ACTA ARITHMETICA
106.2 (2003)

Blocks and progressions in subset sum sets

by

Vsevolod F. Lev (Haifa)

0. The summary. Let A be a finite set of positive integers and let A^* be the set of all subset sums of A. We show that if A is dense enough (say, $A \subseteq [1, l]$ and $|A| \geq 5(l \ln l)^{1/2}$), then A^* contains a long block of consecutive integers or at least a long homogeneous arithmetic progression. This refines earlier results due to Freiman and Sárközy.

1. The background. Most problems in additive number theory share, with some deviations, the same common pattern: given a set A of positive integers, show that the set of all numbers representable as a sum of elements of A is "large". The exact meaning of "large" can vary: one may actually count the numbers representable, or may wish to prove that there are, say, perfect squares, or prime numbers among them. Here we are concerned with the situation when A is finite, the summands are pairwise distinct, and any number of summands is allowed. (For the case of A infinite see Appendix; for representations with repetitions see [S89, Le97]; for representations with a fixed number of summands see Section 4.)

We let

$$A^* := \{a_1 + \ldots + a_k : a_1 < \ldots < a_k,\ a_1, \ldots, a_k \in A\},$$

the set of all integers representable as a sum of a number of pairwise distinct elements of A. Plainly, $A^* \subseteq [0, \sigma(A)]$, where $\sigma(A)$ is the sum of all elements of A; moreover, A^* is symmetric about $\sigma(A)/2$, the midpoint of the interval: $s \in A^*$ if and only if $\sigma(A) - s \in A^*$. It is the set A^* which is to be shown "large".

Perhaps, the most radical approach is to prove that A^* contains long blocks of consecutive integers, provided that A is sufficiently dense. It is easily seen, however, that A^* may fail to contain such blocks; this happens, for instance, if all or almost all elements of A have a common divisor $d > 1$. For this reason, instead of a block of integers one may seek in A^* an arith-

2000 *Mathematics Subject Classification*: Primary 11B05; Secondary 11B13, 11B25, 05D05, 11P99.

metic progression of the form

$$\{(m+1)d, (m+2)d, \dots, (m+M)d\}.$$

Such progressions are called *homogeneous*; they are just sets of consecutive multiples of an integer number d.

Two results in this direction were established independently and almost simultaneously by Freiman and Sárközy.

THEOREM A (Freiman, [Fr93]). *There exist absolute positive constants c_1 and c_2 with the following property. Let $A \subseteq [1,l]$ be a set of $n = |A|$ integers satisfying $n > c_1(l\ln l)^{1/2}$. Write $\sigma(A) = \sum_{a\in A} a$. Then there is a positive integer $d \le 3l/n$ such that A^* contains all multiples of d which fall into the interval $\left[\frac{1}{2}\,\sigma(A) - c_2dn^2, \frac{1}{2}\,\sigma(A) + c_2dn^2\right]$.*

THEOREM B (Sárközy, [S94, Theorem 4]). *Let $A \subseteq [1,l]$ be a set of $n = |A|$ integers satisfying*

$$l > 2500, \quad n > 200(l\ln l)^{1/2}.$$

Then there are integers y, z, and $d \ge 1$ such that

$$d < 10^4\,\frac{l}{n}, \qquad z > 7^{-1}10^{-4}n^2, \qquad \frac{z}{y} > 7^{-1}10^{-4}\,\frac{n^2}{l},$$

and $\{yd, (y+1)d, \dots, zd\} \subseteq A^*$.

Both theorems impose similar density restrictions on A, both give similar bounds for the difference d and the number of terms of the progression. The minor discrepancy is that Theorem A guarantees that the progression is centered about $\sigma(A)/2$, while Theorem B provides, instead, a lower bound for the "logarithmic length" (z/y) of the progression.

There are a number of papers where other results of this kind are obtained; we mention [AF88, Li89, C90, EF90, GM91, Le98, C99]. Theorems A and B, however, are most applicable and strongest up to date in the sense that the assumption $n \gg (l\ln l)^{1/2}$ is less restrictive than those made elsewhere. (The only exception is [C99, Corollary 2.9] which is based on Theorem A and inherits its assumption.)

2. The results. We develop Theorems A and B in the following directions.

First, we extend significantly the length of the progression which is guaranteed to exist in A^* and show that in fact this progression stretches almost onto the whole interval $[0, \sigma(A)]$.

Second, we formulate an arithmetic condition such that if A satisfies this condition, then the progression in question has difference one and is, therefore, a block of consecutive integers. Our condition is, in a sense, the weakest possible.

Third, we get rid of unspecified or excessively large constants replacing them by reasonably small ones.

For y and z real, $y \le z$, we denote by $[y, z]$ the set of all integer numbers between y and z. Given a positive integer q, by $N_q(A)$ we denote the number of elements of A *not* divisible by q.

THEOREM 1. *Let $A \subseteq [1, l]$ be a set of $n = |A| \ge n_0$ integers, where n_0 is a sufficiently large absolute constant. Suppose that*

$$n \ge 20(l \ln n)^{1/2}$$

and that for any positive integer $q < 2l/n$ we have $N_q(A) \ge q - 1$. Then

$$[\lambda\sigma(A), (1-\lambda)\sigma(A)] \subseteq A^*,$$

where $\sigma(A)$ is the sum of all elements of A and $\lambda = 280l/n^2$.

We postpone the proof of this and other results of this paper (that will be presented shortly) until Sections 4–7.

REMARKS. 1. The value of λ cannot be replaced by $\lambda_0 = l/(2n^2)$: for any l and n such that $3n \le l \le n^2/6$ there exist an n-element set $A \subseteq [1, l]$ satisfying $N_q(A) \ge q - 1$ for any $q < 2l/n$ and an integer $s \in [\lambda_0\sigma(A), (1-\lambda_0)\sigma(A)]$ such that $s \notin A^*$. To see this, take $A = [l - (n-1), l]$, $h = \lfloor l/(n-1) \rfloor - 1 > l/(2(n-1)) > l/(2n)$ and $s = hl$, so that $\lambda_0\sigma(A) \le s \le (1-\lambda_0)\sigma(A)$ (as follows by an easy verification). Then s cannot be represented as a sum of $h+1$ or more distinct elements of A, for any such sum is greater than $(h+1)(l-(n-1)) = s + (l - (h+1)(n-1)) \ge s$; and similarly, s cannot be represented as a sum of h or less distinct elements of A.

2. If A contains less than $q-1$ elements not divisible by q for an integer $q \ge 1$, then it can happen that some residue classes modulo q are not represented in A^*. (For instance, consider the situation when each $a \in A$ is either zero or one modulo q.) Clearly, in this case A^* does not contain q consecutive integers.

3. It is easily seen that for any fixed $\varepsilon > 0$ and n large enough, the condition $n \ge 20(l \ln n)^{1/2}$ follows from $n \ge (10\sqrt{2} + \varepsilon)(l \ln l)^{1/2}$. Thus, we have in effect replaced Freiman's constant c_1 and Sárközy's constant 200 by the constant $10\sqrt{2}$. It will be seen later that $10\sqrt{2}$ can be further reduced to a value smaller than 5, but this comes at the expense of decreasing the block length.

As a corollary, we show that even if A fails to satisfy $N_q(A) \ge q - 1$ for all $q < 2l/n$, the subset sum set A^* still contains a long homogeneous progression. Let $\sigma(A)$, λ, and $N_q(A)$ be as in Theorem 1.

COROLLARY 1. *Let $A \subseteq [1,l]$ be a set of $n = |A| \ge n_0$ integers, where n_0 is a sufficiently large absolute constant. Suppose that*

$$n \ge 20(l \ln n)^{1/2}.$$

Then there exists a positive integer $d < 2l/n$ such that A^* contains all multiples of d that fall into the interval

$$I_A := [\lambda\sigma(A), (1-\lambda)\sigma(A)].$$

Moreover, if $N_q(A) \ge q-1$ for any positive integer $q < 2l/n$, then one can take $d = 1$. Otherwise, one can take d to be the maximal number $q < 2l/n$ for which $N_q(A) < q-1$; in this case the set of all multiples of d that fall into I_A is contained already in A_d^*, where $A_d = \{a \in A : a \equiv 0 \pmod d\}$.

In applications, one is often less concerned with the block (or progression) length and cares more about the density requirement and small values of n. For this reason we also provide a result which merely sharpens the constants of Theorem B.

THEOREM 2. *Let $A \subseteq [1,l]$ be a set of $n = |A|$ integers satisfying*

$$n > 10(l \ln n)^{1/2}$$

(equivalently, $l < 0.01n^2/\ln n$). Then there are integers y, z, and $d \ge 1$ such that

$$d < 6\,\frac{l}{n}, \qquad z - y > \frac{n^2}{20}, \qquad \frac{z-y}{y} > \frac{n^2}{115l},$$

and $\{yd, (y+1)d, \dots, zd\} \subseteq A^*$.

Since the factor 10 (in the condition $n > 10(l \ln n)^{1/2}$) is often more important than other constants, we give yet another version of this theorem.

THEOREM 2′. *Let $A \subseteq [1,l]$ be a set of $n = |A| \ge 3803$ integers satisfying*

$$n > 7(l \ln n)^{1/2}.$$

Then there are integers y, z, and $d \ge 1$ such that

$$d < 5\,\frac{l}{n}, \qquad z - y > \frac{n^2}{405}, \qquad \frac{z-y}{y} > \frac{n^2}{2425l},$$

and $\{yd, (y+1)d, \dots, zd\} \subseteq A^*$.

REMARK. Observe that for any fixed $\varepsilon > 0$ and n large enough, the condition $n > 7(l \ln n)^{1/2}$ follows from $n > (7/\sqrt{2}+\varepsilon)(l \ln l)^{1/2}$; furthermore, $7/\sqrt{2} \approx 4.949$.

3. The method. We prove Theorems 2 and 2′ in Section 5 and then derive Theorem 1 in Sections 6 and 7. Our proof of Theorems 2 and 2′ follows the lines of [S94] where their prototype, Theorem B, is established.

To prove Theorem B, Sárközy uses a result about arithmetic progressions in the set

$$h \cdot A := \{a_1 + \ldots + a_h : a_1 < \ldots < a_h,\, a_1, \ldots, a_h \in A\}$$

(all integers representable as a sum of precisely h pairwise distinct elements of A).

THEOREM C (Sárközy, [S94, Theorem 3]). *Let $A \subseteq [1, l]$ be a set of $n = |A|$ integers such that*

$$l > 2500, \qquad n > 100(l \ln l)^{1/2}.$$

Write $L = \ln(13l/n)$. Then for every integer M satisfying

$$l \le M \le 10^{-4}\,\frac{n^2}{L}$$

there exist positive integers d and h such that

$$d < 4828\,\frac{l}{n}, \qquad h < 8496\,\frac{M}{n}$$

and the set $h \cdot A$ contains an M-term homogeneous arithmetic progression of difference d.

As shown in [S94], this theorem is best possible save for the constants and logarithmic factors.

To upgrade Theorem B to Theorems 2 and 2′ we establish in Section 4 a refined version of Theorem C.

THEOREM 3. *Let $A \subseteq [1, l]$ be a set of $n = |A|$ integers. Write $L = \ln(4l/n)$. Then for every M satisfying*

$$4l \le M \le \frac{n^2}{12L}$$

there exist positive integers d and h such that

$$d < 4\,\frac{l}{n}, \qquad h < 6\,\frac{M}{n},$$

and the set $h \cdot A$ contains a homogeneous arithmetic progression of difference d and at least $M + 1$ terms. Moreover, d and h can be so chosen that $2d$ divides h.

REMARKS. 1. Though in Theorem 3 we do not impose any explicit restrictions on l or n, some are implied by the assumptions. For instance, from $l \ge n$ it follows that $L \ge \ln 4 > 4/3$, whence $n^2 \ge 48lL > 64n$, $n > 64$; we use this observation in the proof in Section 4.

2. Some further minor improvements of the constants are possible. Also, there is the usual trade-off: some of the constants can be improved at the expense of the others. In fact, for any $\varepsilon > 0$ we can find $c = c(\varepsilon) > 0$ and $C = C(\varepsilon) > 0$ such that if $cl \le M \le n^2/(CL)$, then d and h exist so that $d < (1 + \varepsilon)l/n$ and $h < (1 + \varepsilon)M/n$.

Once Theorem 3 is proven, Theorems 2 and 2′ are relatively easy to deduce; this is done in Section 5.

Theorem 1 is derived from Theorem 2 in Sections 6 and 7. The proof uses some ideas from [AF88, C99]. Corollary 1 is proven in Section 8.

The reason why we believe that strengthening Theorems A and B is worthwhile is that they are used in the proofs of many important results in additive number theory. Replacing these theorems with Theorems 1, 2, and 2′ can lead to improvements in these results. We give an example in the Appendix.

Notation. Throughout the rest of the paper we continue to use the above introduced notation, unless indicated otherwise. Specifically, A is a finite, non-empty set of positive integers; l is a real number such that $A \subseteq [1, l]$; the cardinality of A is $n = |A|$; the sum of the elements of A is denoted by $\sigma(A)$; the number of elements of A not divisible by an integer $q \geq 1$ is denoted by $N_q(A)$; the set of all numbers representable as a sum of precisely h distinct elements of A is $h \cdot A$, and $A^* = \bigcup_{h=0}^{n} h \cdot A$ is the subset sum set of A; finally, $\lambda = 280l/n^2$. Also, we write hA for the set of all numbers representable as a sum of precisely h *not necessarily distinct* elements of A:

$$hA = \{a_1 + \ldots + a_h : a_1, \ldots, a_h \in A\}.$$

4. Fixed number of summands: proof of Theorem 3. Our proof of Theorem 3 employs the same idea as that of Theorem C. However, we replace the key component of the proof, Sárközy's theorem [S89, Theorem 1], by the following result of ours.

Theorem 4 (Lev, [Le97, Theorem 3(ii)]). *Let $S \subseteq [0, l]$ be a set of $|S| \geq 3$ integers such that $0, l \in S$ and $\gcd(S) = 1$. Write*

$$k = \left\lfloor \frac{l-1}{|S|-2} \right\rfloor, \qquad \varrho = (k+1)(|S|-2) + 2 - l.$$

Suppose that $h \geq 3k$ is an integer number. Then hS contains a block of $(h-k)l + k\varrho + 1$ consecutive integers.

Another new ingredient we introduce in the proof is the following lemma.

Lemma 1. *Let A be a set of $n = |A|$ integers. For $t = 1, 2, \ldots$ denote by S_t the set of all integers $s \in 2A$ which have at least t representations as $s = a_1 + a_2$ $(a_1, a_2 \in A)$, and let $N_t = |S_t|$ be the cardinality of S_t. Then*

$$N_1 + \ldots + N_\tau \geq 2n\tau - \tau^2$$

for any $\tau = 1, \ldots, n$.

Remark. The case $\tau = n$ is immediate (with equality sign) and was utilized in the original argument of Sárközy. The case $\tau = 1$ is easy.

Proof of Lemma 1. The shortest way to prove the lemma is to derive it from a result of Pollard [P74] which says that if p is a prime number, A_1 and A_2 are non-empty sets of residues modulo p, and N_t is the number of residues modulo p with at least t representations as $a_1 + a_2$ ($a_i \in A_i$), then

$$N_1 + \ldots + N_\tau \geq \tau \min\{|A_1| + |A_2| - \tau, p\}$$

for any $\tau = 1, \ldots, \min\{|A_1|, |A_2|\}$. Now choose $A_1 = A_2$ to be the canonical image of A in $\mathbb{Z}/p\mathbb{Z}$, where p is a sufficiently large prime number. ∎

From now on, we adopt S_t and N_t (not to be confused with $N_q(A)$!) as a standard notation.

Corollary 2. *Let A, n, l, and L be as in Theorem 3. Then there exists $t \in [1, n]$ such that*

$$N_t > \frac{n}{2} \max\left\{\frac{n}{tL}, 1\right\} + 2. \tag{1}$$

Proof. Assume the opposite:

$$N_t \leq \frac{n}{2} \max\left\{\frac{n}{tL}, 1\right\} + 2, \quad t = 1, \ldots, n. \tag{2}$$

Set $\tau = \lfloor 3n/4 \rfloor$. Taking into account that $N_t < 2l$ for any t, by Lemma 1 and (2) we get

$$\begin{aligned}
2n\tau - \tau^2 &\leq N_1 + \ldots + N_\tau \\
&< \sum_{1 \leq t \leq n^2/(4lL)} 2l + \sum_{n^2/(4lL) < t \leq n/L} \left(\frac{1}{2} \cdot \frac{n^2}{tL} + 2\right) + \sum_{n/L < t \leq \tau} \left(\frac{1}{2}\, n + 2\right) \\
&\leq \frac{n^2}{2L} + 2(\tau - 1) + \frac{1}{2} \cdot \frac{n^2}{L} \left(\frac{4lL}{n^2} + \ln \frac{4l}{n}\right) + \frac{1}{2}\, n \left(\tau + 1 - \frac{n}{L}\right) \\
&= \frac{1}{2}\, n^2 + \frac{1}{2}\, n\tau + \left(2l + \frac{1}{2}\, n + 2\tau - 2\right).
\end{aligned}$$

(We use here the estimate $\sum_{P < t \leq Q} t^{-1} \leq P^{-1} + \ln(Q/P)$. Also, the number of summands in $\sum_{n/L < t \leq \tau}$ is at most $\tau + 1 - n/L$, as $\tau + 1 > 3n/4 > n/L$ in view of $L > 4/3$.) This implies that

$$\begin{aligned}
\frac{1}{16}\, n^2 &< \left(\tau - \frac{3}{4}\, n\right)^2 + \left(2l + \frac{1}{2}\, n + 2\tau - 2\right) < 2l + \frac{1}{2}\, n + 2\tau - 1 \\
&\leq 2l + 2n - 1, \\
l &> \frac{n^2}{32} - n > \frac{n^2}{64} > \frac{n^2}{48L},
\end{aligned}$$

contrary to the assumption that $4l \leq M \leq n^2/(12L)$. ∎

Our next lemma is parallel to [S94, Lemma 1].

LEMMA 2. *Let A be a set of integers, and suppose that $h < \lceil (t+3)/4 \rceil$ is a positive integer. Then $hS_t \subseteq 2h \cdot A$.*

Sketch of the proof. Use induction on $j = 1, \dots, h$ to show that $jS_t \subseteq 2j \cdot A$. Let $s_1 + \ldots + s_{j-1} + s_j$ be an element of jS_t. By the induction hypothesis, there is a representation $s_1 + \ldots + s_{j-1} = a_1 + \ldots + a_{2j-2}$ with $a_i \in A$ all distinct, and the assumption on h ensures that s_j has a representation $s_j = a' + a''$ with $a' \neq a''$ and both a' and a'' distinct from all a_i. ∎

Proof of Theorem 3. We choose $t \in [1, n]$ satisfying (1), and we put

$$d := \gcd(S_t - S_t), \qquad S_t' := \{d^{-1}(s - \min(S_t)) : s \in S_t\}, \qquad l' := \max(S_t').$$

Thus, $S_t' \subseteq [0, l']$, $0, l' \in S_t'$, $\gcd(S_t') = 1$, and $|S_t'| = N_t$. Set

$$k := \left\lfloor \frac{l'-1}{N_t - 2} \right\rfloor, \qquad \varrho := (k+1)(N_t - 2) + 2 - l'.$$

For future reference we note that

$$k(N_t - 2) < l' \le \frac{1}{d}\,(\max(S_t) - \min(S_t)) < \frac{2l}{d}, \tag{3}$$

hence by (1) and $4l \le M \le n^2/(12L)$,

$$d < \frac{2l}{k(N_t-2)} < \frac{4l}{kn}\,\min\left\{\frac{tL}{n}, 1\right\}, \tag{4}$$

$$k < \frac{4l}{dn}\,\min\left\{\frac{tL}{n}, 1\right\}, \tag{5}$$

$$t > \frac{kdn^2}{4lL} \ge \frac{3kdM}{l} \ge 12kd. \tag{6}$$

We now define

$$h_0 := \max\left\{2k, \left\lceil \frac{M - k\varrho}{l'} \right\rceil\right\} + k + \delta, \qquad h = 2h_0,$$

where $\delta \in [0, d-1]$ is chosen so that $d \mid h_0$. As $h_0 \ge 3k$, by Theorem 4 the set $h_0 S_t'$ contains a block of at least $(h_0 - k)l' + k\varrho + 1 \ge \lceil (M - k\varrho)/l' \rceil l' + k\varrho + 1 \ge M + 1$ consecutive integers, hence $h_0 S_t$ contains a homogeneous arithmetic progression of difference d and at least $M + 1$ terms. To complete the proof it suffices to show that

$$h_0 < \min\left\{\frac{t}{4}, 3\,\frac{M}{n}\right\} \tag{7}$$

(the estimate $d < 4l/n$ was already established in (4)): then by Lemma 2 we have $h \cdot A = 2h_0 \cdot A \supseteq h_0 S_t$, whence $h \cdot A$ contains a homogeneous arithmetic progression of difference d and at least $M + 1$ terms.

Evidently, (7) will follow from

$$3k+d-1 < \min\left\{\frac{t}{4},\, 3\,\frac{M}{n}\right\}, \tag{8}$$

and

$$\frac{M-k\varrho}{l'}+k+d < \min\left\{\frac{t}{4},\, 3\,\frac{M}{n}\right\}. \tag{9}$$

To prove (8) we observe that by (4),

$$3k+d-1 < 3k+\frac{4l}{kn}\,\min\left\{\frac{tL}{n},1\right\}-1.$$

We consider the right-hand side as a function of $k\in[1,K]$, where

$$K := \frac{4l}{n}\,\min\left\{\frac{tL}{n},\,1\right\}$$

(cf. (5)) and we denote this function by $f(k)$. Now (8) follows from the fact that f is convex,

$$f(1)=\frac{4l}{n}\,\min\left\{\frac{tL}{n},1\right\}+2\le\min\left\{\frac{t}{12}+2,\,\frac{M}{n}+2\right\}\le\min\left\{\frac{t}{4},\,3\,\frac{M}{n}\right\}$$

(notice that $t>12$ by (6)), and

$$f(K)=\frac{12l}{n}\,\min\left\{\frac{tL}{n},1\right\}\le\min\left\{\frac{t}{4},3\,\frac{M}{n}\right\}.$$

Finally, to prove (9) we consider two cases: $k=1$ and $k\ge 2$.

If $k=1$ then by (3) and (1),

$$\begin{aligned}\frac{M-k\varrho}{l'}+k+d &< \frac{M-\varrho+2l}{l'}+1=\frac{M-(2N_t-2-l')+2l}{l'}+1\\ &\le\frac{3M/2-2(N_t-1)}{l'}+2<\frac{3M/2}{N_t-2}\\ &<3\,\frac{M}{n}\,\min\left\{\frac{tL}{n},1\right\}\le\min\left\{\frac{t}{4},\,3\,\frac{M}{n}\right\}.\end{aligned}$$

If $k\ge 2$ then by (3) and (5),

$$\begin{aligned}\frac{M-k\varrho}{l'}+k+d &< \frac{M+2l}{l'}+k<\frac{M+2l}{n}\,\min\left\{\frac{tL}{n},\,1\right\}+k\\ &<\frac{M+6l}{n}\,\min\left\{\frac{tL}{n},\,1\right\}<3\,\frac{M}{n}\,\min\left\{\frac{tL}{n},\,1\right\}\\ &\le\min\left\{\frac{t}{4},\,3\,\frac{M}{n}\right\}.\ \blacksquare\end{aligned}$$

5. Proof of Theorems 2 and 2′. We only sketch the proof of Theorem 2, which essentially follows that of Theorem B, and indicate how it is to be modified to obtain Theorem 2′.

We need a lemma which is a minor variation of [S94, Lemma 3].

LEMMA 3. *Let U, V, and d be positive integers and let $B \subseteq [U, V]$ be a set of integers. Then for any $s \in B^*$ such that $d \mid s$, $s \le U(|B| - d)$ there exists $s' \in B^*$ satisfying*

$$d \mid s', \quad 0 < s' - s \le Vd.$$

Proof. Choose $S \subseteq B$ such that $s = \sum_{b \in S} b$. As

$$U|S| \le s \le U(|B| - d),$$

we have $|B \setminus S| \ge d$, and therefore there is a set $S_0 \subseteq B \setminus S$ of cardinality $0 < |S_0| \le d$ such that $\sum_{b \in S_0} b \equiv 0 \pmod d$. We define $S' := S \cup S_0$ and $s' := \sum_{b \in S'} b$; then

$$s' - s = \sum_{b \in S_0} b \le V|S_0| \le Vd$$

and the validity of $d \mid s'$ is obvious from the construction. ∎

To prove Theorem 2 we write $A = \{a_1, \dots, a_n\}$ (the elements being numbered in increasing order) and set

$$j := \lfloor 0.7n \rfloor, \quad A_1 := \{a_1, \dots, a_j\}, \quad A_2 := \{a_{j+1}, , \dots, a_n\}.$$

(For Theorem 2′, set $j := \lfloor 0.99n \rfloor$.) We apply Theorem 3 to the set $A_1 \subseteq [1, a_j]$ of cardinality $j = |A_1|$ and with $M = 4l$. (Verification of the condition $M \le |A_1|^2/(12L)$, or equivalently $j^2 \ge 48l \ln(4a_j/j)$, is left to the reader; hint: $4a_j/j < 4l/j \le n$, and $j^2 > 0.49n^2 - 1.4n$.) We conclude that there exist integers $d, h \ge 1$ and y such that

$$\{yd, (y+1)d, \dots, (y+4l)d\} \subseteq h \cdot A_1 \tag{10}$$

and

$$d < 4\,\frac{a_j}{j}, \quad h < 24\,\frac{l}{j}; \tag{11}$$

we observe, moreover, that (11) implies

$$d < 4\,\frac{l}{j} < 6\,\frac{l}{n}$$

and

$$yd \le ha_j < 24\,\frac{la_j}{j}. \tag{12}$$

Next, applying Lemma 3 to the set $B = A_2 \subseteq [a_j, l]$ we see that for any $s \in A_2^*$ such that $d \mid s$ and $s \le a_j(n - j - d)$ there exists $s' \in A_2^*$ such that

$d \mid s'$ and $s' - s \le ld$. Along with (10) this shows that

$$\{yd, (y+1)d, \dots, (\lfloor a_j(n-j-d)/d \rfloor + 1)d + (y+4l)d\} \subseteq A_1^* + A_2^* = A^*.$$

We put

$$z := \left\lfloor \frac{a_j(n-j-d)}{d} \right\rfloor + 1 + (y+4l),$$

so that by (11),

$$\begin{aligned} z - y &> \frac{a_j(n-j)}{d} + (4l - a_j) > \frac{a_j(n-j)}{d} \\ &> \frac{1}{4}\, j(n-j) \ge \frac{0.21}{4}\, n^2 > \frac{1}{20}\, n^2. \end{aligned}$$

It remains to notice that by (12),

$$\frac{z-y}{y} > \frac{a_j(n-j)}{yd} > \frac{j(n-j)}{24l} \ge \frac{0.21}{24} \cdot \frac{n^2}{l} > \frac{n^2}{115l}.$$

REMARK. It is clear that the above argument actually gives slightly better constants than those indicated in the formulation of Theorems 2 and 2′. In particular, the reader can easily verify that our proof of Theorem 2 remains valid if the assumption $n \ge 10(l \ln n)^{1/2}$ is relaxed to $n \ge 9.95(l \ln n)^{1/2}$. This observation will be used in Section 7 to avoid further loss of accuracy.

6. Auxiliary results on addition in $\mathbb{Z}_q$. We say (modifying slightly the terminology of [C90, C99]) that a set of integers A is *q-complete* if $A^* \pmod q = \mathbb{Z}_q$; in other words, if all residues modulo q are represented in A^*. In the spirit of the two papers just mentioned, we establish here a sufficient condition for q-completeness.

LEMMA 4. *Let A be a finite set of integers and let $q \ge 1$ be an integer. Suppose that $N_d(A) \ge d - 1$ for any $d \mid q$. Then A is q-complete.*

Proof. We can assume that A is finite, and we use induction by $n = |A|$. Write $A = \{a_1, \dots, a_n\}$ and $A_i = \{a_1, \dots, a_i\}$ $(1 \le i \le n)$ and let overlined characters denote canonical images in $\mathbb{Z}_q$. Suppose that a_i are so numbered that $\overline{A}_1^* \subsetneq \dots \subsetneq \overline{A}_j^*$ and that either $A_j = A$, or $(\overline{A}_j \cup \{\overline{a}\})^* = \overline{A}_j^*$ for any $a \in A \setminus A_j$. In the former case we have $j = n$ and accordingly

$$|\overline{A}^*| = |\overline{A}_j^*| \ge j + 1 = |A| + 1 \ge N_q(A) + 1 \ge q,$$

which proves the assertion. We assume, therefore, that the second possibility holds, which translates easily into $\overline{A}_j^* + \overline{a} = \overline{A}_j^*$ (for any $a \in A \setminus A_j$).

Consider the subgroup of all $g \in \mathbb{Z}_q$ satisfying $\overline{A}_j^* + g = \overline{A}_j^*$ and write this subgroup as $d\mathbb{Z}_q$, where d is a divisor of q. (This subgroup is often called the *period* or *stabilizer* of $\overline{A}_j^*$.) The condition $\overline{A}_j^* + \overline{a} = \overline{A}_j^*$ means that $\overline{a} \in d\mathbb{Z}_q$, that is, $d \mid a$, and it follows that $\delta \mid a$ for any divisor $\delta \mid d$. This shows that

$N_\delta(A_j) = N_\delta(A) \ge \delta - 1$, provided $\delta \mid d$. By the induction hypothesis, A_j is d-complete; that is, $\overline{A}_j^*$ contains representatives of all $d\mathbb{Z}_q$-cosets. Now $\overline{A}_j^* = \overline{A}_j^* + d\mathbb{Z}_q = \mathbb{Z}_q$, implying the result. ∎

Corollary 3. *Suppose that $A \subseteq \mathbb{Z}$ satisfies $N_q(A) \ge q - 1$ for any $q \le Q$. Then A is q-complete for any $q \le Q$.*

Lemma 5 (Chaimovich, [C99, Lemma 2.3]). *If A is q-complete then there is a subset $A_0 \subseteq A$ of cardinality $|A_0| \le q - 1$ which is also q-complete.*

Proof. As in the proof of the previous lemma, we use overlined characters to denote canonical images in $\mathbb{Z}_q$. Let $A_0 = \{a_1, \dots, a_k\} \subseteq A$ be a subset of minimum cardinality such that $\overline{A}_0^* = \mathbb{Z}_q$, and let $A_j = \{a_1, \dots, a_j\}$ $(j = 1, \dots, k)$. If $\overline{A}_1^* \subsetneq \dots \subsetneq \overline{A}_k^*$ then $q = |\overline{A}_k^*| \ge k + 1$, as required. Otherwise, there is an index $j \le k - 1$ such that $\overline{A}_j^* = \overline{A}_{j+1}^*$. But in this case we would have

$$\begin{aligned}\overline{A}_0^* &= \overline{A}_{j+1}^* + \{\overline{a}_{j+2}, \dots, \overline{a}_k\}^* = \overline{A}_j^* + \{\overline{a}_{j+2}, \dots, \overline{a}_k\}^* \\ &= \{\overline{a}_1, \dots, \overline{a}_j, \overline{a}_{j+2}, \dots, \overline{a}_k\}^*\end{aligned}$$

contradicting minimality of A_0. ∎

7. Proof of Theorem 1

Lemma 6. *For any $C, \varepsilon > 0$ and $K \ge 2$ there exists $n_0 = n_0(C, \varepsilon, K)$ with the following property. Let $Q \ge 2$ and let $A \subseteq [1, l]$ be a set of integers of cardinality $n = |A| \ge C(l \ln l)^{1/2}$, $n \ge n_0$, satisfying $N_q(A) \ge Kq$ for all $q \in [2, Q]$. Then there is a subset $A_0 \subseteq A$ such that*

(i) $K^{-1}n < |A_0| < (1 + \varepsilon)K^{-1}n$;
(ii) $\sigma(A_0) < (1 + \varepsilon)K^{-1}\sigma(A)$;
(iii) $N_q(A_0) \ge q$ *for all* $q \in [2, Q]$.

Proof. Fix arbitrarily $p \in (K^{-1}, (1 + \varepsilon)K^{-1})$, $p < 1$ and set

$$\delta := \frac{1}{2} \min\{1 - (pK)^{-1}, (1 + \varepsilon)(pK)^{-1} - 1\}.$$

Define $P(q) := e^{-\delta^2 pKq/2}$. As the series $\sum_{q=2}^{\infty} P(q)$ converges, there is an integer $Q_0 = Q_0(p, K)$ such that $\sum_{q=Q_0+1}^{\infty} P(q) < 0.1$.

The set A_0 will be comprised of two parts. First, for each integer $q \in [2, Q_0]$ we choose q arbitrary elements of A, not divisible by q, and we denote by A_1 the set of all elements chosen. Plainly, $|A_1| < \frac{1}{2} Q_0(Q_0 + 1)$. Second, we let A_2 be the random subset of A to which any $a \in A$ belongs with probability p (with all $a \in A$ being selected independently). Thus $|A_2|$ is distributed binomially with parameters n and p, and $\mathrm{E}|A_2| = pn$. Eventually, we define $A_0 := A_1 \cup A_2$.

Using an estimate for the right tail of binomial distribution (see, for instance, [JŁR, Theorems 2.1 and 2.10]) we get

$$\Pr\{|A_2| \le (1-\delta)pn\} = \Pr\{|A_2| \le (1-\delta)\mathrm{E}|A_2|\} \le e^{-\delta^2 \mathrm{E}|A_2|/2} = e^{-\delta^2 pn/2}.$$

As $(1-\delta)p > K^{-1}$, for n large enough this implies

$$\Pr\{|A_0| \le K^{-1}n\} \le \Pr\{|A_2| \le (1-\delta)pn\} < 0.1.$$

Similarly, in view of $|A_1| = O_{p,K}(1)$ we have

$$\Pr\{|A_0| \ge (1+\varepsilon)K^{-1}n\} \le \Pr\{|A_2| \ge (1+\delta)pn\} < 0.1;$$

therefore, (i) fails with probability at most 0.2.

Furthermore, $\sigma(A_2) := \sum_{a\in A_2} a$ is a random variable with expectation $\mathrm{E}\sigma(A_2) = p\sigma(A)$ and variance

$$\operatorname{Var}\sigma(A_2) = p(1-p)\sum_{a\in A} a^2 < pl\sigma(A).$$

By Chebyshev's inequality we have

$$\begin{aligned}\Pr\{\sigma(A_2) \ge (1+\delta)p\sigma(A)\} &= \Pr\{\sigma(A_2) \ge (1+\delta)\mathrm{E}\sigma(A_2)\}\\ &< \frac{pl\sigma(A)}{(\delta\mathrm{E}\sigma(A_2))^2} = (\delta^2 p)^{-1}\frac{l}{\sigma(A)} < 2(\delta^2 p)^{-1}\frac{l}{n^2} < 0.1,\end{aligned}$$

provided that n is large enough. As $(1+\delta)p < (1+\varepsilon)K^{-1}$ and in view of $\sigma(A_1) = O(l) = o(\sigma(A))$ this implies that

$$\Pr\{\sigma(A_0) \ge (1+\varepsilon)K^{-1}\sigma(A)\} < 0.1;$$

that is, (ii) fails with probability at most 0.1.

Finally, we take care of (iii). For any integer $q \in [Q_0+1, Q]$ the random variable $N_q(A_2)$ is distributed binomially with parameters $N_q(A)$ and p, whence in view of $q \le K^{-1}N_q(A) = (Kp)^{-1}\mathrm{E}N_q(A_2) < (1-\delta)\mathrm{E}N_q(A_2)$ we get

$$\begin{aligned}\Pr\{N_q(A_2) < q\} \le\ &\Pr\{N_q(A_2) < (1-\delta)\mathrm{E}N_q(A_2)\}\\ &< e^{-\delta^2\mathrm{E}N_q(A_2)/2} \le e^{-\delta^2 pKq/2} = P(q)\end{aligned}$$

(estimating the left binomial tail). It follows that the probability that there is a value of $q \in [Q_0+1, Q]$ such that $N_q(A_2) < q$, and therefore the probability that there is a value of $q \in [2, Q]$ such that $N_q(A_0) < q$, is at most $\sum_{q=Q_0+1}^{\infty} P(q) < 0.1$ (by the choice of Q_0). Therefore, (iii) fails with probability 0.1 at most.

We see that the probability that any of conditions (i)–(iii) fails to hold is $0.2+0.1+0.1 < 1$ at most. Hence, with a positive probability A_0 satisfies all the properties required. ∎

PROPOSITION 1. *There exist positive constants K and n_0 with the following property. Let $A \subseteq [1,l]$ be a set of $n = |A| \ge n_0$ integers such that*

$$n \ge 9.96(l \ln n)^{1/2}$$

and $N_q(A) \ge Kq$ for any integer $q \in [2, 7l/n]$. Then there are integers y and z such that $[y,z] \subseteq A^*$ and

$$z - y > \frac{n^2}{21}, \qquad \frac{z-y}{y} > \frac{n^2}{138l}.$$

Proof. Choose positive K and ε so that the former is sufficiently large and the latter is sufficiently small. By Lemma 6 and Corollary 3 there is a subset $A_0 \subseteq A$ of cardinality $|A_0| < (1+\varepsilon)K^{-1}n < 0.001n$ which is q-complete for any $q \le 7l/n$. Write $A_1 := A \setminus A_0$, so that $n_1 := |A_1| > 0.999n$. We have

$$n_1 > 0.999 \cdot 9.96(l \ln n)^{1/2} > 9.95(l \ln n_1)^{1/2}$$

and by Theorem 2 as applied to the set A_1 (see also the Remark at the end of Section 5), there exist integers y_1, z_1, and $d \ge 1$ such that $\{y_1 d, (y_1+1)d, \dots, z_1 d\} \subseteq A_1^*$ and

$$d < 6\,\frac{l}{n_1} < 7\,\frac{l}{n}, \qquad z_1 - y_1 > \frac{n_1^2}{20} > \frac{n^2}{21}, \qquad \frac{z_1 - y_1}{y_1} > \frac{n_1^2}{115l} > \frac{n^2}{116l}.$$

If $d = 1$ the assertion follows. Assuming that $d \ge 2$, find $B_0 \subseteq A_0$ which is d-complete and such that $|B_0| \le d-1$: this is possible by Lemma 5. Notice that to any integer $r \in [0, d-1]$ there corresponds $k_r \in [0, l-1]$ such that $r + k_r d \in B_0^*$. Now set $y := (y_1 + l)d$ and $z := z_1 d$; then

$$z - y = (z_1 - y_1 - l)d > z_1 - y_1 > n^2/21,$$

as $z_1 - y_1 > n^2/21 > 2l \ge ld/(d-1)$. Furthermore, let $\kappa = 116/21$. If $y_1 \ge \kappa l$ then

$$\frac{z-y}{y} = \frac{z_1}{y_1 + l} - 1 \ge \frac{\kappa}{\kappa+1} \cdot \frac{z_1}{y_1} - 1 > \frac{\kappa}{\kappa+1} \cdot \frac{n^2}{116l} - 1 > \frac{n^2}{138l},$$

and if $y_1 \le \kappa l$ then

$$\frac{z-y}{y} = \frac{z_1}{y_1 + l} - 1 \ge \frac{1}{\kappa+1} \cdot \frac{z_1}{l} - 1 > \frac{1}{\kappa+1} \cdot \frac{n^2}{21l} - 1 > \frac{n^2}{138l}.$$

To complete the proof we show that $[y,z] \subseteq B_0^* + A_1^*$. Indeed, given $x \in [y,z]$ write $x = r + kd$ where $r \in [0, d-1]$ and $y_1 + l \le k \le z_1$. Then $x = (r + k_r d) + (k - k_r)d$ and it suffices to notice that $y_1 \le k - l < k - k_r \le z_1$. ∎

As a next step, we show that the condition $N_q(A) \ge Kq$ of Proposition 1 can be relaxed considerably.

PROPOSITION 2. *Let $A \subseteq [1,l]$ be a set of $n = |A| \geq n_0$ integers, where n_0 is a sufficiently large absolute constant. Suppose that*

$$n \geq 9.96(l \ln n)^{1/2}$$

and that $N_q(A) \geq q-1$ for any integer $q < 2l/n$. Then there are integers y and z such that $[y,z] \subseteq A^*$ and

$$z - y > \frac{n^2}{21}, \qquad \frac{z-y}{y} > \frac{n^2}{138l}.$$

Proof. Let K be the constant of Proposition 1. If $N_q(A) \geq Kq$ for all integer $q \in [2, 7l/n]$ then the result follows from Proposition 1. Otherwise, denote by q the maximal number such that $2 \leq q \leq 7l/n$ and $N_q(A) < Kq$. Let A_0 be the set of all elements of A, not divisible by q, and let A_1 be the set of integers such that $A \setminus A_0 = \{qa : a \in A_1\}$. Evidently, we have $A_1 \subseteq [1, l_1]$ where $l_1 = \lfloor l/q \rfloor \leq l/2$. Furthermore, the cardinality of A_1 satisfies $n_1 := |A_1| > n - Kq$, and if n is large enough then $q < 0.1K^{-1}n$ and $n_1 > 0.9n$.

We want to apply Proposition 1 to the set A_1. We have

$$n_1 > 0.9n \geq 0.9 \cdot 0.96(l \ln n)^{1/2} > 0.96((l/2) \ln n_1)^{1/2} \geq 0.96(l_1 \ln n_1)^{1/2}$$

and for Proposition 1 to be applicable it is necessary and sufficient that $N_{q_1}(A_1) \geq Kq_1$ for any $q_1 \in [2, 7l_1/n_1]$. Assume that this does *not* hold for some q_1. Then $q_1 \leq 7l/(0.9n) \leq 0.1K^{-1}n$ (again, assuming that n is sufficiently large), hence $K(q+q_1) < 0.2n$ and therefore

$$\frac{l}{q} \geq l_1 > (n_1 - Kq_1)q_1 > (n - K(q+q_1))q_1 > 0.8nq_1, \qquad qq_1 < 2\,\frac{l}{n}.$$

Along with $N_{qq_1}(A) < Kq + Kq_1 \leq Kqq_1$ this contradicts maximality of q.

We see that the conditions of Proposition 1 are fulfilled and there exist integers y_1 and z_1 such that

$$z_1 - y_1 > \frac{n_1^2}{21} > \frac{n^2}{26}, \qquad \frac{z_1 - y_1}{y_1} > \frac{n_1^2}{138l_1} > \frac{n^2}{86l}$$

and $[y_1, z_1] \subseteq A_1^*$. Notice that the latter inclusion yields $\{y_1q, (y_1+1)q, \ldots, z_1q\} \subseteq (A \setminus A_0)^*$.

We have assumed that $q \leq 7l/n$ and we now observe that in fact, $q < 2l/n$ holds, for $\lfloor l/q \rfloor = l_1 \geq n_1 > 0.9n$. It follows that A, and therefore A_0 also, is q-complete and, as in the proof of Proposition 1, we conclude that $[y,z] \subseteq A^*$ with $y = (y_1 + l)q$ and $z = z_1q$. Finally, we have

$$z - y \geq 2(z_1 - y_1 - l) \geq 2(n^2/26 - l) > n^2/21$$

and to estimate $(z-y)/y = z_1/(y_1+l) - 1$ one can define $\kappa = 86/26$ and act as in the proof of Proposition 1. ∎

PROPOSITION 3. *Let $A \subseteq [1,l]$ be a set of $n = |A| \ge n_0$ integers, where n_0 is a sufficiently large absolute constant. Suppose that*

$$n \ge 20(l \ln n)^{1/2}$$

and that $N_q(A) \ge 3q$ for any integer $q \in [2, 5l/n]$. Then

$$[\lambda\sigma(A), (1-\lambda)\sigma(A)] \subseteq A^*.$$

Proof. Set $K = 2.008$, fix ε positive and small enough and find a subset $A_0 \subseteq A$ satisfying conditions of Lemma 6. Write $A_1 := A \setminus A_0$. We have

$$n_0 := |A_0| > K^{-1}n \ge 9.96(l \ln n)^{1/2} \ge 9.96(l \ln n_0)^{1/2}$$

and $N_q(A_0) \ge q-1$ for any $q \le 5l/n$, hence for any $q < 2l/n_0$. Proposition 2 shows, therefore, that there is an interval $[y,z] \subseteq A_0^*$ such that $z - y > n_0^2/21 > n^2/(21K^2) > l$ and $(z-y)/y > n_0^2/(138l)$. Since the gaps between consecutive elements of A_1^* do not exceed l, it follows that $[y, \sigma(A_1) + z] \subseteq A_0^* + A_1^* = A^*$. We now observe that

$$y < 138\,\frac{l}{n_0^2}\,z < 138K^2\,\frac{l}{n^2}\,\sigma(A_0) < 138(1+\varepsilon)K\,\frac{l}{n^2}\,\sigma(A) < \lambda\sigma(A)$$

and that

$$\sigma(A_1) + z \ge \sigma(A) - \sigma(A_0) > (1 - (1+\varepsilon)K^{-1})\sigma(A) > \sigma(A)/2.$$

Hence, $[\lambda\sigma(A), \sigma(A)/2] \subseteq A^*$ and it remains to notice that A^* is symmetric about $\sigma(A)/2$: if $s \in A^*$ then also $\sigma(A) - s \in A^*$ for any integer s. ∎

Finally, we are ready to prove Theorem 1. For this, we "improve" Proposition 3 in the same spirit as we did with Proposition 1.

Proof of Theorem 1. The beginning of the proof runs as that of Proposition 2. If for any $q \in [2, 5l/n]$ we have $N_q(A) \ge 3q$, then the result follows at once by Proposition 3. Otherwise, let $q \in [2, 5l/n]$ be the maximal integer satisfying $N_q(A) < 3q$. Define A_0 to be the set of all elements of A not divisible by q (thus $|A_0| = N_q(A) < 3q$), and let A_1 be the set of integers such that $A \setminus A_0 = \{qa : a \in A_1\}$. Evidently, we have $A_1 \subseteq [1, l_1]$, where $l_1 = \lfloor l/q \rfloor$, and $n_1 := |A_1| > n - 3q \ge 0.9n$ for n large enough, as then $q \le 5l/n \le 0.03n$. Therefore,

$$n_1 > 0.9n \ge 18(l \ln n)^{1/2} > 20((l/2)\ln n_1)^{1/2} \ge 20(l_1 \ln n_1)^{1/2}$$

and also $N_{q_1}(A_1) \ge 3q_1$ for $q_1 \in [2, 5l_1/n_1]$, as in the proof of Proposition 2. By Proposition 3 we have $[y,z] \in A_1^*$ with y and z integer and satisfying $y \le \lceil \lambda_1\sigma(A_1) \rceil$, $z \ge \lfloor (1-\lambda_1)\sigma(A_1) \rfloor$, where $\lambda_1 = 280l_1/n_1^2$. Furthermore, $q \le l/l_1 \le l/n_1 < 2l/n$, whence A_0 is q-complete. It follows that $[(y+l)q, zq] \in A^*$. Next, we have

$$lq = 2\,\frac{l}{n_1^2} \cdot \frac{n_1^2}{2}\,q < 3\,\frac{l}{n^2}\,\sigma(A_1)q \le 3\,\frac{l}{n^2}\,\sigma(A), \tag{13}$$

implying

$$(y+l)q \le (\lambda_1\sigma(A_1)+l+1)q \le 280\,\frac{l}{qn_1^2}\,\sigma(A)+2lq$$
$$\le 200\,\frac{l}{n^2}\,\sigma(A)+6\,\frac{l}{n^2}\,\sigma(A) \le \lambda\sigma(A)$$

and

$$zq \ge ((1-\lambda_1)\sigma(A_1)-1)q \ge (1-\lambda_1)\sigma(A)-\sigma(A_0)-q$$
$$\ge (1-\lambda_1)\sigma(A)-4lq \ge (1-280l_1/n_1^2-12l/n^2)\sigma(A)$$
$$\ge (1-212l/n^2)\sigma(A) > (1-\lambda)\sigma(A). \blacksquare$$

8. Proof of Corollary 1. We can assume that $q < 2l/n$ with the property that $N_q(A) < q-1$ do exist, and let d denote the maximal such q. Furthermore, let A_0 be the set of all elements of A not divisible by d (so that $|A_0| = N_d(A) < d-1$), and let A_1 be the set of integers such that $A \setminus A_0 = \{ad : a \in A_1\}$; thus, $n_1 := |A_1| = |A_d| > n-d$. We have $A_1 \subseteq [1, l_1]$, where $l_1 = \lfloor l/d \rfloor$, and we set $\lambda_1 := 280l_1/n_1^2$. If n is large enough, then $d < 2l/n < 0.1n$ and

$$n_1 > 0.9n \ge 18(l\ln n)^{1/2} > 20(l_1\ln n_1)^{1/2}.$$

As in Proposition 2, maximality of d implies that $N_q(A_1) \ge q-1$ for any $q < 2l_1/n_1$; hence $[\lambda_1\sigma(A_1), (1-\lambda_1)\sigma(A_1)] \subseteq A_1^*$ by Theorem 1, and A_d^* contains all multiples of d from the interval $[\lambda_1 d\sigma(A_1), (1-\lambda_1)d\sigma(A_1)]$. It remains to observe that

$$\lambda_1 d\sigma(A_1) \le 280\,\frac{l}{dn_1^2}\,\sigma(A) \le 200\,\frac{l}{n^2}\,\sigma(A)$$

and that, in view of $ld < (3l/n^2)\sigma(A)$ (cf. (13)),

$$(1-\lambda_1)d\sigma(A_1) \ge (1-\lambda_1)\sigma(A)-\sigma(A_0)$$
$$> (1-200l/n^2)\sigma(A)-ld > (1-203l/n^2)\sigma(A).$$

Appendix. Infinite progressions in subset sums set. Let $\mathcal{A}$ be a (strictly increasing, infinite) sequence of positive integers. Write $\mathcal{A}(n) = |\mathcal{A} \cap [1,n]|$, the *counting function* of $\mathcal{A}$. A beautiful result of Folkman [Fo66, Theorem 1.3] is that $\mathcal{A}^*$ contains an infinite arithmetic progression, provided that $\mathcal{A}(n) > n^{1/2+\varepsilon}$ for some fixed $\varepsilon > 0$ and all n large enough. In a striking development, Łuczak and Schoen showed recently that the progression can be chosen to be homogeneous and even under a somewhat relaxed assumption on $\mathcal{A}(n)$; their result is, in fact, the infinite version of Theorems A and B.

THEOREM 5 (Łuczak and Schoen, [ŁS00, Theorem 2]). *Let $\mathcal{A}$ be a sequence of positive integers satisfying $\mathcal{A}(n) \ge 402\sqrt{n\ln n}$ for all n large enough. Then there exists an integer d such that $\{d, 2d, 3d, \ldots\} \subseteq \mathcal{A}^*$.*

In this context we also mention a paper of Hegyvári [H00] where under a similar restriction on $\mathcal{A}(n)$ it is shown that $\mathcal{A}^*$ contains an infinite arithmetic progression. However, the progression is not guaranteed to be homogeneous.

An almost immediate consequence of Theorem 5, indicated in [ŁS00], is a (nearly sharp) estimate of the maximum possible density of a strongly sum-free sequence of integers.

THEOREM 6 (Łuczak and Schoen, [ŁS00, Theorem 3]). *Let $\mathcal{A}$ be a sequence of positive integers such that the equation $a_1 + \ldots + a_k = a_0$ has no solutions in $a_i \in \mathcal{A}$ (for any $k \geq 2$). Then $\mathcal{A}(n) \leq 403\sqrt{n \ln n}$ for all n large enough.*

Using Theorem 2′ one can reduce the constants as follows.

THEOREM 5′. *Let $\mathcal{A}$ be a sequence of positive integers satisfying $\mathcal{A}(n) \geq 5\sqrt{n \ln n}$ for all n large enough. Then there exists an integer d such that $\{d, 2d, 3d, \ldots\} \subseteq \mathcal{A}^*$.*

THEOREM 6′. *Let $\mathcal{A}$ be a sequence of positive integers such that the equation $a_1 + \ldots + a_k = a_0$ has no solutions in $a_i \in \mathcal{A}$ (for any $k \geq 2$). Then $\mathcal{A}(n) < 5\sqrt{n \ln n}$ for all n large enough.*

Theorem 6′ can be deduced from Theorem 5′ exactly in the same way as Theorem 6 is deduced from Theorem 5, and we refer the reader to [ŁS00] for this deduction ([1]). On the other hand, modifications to be made to the original proof of Theorem 5 (to upgrade it to Theorem 5′) are minor but numerous, and for this reason we provide a sketch of the modified proof.

Sketch of the proof of Theorem 5′. Write $\mathcal{A} = \{a_1, a_2, \ldots\}$, set $\varphi(i) := \lfloor i \ln^{1/4}(i+1) \rfloor$ and consider the decomposition $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$ defined by

$$\mathcal{A}_1 := \{a_{\varphi(1)}, a_{\varphi(2)}, \ldots\}, \quad \mathcal{A}_0 := \mathcal{A} \setminus \mathcal{A}_1.$$

Evidently,

$$\begin{aligned}\mathcal{A}_1(n) &= \#\{i : \varphi(i) \leq \mathcal{A}(n)\} = \#\{i : i \ln^{1/4}(i+1) \leq \mathcal{A}(n)\} \\ &= \frac{\mathcal{A}(n)}{\ln^{1/4} \mathcal{A}(n)} \, (1 + o(1))\end{aligned}$$

whence for n large enough we have $\mathcal{A}_1(n) \gg \sqrt{n} \, \ln^{1/4} n$ and $\mathcal{A}_0(n) = \mathcal{A}(n) - \mathcal{A}_1(n) \geq 4.99\sqrt{n \ln n}$.

From [Fo66, Lemma 2.2] it follows that the condition

$$\limsup_{n \to \infty} \mathcal{A}_1(n)/\sqrt{n} = \infty$$

([1]) We have simplified the logic for the sake of clarity. In fact, to deduce Theorem 6′ one needs a version of Theorem 5′ with a constant strictly less than 5. The reader will see, however, that the argument we present below allows one to obtain any constant greater than $7/\sqrt{2} = 4.949\ldots$

guarantees that for any fixed integer $d \geq 1$ the subset sum set $\mathcal{A}_1^*$ has bounded gaps on the sequence of all multiples of d. That is, to any d there corresponds a constant C_d such that $\{(y+1)d, (y+2)d, \ldots, (y+C_d)d\} \cap \mathcal{A}_1^* \neq \emptyset$ for any integer y. On the other hand, we show that the condition $\mathcal{A}_0(n) \geq 4.99\sqrt{n \ln n}$ ensures the existence of d such that $\mathcal{A}_0^*$ contains a homogeneous progression with difference d of any pre-assigned length. This will imply that $\mathcal{A}^* = \mathcal{A}_0^* + \mathcal{A}_1^*$ contains an infinite progression of the form $\{(m+1)d, (m+2)d, \ldots\}$ and therefore the infinite progression $\{d', 2d', \ldots\}$ where $d' = (m+1)d$.

We need a lemma which is, essentially, [ŁS00, Fact 5].

LEMMA 7. *Let P_1 and P_2 be homogeneous arithmetic progressions with differences d_1 and d_2 and at least m_1 and m_2 terms, respectively. Suppose that $m_2 \geq d_1$, $m_1 \geq d_2$, and $d_1 \leq d_2$. Then $P_1 + P_2$ contains a homogeneous arithmetic progression of difference d_1 with at least $m_1 + m_2 - 2d_2$ terms.*

For the proof observe that P_2 contains a homogeneous progression P_2' of difference $d_1 d_2$ with $\lfloor m_2/d_1 \rfloor \geq 1$ terms, and the sum of P_1 and P_2' is a homogeneous progression of difference d_1 with at least $(\lfloor m_2/d_1 \rfloor - 1)d_2 + m_1 \geq m_1 + m_2 - 2d_2$ terms. (We use here the observation that the difference between consecutive terms of P_2' does not exceed the "length" of P_1.)

Back to the proof of Theorem 5′, for $i = 0, 1, \ldots$ we define $l_i := 2^{2^i}$ (so that $l_{i+1} = l_i^2$) and we set $A_i := \mathcal{A}_0 \cap (l_{i-1}, l_i]$, $n_i := |A_i|$. For i large enough we then have $A_i \subseteq [1, l_i]$ and

$$n_i = \mathcal{A}_0(l_i) - \mathcal{A}_0(l_{i-1}) \geq 4.99\sqrt{l_i \ln l_i} - \sqrt{l_i} \geq 4.98\sqrt{l_i \ln l_i},$$
$$\frac{n_i^2}{49 \ln n_i} \geq \frac{4.98^2\, l_i \ln l_i}{49\,(\ln l_i)/2}\,(1 + o(1)) \geq 1.01(1 + o(1))\, l_i > l_i,$$
$$n_i > 7(l_i \ln n_i)^{1/2}.$$

Thus Theorem 2′ is applicable and A_i^* contains a homogeneous progression of difference $d_i < 5l_i/n_i < 2\sqrt{l_i/\ln l_i}$ and with at least $n_i^2/405 > (l_i \ln l_i)/20 > 4l_i$ terms.

Fix an integer j such that the above conclusions hold true and moreover, $d_i \geq d_j$ for $i \geq j$. Using induction and Lemma 7 it is easy to verify that for any $i \geq j$ the set $A_j^* + \ldots + A_i^*$ contains a homogeneous progression of difference d_j with at least $4l_i$ terms: put $P_1 = A_j^* + \ldots + A_{i-1}^*$, $P_2 = A_i^*$ and observe that

$$d_i < 2\,\frac{l_{i-1}}{\sqrt{2^i \ln 2}} < 2l_{i-1}, \qquad d_j < 2\sqrt{l_j/(2 \ln 2)} < 2l_i,$$
$$4l_{i-1} + 4l_i - 2d_i = 4l_i + 2(2l_{i-1} - d_i) > 4l_i.$$

Therefore $\mathcal{A}_0^*$ contains arbitrarily long homogeneous progressions of difference $d = d_j$, as required. ∎

References

[AF88] N. Alon and G. A. Freiman, *On sums of subsets of a set of integers*, Combinatorica 8 (1988), 297–306.

[C90] M. Chaimovich, *Subset-sums problems with different summands*: *computation*, Discrete Math. Appl. 27 (1990), 277–282.

[C99] —, *New algorithm for dense subset-sum problem*, Astérisque 258 (1999), 363–373.

[EF90] P. Erdős and G. Freiman, *On two additive problems*, J. Number Theory 34 (1990), 1–12.

[Fo66] J. Folkman, *On the representation of integers as sums of distinct terms from a fixed sequence*, Canad. J. Math. 18 (1966), 643–655.

[Fr93] G. A. Freiman, *New analytical results in subset-sum problem*, Discrete Math. 114 (1993), 205–217.

[GM91] Z. Galil and O. Margalit, *An almost linear-time algorithm for the dense subset sum problem*, SIAM J. Comput. 20 (1991), 1157–1189.

[H00] N. Hegyvári, *On the representation of integers as sums of distinct terms of a fixed set*, Acta Arith. 92 (2000), 99–104.

[JŁR] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[Le97] V. Lev, *Optimal representations by sumsets and subset sums*, J. Number Theory 62 (1997), 127–143.

[Le98] —, *On consecutive subset sums*, Discrete Math. 187 (1998), 151–160.

[Li89] E. Lipkin, *On representation of r-th powers by subset sums*, Acta Arith. 52 (1989), 353–366.

[ŁS00] T. Łuczak and T. Schoen, *On the maximal density of sum-free sets*, ibid. 95 (2000), 225–229.

[P74] J. M. Pollard, *A generalization of the theorem of Cauchy and Davenport*, J. London Math. Soc. (2) 8 (1974), 460–462.

[S89] A. Sárközy, *Finite addition theorems*, *I*, J. Number Theory 32 (1989), 114–130.

[S94] —, *Finite addition theorems*, *II*, ibid. 48 (1994), 197–218.

Department of Mathematics
Haifa University at Oranim
Tivon 36006, Israel
E-mail: seva@math.haifa.ac.il

Received on 21.5.2001
and in revised form on 31.3.2002 (4034)