CONSECUTIVE INTEGERS IN HIGH-MULTIPLICITY SUMSETS

VSEVOLOD F. LEV

Abstract. Sharpening (a particular case of) a result of Szemerédi and Vu [4] and extending earlier results of Sárközy [3] and ourselves [2], we find, subject to some technical restrictions, a sharp threshold for the number of integer sets needed for their sumset to contain a block of consecutive integers, whose length is comparable with the lengths of the set summands.

A corollary of our main result is as follows. Let $k, l \geqslant 1$ and $n \geqslant 3$ be integers, and suppose that $A_1, \dots, A_k \subseteq [0, l]$ are integer sets of size at least n, none of which is contained in an arithmetic progression with difference greater than 1. If $k \geqslant 2\lceil (l-1)/(n-2) \rceil$, then the sumset $A_1 + \dots + A_k$ contains a block of at least $k(n-1)+1$ consecutive integers.

1. Background and summary of results

The *sumset* of the subsets $A_1, \dots, A_k$ of an additively written group is defined by

$$A_1 + \dots + A_k := \{a_1 + \dots + a_k \colon a_1 \in A_1, \dots, a_k \in A_k\};$$

if $A_1 = \dots = A_k = A$, this is commonly abbreviated as kA. In the present paper we will be concerned exclusively with the group of integers, in which case a well-known phenomenon occurs: if all sets A_i are dense, and their number k is large, then the sumset $A_1 + \dots + A_k$ contains long arithmetic progressions. There are numerous ways to specialize this statement by indicating the exact meaning of "dense", "large", and "long", but in our present context the following result of Sárközy is the origin of things.

Theorem 1 (Sárközy [3, Theorem 1]). *Let $l \geqslant n \geqslant 2$ be integers and write $\kappa := \lceil (l+1)/(n-1) \rceil$. Then, for every integer set $A \subseteq [1, l]$ with $|A| = n$, there exist positive integers $d \leqslant \kappa - 1$ and $k < 118\kappa$ such that the sumset kA contains l consecutive multiples of d.*

2010 *Mathematics Subject Classification.* Primary: 11B13; secondary: 11B25, 11P70, 11B75.

Key words and phrases. Sumset, set addition, arithmetic progression.

The author gratefully acknowledges the support of the Georgia Institute of Technology and the Fields Institute, which he was visiting while conducting his research.

In [2] we established a sharp version of this result, replacing the factor 118 with 2 (which is the best possible value, as conjectured by Sárközy) and indeed, going somewhat further.

Theorem 2 (Lev [2, Theorem 1]). *Let $n \geqslant 3$ and $l \geqslant n-1$ be integers, and write $\kappa := \lfloor (l-1)/(n-2) \rfloor$. Then for every integer set $A \subseteq [0, l]$ with $0, l \in A$, $\gcd(A) = 1$, and $|A| = n$, and every integer $k \geqslant 2\kappa$, we have*

$$[\kappa(2l-2-(\kappa+1)(n-2)), kl-\kappa(2l-2-(\kappa+1)(n-2))] \subseteq kA.$$

The complicated-looking interval appearing in the statement of Theorem 2 is best possible and, in general, cannot be extended even by 1 in either direction. As the interested reader will easily check, this interval (strictly) includes $[\kappa l, (k-\kappa)l]$ as a subinterval; consequently, if $k \geqslant 2\kappa+1$, then its length exceeds l. At the same time, if $k \leqslant 2\kappa$, then the sumset kA may fail to contain a block of consecutive integers of length l; see the example after the statement of Theorem 4 below. Thus, $2\kappa+1$ is the smallest value of k such that, with A as in Theorem 2, the sumset kA is guaranteed to contain a block of consecutive integers of length l.

At first sight, Theorem 2 is weaker than Theorem 1 in imposing the extra assumptions $0, l \in A$ and $\gcd(A) = 1$. It is explained in [2], however, that these assumptions are merely of normalization nature, and a refinement of Theorem 1, with the bound 118κ replaced by $2\kappa+2$ (and some other improvements), is deduced from Theorem 2 in a relatively straightforward way.

We notice that Theorem 2 yields a sharp result about the function $f(n, k, l)$, introduced in [4]. This function is defined for positive integers k and $l \geqslant n \geqslant 2$ to be the largest number f such that for every n-element integer set $A \subseteq [1, l]$, the sumset kA contains an arithmetic progression of length f. (The *length* of an arithmetic progression is the number of its terms, less 1). As indicated in [4], "many estimates for $f(n, k, l)$ have been discovered by Bourgain, Freiman, Halberstam, Green, Ruzsa, and Sárközy". It is worth noting in this connection that Theorem 2 establishes the *exact value* of this function for k large; namely, it is easy to deduce from Theorem 2 (and keeping in mind the trivial example $A = [1, n]$) that

$$f(n, k, l) = k(n-1); \quad k \geqslant 2\lfloor (l-2)/(n-2) \rfloor + 2.$$

This, to our knowledge, remains the only situation where the value of $f(n, k, l)$ is known precisely.

An obvious shortcoming of Theorem 2 is that it applies only to identical set summands. Potentially distinct summand were dealt with by Szemerédi and Vu in [4]. A particular case of their result, to be compared with Theorems 1 and 2, is as follows.

Theorem 3 (Szemerédi-Vu [4, a particular case of Corollary 5.2])**.** *There exist positive absolute constants C and c with the following property. Suppose that $k, l \geqslant 1$ and $n \geqslant 2$ are integers, and $A_1, \dots, A_k \subseteq [0, l]$ are integer sets, having at least n elements each. If $k > Cl/n$, then the sumset $A_1 + \dots + A_k$ contains an arithmetic progression of length at least ckn.*

Though the proof of Theorem 3, presented in [4], is constructive, the constants C and c are not computed explicitly. Indeed, the argument leads to excessively large values of these constants, which may present a problem in some applications.

The goal of this note is to merge the best of the two worlds, extending Theorem 2 onto distinct set summands, or, equivalently, proving a sharp analogue if Theorem 3, with Cl/n and ckn replaced with best possible expressions. A result of this sort (up to a technical restriction, addressed below) is an almost immediate corollary from the following theorem, proven in Section 2.

Theorem 4. *Let $l \geqslant 1$ and $n \geqslant 3$ be integers, and write $\kappa := \lceil (l-1)/(n-2) \rceil - 1$. Suppose that $A_1, \dots, A_{2\kappa+1} \subseteq [0, l]$ are integer sets, having at least n elements each, such that none of them are contained in an arithmetic progression with difference greater than 1. Then the sumset $A_1 + \dots + A_{2\kappa+1}$ contains a block of consecutive integers of length $2(\kappa+1)(n-1) - l \geqslant l$.*

Observe, that Theorem 4 guarantees the existence of a block of consecutive integers of length l in $A_1 + \dots + A_k$ for $k = 2\kappa + 1$, which is a sharp threshold: for $k = 2\kappa$ this sumset may fail to contain such a block, as witnessed, say, by the system of identical sets

$$A_1 = \dots = A_{2\kappa} = [0, m] \cup [l - m, l]$$

with $m \in [1, l/2)$ integer and $n = 2m + 2$. Indeed, in this case we have

$$A_1 + \dots + A_{2\kappa} = \bigcup_{j=0}^{2\kappa} [j(l-m), j(l-m) + 2\kappa m],$$

the length of each individual segment being $2\kappa m = \kappa(n-2) < l$, and the segments not abutting, provided $\{(l-1)/(n-2)\} > 1/2$.

Unfortunately, we were unable to eliminate the assumption that none of the sets A_j are contained in an arithmetic progression with difference greater than 1. This reminiscent of the condition $\gcd(A) = 1$ from Theorem 2 will be discussed in Section 3.

The reader may compare the following corollary of Theorem 4 against Theorem 3.

Corollary 1. *Suppose that $k, l \geqslant 1$ and $n \geqslant 3$ are integers, and $A_1, \dots, A_k \subseteq [0, l]$ are integer sets, having at least n elements each, such that none of them are contained in an arithmetic progression with difference greater than 1. If $k \geqslant 2\lceil (l-1)/(n-2) \rceil$, then the sumset $A_1 + \dots + A_k$ contains a block of consecutive integers of length $k(n-1)$.*

Proof. Reducing the value of l and renumbering the sets, if necessary, we can assume that $0, l \in A_k$. Write $\kappa := \lceil (l-1)/(n-2) \rceil - 1$. By Theorem 4, the sumset $A_1 + \dots + A_{2\kappa+1}$ contains a block of consecutive integers of length at least $2(\kappa+1)(n-1) - l \geqslant l$. Adding one by one the sets $A_{2\kappa+2}, \dots, A_{k-1}$ to this sumset, we increase the length of the block by at least $n-1$ each time, and adding A_k at the last step we increase the length by l. Consequently, $A_1 + \dots + A_k$ contains a block of length at least

$$(2(\kappa+1)(n-1) - l) + (k - 2\kappa - 2)(n-1) + l = k(n-1).$$

□

2. Proof of Theorem 4

For a finite, non-empty integer set A, let $\ell(A)$ denote the difference of the largest and the smallest elements of A.

Our approach is fairly close to that employed in [2], with the following result in its heart.

Theorem 5 (Lev [1, Theorem 1])**.** *Let $k \geqslant 2$ be an integer, and suppose that $A_1, \dots, A_k$ are finite, non-empty integer sets. If $\ell(A_j) \leqslant \ell(A_k)$ for $j = 1, \dots, k-1$ and A_k is not contained in an arithmetic progression with difference greater than 1, then*

$$|A_1 + \dots + A_k| \geqslant |A_1 + \dots + A_{k-1}| + \min\{\ell(A_k),\, n_1 + \dots + n_k - k + 1\},$$

where

$$n_j = \begin{cases} |A_j| & \text{if } \ell(A_j) < \ell(A_k) \\ |A_j| - 1 & \text{if } \ell(A_j) = \ell(A_k) \end{cases}; \quad j = 1, \dots, k.$$

Up to some subtlety which we suppress for the moment, the strategy pursued below is to apply Theorem 5 to show that if k is large enough (which in practice means $k \geqslant 2\kappa + O(1)$), then the densities of the sumsets $A_1 + \dots + A_{\lfloor k/2 \rfloor}$ and $A_{\lfloor k/2 \rfloor + 1} + \dots + A_k$ exceed $1/2$; hence, the box principle leads to the conclusion that the sumset of the two, which is $A_1 + \dots + A_k$, contains a long block of consecutive integers. We start with the second, technically simpler, component.

Lemma 1. *Let L_1 and L_2 be positive integers and suppose that $S_1 \subseteq [0, L_1]$ and $S_2 \subseteq [0, L_2]$ are integer sets. If* $\max\{L_1, L_2\} \leqslant |S_1| + |S_2| - 2$*, then*

$$[L_1 + L_2 - (|S_1| + |S_2| - 2), |S_1| + |S_2| - 2] \subseteq S_1 + S_2.$$

Proof. Given an integer $g \in [0, L_1 + L_2]$, the number of representations $g = s_1 + s_2$ with arbitrary $s_1 \in [0, L_1]$ and $s_2 \in [0, L_2]$ is

$$\begin{aligned} |(g - [0, L_1]) \cap [0, L_2]| &= |[g - L_1, g] \cap [0, L_2]| \\ &= \min\{g, L_2\} - \max\{g - L_1, 0\} + 1 \\ &= \min\{g, L_2\} + \min\{g, L_1\} - g + 1. \end{aligned}$$

In order for $g \in S_1 + S_2$ to hold, it suffices that this number of representations exceeds the total number of "gaps" in S_1 and S_2; that is,

$$\min\{g, L_2\} + \min\{g, L_1\} - g + 1 > L_1 + L_2 + 2 - |S_1| - |S_2|.$$

This, however, follows immediately for every g in the interval $[L_1 + L_2 - (|S_1| + |S_2| - 2), |S_1| + |S_2| - 2]$, by considering the location of g relative to L_1 and L_2. $\square$

We now turn to the more technical part of the argument, consisting in inductive application of Theorem 5.

Proposition 1. *Let $k, l \geqslant 1$, and $n \geqslant 3$ be integers, and suppose that $A_1, \dots, A_k$ are integer sets with $|A_i| \geqslant n$ and $\ell(A_i) \leqslant l$ for $i = 1, \dots, k$, such that none of these sets are contained in an arithmetic progression with difference greater than* 1*. Write $S = A_1 + \dots + A_k$.*

(i) *If $k \geqslant (l-1)/(n-2) - 1$, then $|S| \geqslant \frac{1}{2}\left(\ell(S) + (k+1)(n-1) - l + 2\right)$;*

(ii) *if $k \geqslant (l-1)/(n-2)$, then $|S| \geqslant \frac{1}{2}\left(\ell(S) + k(n-1) + 2\right)$.*

Observe, that under assumption (i) we have $(k+1)(n-1) - l + 2 > 0$, so that $|S| > \ell(S)/2$ in this case. Similarly, under assumption (ii) we have $k(n-1) + 2 > l$, and hence in this case the stronger estimate $|S| > (\ell(S) + l)/2$ holds.

Proof of Proposition 1. Write $l_j = \ell(A_j)$. Without loss of generality, we can assume that $l_1 \leqslant \dots \leqslant l_k$. By Theorem 5 and in view of $\ell(S) = l_1 + \dots + l_k$, we have

$$\begin{aligned} |S| - \frac{1}{2}\,\ell(S) &\geqslant \sum_{j=1}^{k} \big(\min\{l_j - 1, j(n-2)\} + 1\big) + 1 - \frac{1}{2}\,(l_1 + \dots + l_k) \\ &= \sum_{j=1}^{k} \min\left\{\frac{l_j - 1}{2}, j(n-2) - \frac{l_j - 1}{2}\right\} + \frac{k}{2} + 1. \end{aligned} \tag{1}$$

Starting with assertion (ii), assume that $k \geqslant (l-1)/(n-2)$. Color all integers $j \in [1,k]$ red or blue, according to whether $l_j - 1 \geqslant j(n-2)$ or $l_j - 1 < j(n-2)$. Notice that the integer 1 is then colored red, hence the interval $[1,k]$ can be partitioned into a union of adjacent subintervals $J_1 \cup \dots \cup J_K$ so that each subinterval consists of a block of consecutive red integers followed by a block of consecutive blue integers, with all blocks non-empty — except that the rightmost subinterval J_K may consist of a "red block" only. Accordingly, we write the sum over $j \in [1,k]$ in the right-hand side of (1) as $\sigma_1 + \dots + \sigma_K$, where for each $\nu \in [1,K]$ by σ_ν we denote the sum over all $j \in J_\nu$.

Fixing $\nu \in [1,K]$, write $J_\nu = [s,t]$ and define q to be the largest red-colored number of the interval $[s,t-1]$. This *does not* define q properly if $\nu = K$ and $s = t = k$; postponing the treatment of this exceptional case, suppose for the moment that q is well defined. Thus either $q+1$ is blue, whence $l_{q+1} - 1 < (q+1)(n-2)$, or $\nu = K$, $t = k$, and $q = k-1$, whence

$$l_{q+1} - 1 = l_k - 1 \leqslant l - 1 \leqslant k(n-2) = (q+1)(n-2).$$

Observe, that

$$l_{q+1} - 1 \leqslant (q+1)(n-2) \tag{2}$$

holds in either case, and it follows that

$$\begin{aligned}
\sigma_\nu &= \sum_{j=s}^{q} \left(j(n-2) - \frac{l_j - 1}{2} \right) + \sum_{j=q+1}^{t} \frac{l_j - 1}{2} \\
&\geqslant \frac{n-2}{2}\,(q^2 + q - s^2 + s) - \frac{l_q - 1}{2}\,(q + 1 - s) + \frac{l_q - 1}{2}\,(t - q) \\
&= \frac{n-2}{2}\,(q^2 + q - s^2 + s) + \frac{l_q - 1}{2}\,(s + t - 1 - 2q).
\end{aligned}$$

We now distinguish two cases: $q \leqslant (s+t-1)/2$ and $q \geqslant (s+t-1)/2$. In the former case we have

$$\begin{aligned}
\sigma_\nu &\geqslant \frac{n-2}{2}\,(q^2 + q - s^2 + s + q(s + t - 1 - 2q)) \\
&= \frac{n-2}{2}\,(q(s + t - q) - s^2 + s) \\
&= \frac{n-2}{2}\,(st + (q - s)(t - q) - s^2 + s) \\
&\geqslant \frac{n-2}{2}\,(st - s^2 + s) \\
&= \frac{n-2}{2}\,s|J_\nu|.
\end{aligned}$$

In the latter case, taking into account that $l_q - 1 \leqslant l_{q+1} - 1 \leqslant (q+1)(n-2)$ by (2), we obtain the same estimate:

$$\begin{aligned}\sigma_\nu &\geqslant \frac{n-2}{2}\,(q^2+q-s^2+s+(q+1)(s+t-1-2q)) \\ &= \frac{n-2}{2}\,(q(s+t-2-q)-s^2+2s-1+t) \\ &\geqslant \frac{n-2}{2}\,((t-1)(s-1)-s^2+2s-1+t) \\ &= \frac{n-2}{2}\,(st-s^2+s) \\ &= \frac{n-2}{2}\,s|J_\nu|.\end{aligned}$$

Addressing finally the situation where $\nu = K$ and $s = t = k$, we observe that in this case

$$\begin{aligned}\sigma_\nu &= k(n-2) - \frac{l_k - 1}{2} \\ &\geqslant k(n-2) - \frac{l-1}{2} \\ &\geqslant \frac{k(n-2)}{2} \\ &= \frac{n-2}{2}\,s|J_\nu|,\end{aligned}$$

as above. Thus,

$$\sigma_1 + \cdots + \sigma_K \geqslant \frac{n-2}{2}\,\big(|J_1| + \cdots + |J_K|\big) = k\,\frac{n-2}{2},$$

and substituting this into (1) we obtain assertion (ii).

To prove assertion (i), instead of $k \geqslant (l-1)/(n-2)$ assume now the weaker bound $k \geqslant (l-1)/(n-2) - 1$. Set $l_{k+1} = l_k$, so that $l_{k+1} - 1 \leqslant l - 1 \leqslant (k+1)(n-2)$. From (1) we get

$$|S| - \frac{1}{2}\,\ell(S) \geqslant \sum_{j=1}^{k+1} \min\left\{\frac{l_j-1}{2}, j(n-2) - \frac{l_j-1}{2}\right\} - \frac{l_{k+1}-1}{2} + \frac{k}{2} + 1.$$

Since $k+1 \geqslant (l-1)/(n-2)$, the sum over j can be estimated as above, and it is at least $(k+1)(n-2)/2$. Assertion (i) now follows in view of

$$\begin{aligned}&\frac{(k+1)(n-2)}{2} - \frac{l_{k+1}-1}{2} + \frac{k}{2} + 1 \\ &\qquad \geqslant \frac{1}{2}\,\big((k+1)(n-2) - (l-1) + k + 2\big) = \frac{1}{2}\,\big((k+1)(n-1) - l + 2\big).\end{aligned}$$

$\square$

Proposition 1 took us most of the way to the proof of Theorem 4.

Proof of Theorem 4. Assume that the sets $A_1, \dots, A_{2\kappa+1}$ are so numbered that, letting $l_j = \ell(A_j)$ for $j \in [1, 2\kappa+1]$, we have $l_1 \geqslant l_2 \geqslant \dots \geqslant l_{2\kappa+1}$. We are going to partition our sets into two groups to apply Lemma 1 to the sumsets S_1 and S_2 of these groups, and this is to be done rather carefully as effective application of the lemma requires that S_1 and S_2 be of nearly equal length.

Accordingly, we let $S_1 := A_1 + A_3 + \dots + A_{2\kappa-1}$ (including all sets A_j with odd indices $j < 2\kappa$) and $S_2 := (A_2 + A_4 + \dots + A_{2\kappa}) + A_{2\kappa+1}$ (all sets A_j with even indices $j \leqslant 2\kappa$ and $A_{2\kappa+1}$). By Proposition 1 we have

$$|S_1| \geqslant \frac{1}{2}\left(\ell(S_1) + (\kappa+1)(n-1) - l + 2\right) \tag{3}$$

and

$$|S_2| \geqslant \frac{1}{2}\left(\ell(S_2) + (\kappa+1)(n-1) + 2\right). \tag{4}$$

Furthermore, from

$$\ell(S_1) - \ell(S_2) = (l_1 - l_2) + \dots + (l_{2\kappa-1} - l_{2\kappa}) - l_{2\kappa+1} \in [-l, l],$$

using (3) and (4) we get

$$\begin{aligned} \max\{\ell(S_1), \ell(S_2)\} &\leqslant \frac{1}{2}\left(\ell(S_1) + \ell(S_2) + l\right) \\ &\leqslant |S_1| + |S_2| + l - (\kappa+1)(n-1) - 2 \leqslant |S_1| + |S_2| - 2. \end{aligned}$$

Applying Lemma 1 and using again (3) and (4), we conclude that $A_1 + \dots + A_k = S_1 + S_2$ contains a block of consecutive integers of length at least

$$2(|S_1| + |S_2| - 2) - (\ell(S_1) + \ell(S_2)) \geqslant 2(\kappa+1)(n-1) - l,$$

as required. □

3. Concluding remarks and open problems

The major challenge arising in connection with the main results of this paper (which are Theorem 4 and Corollary 1) is to get rid of the assumption that none of the sets involved are contained in an arithmetic progression with difference greater than 1. One can expect that a vital ingredient of such an improvement would be a suitable generalization of Theorem 5. Indeed, we were able to generalize Theorem 5 in what seems to be the right direction.

Theorem 5′. *Let $k \geqslant 2$ be an integer, and let $A_1, \dots, A_k$ be finite, non-empty integer sets. Suppose that $l \in A_k - A_k$ is a positive integer, and define d to be the largest integer such that A_k is contained in an arithmetic progression with difference d. Then*

$$|A_1 + \dots + A_k| \geqslant |A_1 + \dots + A_{k-1}| + \min\{hl/d,\ n_1 + \dots + n_k - k + 1\},$$

where h is the number of residue classes modulo d, represented in $A_1 + \dots + A_{k-1}$, and n_j $(j = 1, \dots, k)$ is the number of residue classes modulo l, represented in A_j.

Clearly, Theorem 5′ implies Theorem 5 and, furthermore, shows that the assumption of Theorem 4 that none of the sets A_j are contained in an arithmetic progression with difference greater than 1 can be slightly relaxed. Specifically, it suffices to request that the A_j can be so ordered that $\ell(A_j)$ increase, and for every $k \in [1, 2\kappa]$ the sumset $A_1 + \dots + A_k$ represents all residue classes modulo $\gcd(A_{k+1} - A_{k+1})$. It seems, however, that Theorem 5′ by itself fails short to extend Theorem 4 the desired way, dropping the modular restriction altogether, and in the absence of applications we do not present here the proof of the former theorem.

Another interesting direction is to refine Theorem 4 as to the length of the block, contained in the sumset $A_1 + \dots + A_{2\kappa+1}$. While we observed that $2\kappa + 1$ is the smallest number of summands which ensures a block of length l, it is quite possible that the existence of a longer block can be guaranteed. In this connection we mention the following conjecture from [2], referring to the equal summands situation.

Conjecture 1. *Let $k, l \geqslant 1$ and $n \geqslant 3$ be integers, and write $\kappa := \lfloor (l-1)/(n-2) \rfloor$. Suppose that $A \subseteq [0, l]$ is an integer set with $0, l \in A$, $\gcd(A) = 1$, and $|A| = n$. If $k \geqslant 2\kappa + 1$, then kA contains a block of consecutive integers of length $(k-\kappa)l + k((\kappa+1)(n-2) + 2 - l)$.*

In fact, Conjecture 1 is established in [2] in the case where $k \geqslant 3\kappa$, but the case $2\kappa + 1 \leqslant k < 3\kappa$, to our knowledge, remains open.

References

[1] V. Lev, Addendum to "Structure theorem for multiple addition", *Journal of Number Theory* **65** (1) (1997), 96–100.
[2] ———, Optimal representations by sumsets and subset sums, *Journal of Number Theory* **62** (1) (1997), 127–143.
[3] A. Sárközy, Finite Addition Theorems, I, *J. Number Theory* **32** (1989), 114–130.
[4] E. Szemerédi and V. Vu, Long arithmetic progressions in sumsets: thresholds and bounds, *J. Amer. Math. Soc.* **19** (1) (2006), 119–169.

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