and*

$$\mathsf{L}(\mathcal{S}(A)) \le 10 \frac{|A|}{\log_2 n} \Big(\log_2 \frac{|A|}{\log_2 n} \Big)^2.$$

Proof. Put

$$K = \frac{10t}{\log_2 n} \Big(\log_2 \frac{t}{\log_2 n} \Big)^2 \quad \text{and} \quad k = \lfloor \log K \rfloor.$$

By the construction above, there is a set $X \subseteq \{1, \dots, 2^{k^2}\}$ such that $|X| = 2^{k-1}$ and $\mathsf{L}(\mathcal{S}(X)) \le 2^k$. Let A consist of all numbers of the form $x 2^{i(k^2+k)}$, where $x \in X$ and $0 \le i \le l = \lfloor \log_2 n/(k^2+k) - 1 \rfloor$. Clearly, $A \subseteq \{1, \dots, n\}$ and

$$|A| = (l+1) 2^{k-1} \ge \frac{\log_2 n}{4 \log_2^2 K} \frac{10t}{\log_2 n} \Big(\log_2 \frac{t}{\log_2 n} \Big)^2 \ge t.$$

Observe that

$$\mathcal{S}(A) = \Big\{ \sum_{i=0}^{l} x 2^{i(k^2+k)} : x \in \mathcal{S}(X) \Big\}.$$

Therefore, from $\sum_{x \in X} x < 2^{k^2+k-1}$, we deduce that $\mathcal{S}(A)$ is 2-Freiman isomorphic to $\mathcal{S}(X) \times \dots \times \mathcal{S}(X)$ ($l+1$ times), hence

$$\mathsf{L}(\mathcal{S}(A)) = \mathsf{L}(\mathcal{S}(X)) \le 2^k \le K = \frac{10t}{\log_2 n} \Big(\log_2 \frac{t}{\log_2 n} \Big)^2,$$

and the assertion follows. ∎

5. Concluding remarks. The bound (1.1) of Szemerédi and Vu and the examples above support the claim that our estimates are not optimal. One would like to replace d by $d-1$ and remove the $\log n$ factor in Theorem 3.2. It also seems that Corollary 2.2 is not best possible for sets of size $\omega(n)\log n$ if $\omega(n)\to\infty$ as $n\to\infty$. Essentially there are two places where our argument could be refined. More specifically, the results can be strengthen provided we can solve any of the following three problems.

PROBLEM 5.1. *Is it true that for every $A\subseteq[n]$, there is a subset $A'\subseteq A$ of size (roughly) $|A|/2$ such that*

$$|\mathcal{S}(A')|\gg|\mathcal{S}(A)|^{1-\varepsilon},$$

where $\varepsilon\to 0$ as $|A|/\log n\to\infty$?

If we can answer the above question in the affirmative, then by Lemma 2.1 applied with $\mathcal{S}(A')$, $\mathcal{S}(A\setminus A')$ and $K=|\mathcal{S}(A)|^{\varepsilon}$ we have

$$|\mathcal{S}(A')+\mathcal{S}(A\setminus A')|\le|\mathcal{S}(A)|\ll K|\mathcal{S}(A')|,$$

so

$$\mathsf{L}(\mathcal{S}(A'))\gg|A|/(\varepsilon\log n),$$

which improves Corollary 2.2. It would even be sufficient for us if the following weaker question had a positive answer.

PROBLEM 5.2. *Is it true that for every set A there is a subset $B\subseteq A$ with $|B|\ge|A|/2$ having the property described in Problem 5.1?*

Probably one can also replace d by $d-1$. To do it one has to find a more efficient argument than the one we used, based on solutions to a linear equation. This could be done if the following question had a positive answer.

PROBLEM 5.3. *Is it true that there exists an absolute constant $C>0$ such that for every set $A\subseteq[n]$ with $|A|>Cn^{1/d}$ (we may allow C to depend at most exponentially on d) there are $a_1,\dots,a_u,b_1,\dots,b_v\in\pm A$ and $M\in\mathbb{Z}$ such that*

$$M\sum_{i=1}^{u}a_i=\sum_{j=1}^{v}b_j\neq 0$$

and $|A|/u>M>|A|^{1/(d-1)}v$?

Acknowledgements. This research was partly supported by MNSW grant N N201 543538.

References

[1] N. Alon and G. Freiman, *On sums of subsets of a set of integers*, Combinatorica 8 (1988), 297–306.

[2] P. Erdős and A. Sárközy, *Arithmetic progressions in subset sums*, Discrete Math. 102 (1992), 249–264.
[3] G. Freiman, *New analytical results in subset-sum problem*, ibid. 114 (1993), 205–217.
[4] E. Lipkin, *On representation of rth powers by subset sums*, Acta Arith. 52 (1989), 353–365.
[5] A. Sárközy, *Finite addition theorems. II*, J. Number Theory 48 (1994), 197–218.
[6] E. Szemerédi and V. Vu, *Finite and infinite arithmetic progressions in sumsets*, Ann. of Math. (2) 163 (2006), 1–35.
[7] —, —, *Long arithmetic progressions in sumsets: thresholds and bounds*, J. Amer. Math. Soc. 19 (2006), 119–169.
[8] —, —, *Long arithmetic progressions in sum-sets and the number of x-sum-free sets*, Proc. London Math. Soc. (3) 90 (2005), 273–296.
[9] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Univ. Press, 2006.

Tomasz Schoen
Faculty of Mathematics and Computer Science
Adam Mickiewicz University
Umultowska 87
61-614 Poznań, Poland
E-mail: schoen@amu.edu.pl

Received on 18.3.2010
and in revised form on 29.6.2010 (6336)