SUMSET GROWTH IN PROGRESSION-FREE SETS

CHRISTIAN ELSHOLTZ, IMRE Z. RUZSA, AND LENA WURZINGER

Abstract. We study the growth of sumsets $\mathcal{A}+\mathcal{B}\subset\mathcal{S}\subset G$, where $\mathcal{S}$ does not contain an arithmetic progression of length $2k+1$, and where G is a commutative group, in which every nonzero element has an order of at least $2k+1$. More specifically, we show the following: if $\mathcal{A},\mathcal{B}\subset G$ are sets such that $\mathcal{A}+\mathcal{B}$ does not contain an arithmetic progression of length $2k+1$, then we have that

$$|\mathcal{A}+\mathcal{B}|\geq|\mathcal{A}|^{\frac{2k-1}{3k-2}}|\mathcal{B}|^{\frac{k}{3k-2}}.$$

As an application we derive upper bounds on the cardinality of the summands in sumsets $\mathcal{A}+\mathcal{B}+\mathcal{C}$ contained in the set of t-th powers, where $t\geq 2$ is an integer. In particular, we show that $\min(|\mathcal{A}|,|\mathcal{B}|,|\mathcal{C}|)\ll(\log N)^{4/5}$ for $t=2$, and $\min(|\mathcal{A}|,|\mathcal{B}|,|\mathcal{C}|)\ll_t(\log N)^{1/2}$ for $t\geq 3$.

1. Introduction

In this paper we study the growth of sumsets which lie in progression-free sets. Below we give a survey of the existing literature, and surprisingly it seems that this question has not been studied before. An application to sumsets in the set of t-th power is discussed in section 3, and we hope that other applications will be found soon.

Here is our main result:

Theorem 1.1. *(1) Let G be a commutative group, in which every nonzero element has order at least $2k+1$. Let $\mathcal{A},\mathcal{B}\subset G$ be finite sets such that $\mathcal{A}+\mathcal{B}=\{a+b: a\in\mathcal{A}, b\in\mathcal{B}\}$ does not contain an arithmetic progression of length $2k+1$. Then the following estimate on the size of the sumset $|\mathcal{A}+\mathcal{B}|$ holds:*

$$|\mathcal{A}+\mathcal{B}|\geq|\mathcal{A}|^{\frac{2k-1}{3k-2}}|\mathcal{B}|^{\frac{k}{3k-2}}.$$

(2) The special case $k=1$ also holds in the non-abelian case.[1] *Indeed, if $\mathcal{A}+\mathcal{B}$ does not contain three elements x,y,z with $z-y=y-x$, then*

$$|\mathcal{A}+\mathcal{B}|=|\mathcal{A}||\mathcal{B}|,$$

2020 *Mathematics Subject Classification.* primary: 11P70; secondary: 11B25, 11B30, 20K99.

Key words and phrases. sumset growth, sumsets in the integer squares, progression-free sets.

[1] As one could define arithmetic 3-progressions in different ways in the non-abelian case, we specify which type we use. See also Remark 4.1.

which is clearly the best possible.

This gives an immediate

Corollary 1.2. *Let G be a group (non-abelian allowed), in which every nonzero element has order at least 3. If $\mathcal{C} \subset G$ does not contain three elements x, y, z with $z - y = y - x$, and if $|\mathcal{C}|$ is a prime, then $\mathcal{C}$ is additively irreducible, i.e., it cannot be written as a sumset in a nontrivial way.*

Part (2) of the Theorem looks as if it should be known, but we are not aware of a reference. In general it is very difficult to decide whether a given set is additively irreducible or not.

For $k > 1$ the result is asymmetric: when choosing $\mathcal{A}$ as the larger of the two summands, one achieves the stronger estimate. Still, if $|\mathcal{B}|$ is much smaller than $|\mathcal{A}|$, then the estimate on $|\mathcal{A} + \mathcal{B}|$ can be weaker than the obvious lower bound $|\mathcal{A}|$.

Also note that density does not play any role here, there is no reference to the size of an ambient interval or space. This is e.g. in contrast to [1, 6]. For the applications given in section 3 (whose proofs can be found in section 5) however, we will apply the results to the case of integers in a fixed interval $[1, N]$.

All commutative torsion-free groups are covered by the theorem, but in this paper we mainly think of the cases $G = \mathbb{Z}$ or $G = \mathbb{Z}^d$. It is surprising that to the best of our knowledge a result of this type does not exist in the literature, not even in the case of integers, given that sumset growth and progression-free sets are well-established topics.

Bourgain, Dilworth, Ford, Konyagin and Kutzarova [5] investigated a quite special class of this problem on sumsets with restricted length of arithmetic progressions. Here both summands $\mathcal{A}, \mathcal{B} \subset \{0, \ldots, r-1\}^d$ are restricted to a multi-dimensional cube of sidelength r, which restricts the maximal length of progressions in the summands to r and which also restricts the maximal length of a progression in $\mathcal{A}+\mathcal{B}$ to at most $2r-1$. (Also variants with $\mathcal{A}$ and $\mathcal{B}$ in two different cubes are considered.) They provide the lower bound

$$|\mathcal{A} + \mathcal{B}| \geq (|\mathcal{A}||\mathcal{B}|)^{\tau},$$

where τ is the real solution of the equation $\frac{1}{r^{2\tau}} + (\frac{r-1}{r})^{\tau} = 1$. If one restricts our general result to the special case $\mathcal{A}, \mathcal{B} \subset \{0, \ldots, r-1\}^d$, then their bound is stronger than ours. They also conjecture that the value $\tau' = \frac{\log(2r-1)}{2 \log r}$

would be admissible, replacing τ. This would be the best possible as can be seen from $\mathcal{A} = \mathcal{B} = \{0, \ldots, r-1\}$.

The special case $r = 2$ of this conjecture had been proved much earlier by Woodall [33] and independently by Hajela and Seymour [18]. They proved the inequality $|\mathcal{A}+\mathcal{B}| \geq (|\mathcal{A}||\mathcal{B}|)^{\tau'}$ in the case where $\mathcal{A}$ and $\mathcal{B}$ are contained in $\{0,1\}^d$ with $\tau' = \frac{\log 3}{2\log 2}$. Becker, Ivanisvili, Krachun and Madrid [2] proved the conjecture when $r = 3$, namely that for all $\mathcal{A}, \mathcal{B} \subset \{0,1,2\}^d$ one has $|\mathcal{A}+\mathcal{B}| \geq (|\mathcal{A}||\mathcal{B}|)^{\log 5/(2\log 3)}$ and give optimal lower bounds for iterated sumsets $\mathcal{A}_1 + \cdots + \mathcal{A}_k$, where $\mathcal{A}_i \in \{0,1\}^d$.

Example: (Progression-free sets in $\mathbb{Z}_m^d$ or $\mathbb{Z}^d$)
It is known (more or less by Salem and Spencer [28], see also Ruzsa [25, p. 147]), that the set of integers of the type $n = \sum_{i=0}^k a_i 5^i$, with $a_i \in \{0,1,2\}$ using a fixed number (say $\lfloor k/3 \rfloor$) of $a_i = 1$, is 3-progression-free. As the infinite set of integers

$$\bigcup_{k=0}^{\infty}\{\sum_{i=0}^{k} a_i 5^i, a_i \in \{0,1,2\}\} = \mathcal{A} + \mathcal{A}$$

with $\mathcal{A} = \bigcup_{k=0}^{\infty}\{\sum_{i=0}^{k} a_i 5^i, a_i \in \{0,1\}\}$ decomposes into a sumset, it is a natural question if progression-free sets of integers also decompose, and if so, what are the size constraints on the summands. The well-known 3-progression-free set by Szekeres [14], namely $\mathcal{M} = \{\sum_{i=0}^{t} a_i 3^i : a_i \in \{0,1\}\}$, where t is a fixed positive integer, naturally decomposes as $\mathcal{M} = \mathcal{A} + \mathcal{B}$ in several ways: for example, when $\mathcal{A}$ has the indices with $i \in I \subset \{0, 1, \ldots, t'\}$, and $\mathcal{B}$ has the remaining indices, (or $\mathcal{A}$ has the odd indices, etc.). It follows in these examples, but more generally by Theorem 1.1 also for an *arbitrary* additive decomposition (independent of the above natural decompositions), that $|\mathcal{M}| = |\mathcal{A}+\mathcal{B}| = |\mathcal{A}||\mathcal{B}|$ holds.

It is known from recent results [12] around Behrend's construction that in the cube $\mathbb{Z}_m^d$ there exists a 3-progression-free subset $\mathcal{C}$ of size about $((c+o(1))m)^d$, for some $c > 0.54$. One may wonder about possible sizes of the summands in a decomposition $\mathcal{C} = \mathcal{A} + \mathcal{B}$ (if it exists); Theorem 1.1 implies $|\mathcal{C}| = |\mathcal{A}||\mathcal{B}|$.

2. Related results in the literature

There is a large number of different types of results on sumsets and progression-free sets that have been studied before. In view of this extensive list one can certainly conclude that Theorem 1.1 deals with a natural question that was overlooked so far. Here we list several related results:

(1) When at least one of the summands is progression-free, Ruzsa [26, Theorem 2.9.1] and [24] proved: Let $\mathcal{A}$ and $\mathcal{B}$ be finite sets of integers and assume that $\mathcal{A}$ does not contain any arithmetic progressions of length k. Then

$$|\mathcal{A}+\mathcal{B}| \geq \frac{1}{\sqrt{2}}\omega_k(|\mathcal{A}|)^{1/4}|\mathcal{A}|^{1/4}|\mathcal{B}|^{3/4},$$

where $\omega_k(N) = N/r_k(N)$, and $r_k(N)$ denotes the largest cardinality of a subset $A \subset [1, N]$ that does not contain any arithmetic progressions of length k.

(2) Let $L(\mathcal{C})$ denote the size of the longest arithmetic progression of $\mathcal{C}$. Improving a result of Bourgain [4], Green [16] proved: Suppose $\mathcal{A}, \mathcal{B}$ are subsets of $\mathbb{Z}/N\,\mathbb{Z}$ having cardinalities γN and δN, respectively. Then there is an absolute constant $c > 0$ such that

$$L(\mathcal{A}+\mathcal{B}) > \exp(c((\gamma\delta\log N)^{1/2} - \log\log N)).$$

A limit to such results comes from a bound of Ruzsa [23]: There exists a set $\mathcal{A} \subset [1, N]$ with $|\mathcal{A}| > (\frac{1}{2} - \varepsilon)N$, such that $L(2\mathcal{A}) < \exp((\log N)^{2/3+\varepsilon})$.

(3) Sanders [29] proved: if $\mathcal{A}$ is a finite subset of an abelian group, and if $\mathcal{A}$ does not contain any arithmetic progression of length 3, then

$$|\mathcal{A}+\mathcal{A}| \gg |\mathcal{A}|(\log|\mathcal{A}|)^{\frac{1}{3}}/\log\log|\mathcal{A}|.$$

This was improved by Henriot [20] to about $|\mathcal{A}|(\log|A|)^{1+o(1)}$. Very recently, this was updated by the late Tomasz Schoen [30] based on the new Kelley-Meka bounds [21]. Related earlier results are by Ruzsa [24] and Stanchescu [31].

(4) Freiman, Halberstam and Ruzsa [15] show that long arithmetic progressions exist in iterated sums of high order. Further results on ternary sumsets are by Henriot [19], Kelley and Meka [21], Bloom and Sisask [3].

(5) Croot, Ruzsa and Schoen [6] proved: For every odd $k' \geq 1$ and N sufficiently large, if $\mathcal{A}, \mathcal{B} \subset [1, N]$ and $L(\mathcal{A}+\mathcal{B}) < k'$, then $|\mathcal{A}|\,|\mathcal{B}| < 6N^{2-2/(k'-1)}$.

(6) Dietmann and Elsholtz [9, 10] proved that in the case of Hilbert cubes $\mathcal{H} = a_0 + \{0, a_1\} + \cdots + \{0, a_d\} \subset \mathcal{S}$, where $a_0 \in \mathbb{Z}, a_1, \ldots, a_d \in \mathbb{N}$ and where $\mathcal{S}$ does not contain any arithmetic progression of length $k \geq 3$, one actually gets an exponential growth:

$$|\,\mathcal{H}| \geq 2\left(\frac{k}{k-1}\right)^{d-1} - 1.$$

(7) Szemerédi and Vu [32] have a series of three papers on long arithmetic progressions in sumsets.

(8) Alon et al. [1] prove that arithmetic progressions of sumsets of n sets, each logarithmically sparse, have length bounded by $n^{n(1+o(1))}$, which is essentially the best possible.

Also other types of conditions have been studied that guarantee sumset growth, see for example for convex sets Elekes, Nathanson and Ruzsa [11], for semiconvex sets Ruzsa and Solymosi [27] or for ternary sumsets including a quasicube Matolcsi, Ruzsa, Shakan, and Zhelezov [22].

3. Application to sumsets in the set of t-th powers

Theorem 1.1 has an immediate application to the study of ternary sumsets in the set of squares or t-th powers, which did not seem available otherwise. Note that the set of t-th powers is four-progression-free for the set of squares ($t = 2$), proved by Fermat, see Dickson [8, page 440], and even three-progression-free for $t \geq 3$ [7]. This application deals with one of the many problems in additive number theory where the size of an interval and counting functions come in naturally, even though this is not the case in Theorem 1.1.

Let us first recall a result of Gyarmati [17] on binary sums in the set of squares.

Theorem (Theorem 9 in [17])**.** Let $\mathcal{A}, \mathcal{B} \subset [1, N]$ with $\mathcal{A}+\mathcal{B} \subset \mathcal{S}_t$, where $\mathcal{S}_t$ denotes the set of t-th powers in the positive integers. Then, for sufficiently large N, the following holds: $\min(|\mathcal{A}|, |\mathcal{B}|) \leq 4t \log N$.

From the methods of that paper, it does not seem possible to improve this minimal size for binary and even for ternary sumsets. Here we show that if $\mathcal{A} + \mathcal{B} + \mathcal{C} \subset S_t$, then

$$\min(|\mathcal{A}|, |\mathcal{B}|, |\mathcal{C}|) = O((\log N)^{\frac{4}{5}})$$

when $t = 2$, i.e., the case where $\mathcal{A}+\mathcal{B}+\mathcal{C}$ is contained in the set of squares, and

$$\min(|\mathcal{A}|, |\mathcal{B}|, |\mathcal{C}|) = O_t((\log N)^{\frac{1}{2}})$$

for $\mathcal{A} + \mathcal{B} + \mathcal{C} \subset \mathcal{S}_t$ with $t \geq 3$.

In order to do so, we need two new ingredients. The first is a binary sumset estimate in the asymmetric case for sumsets $\mathcal{A}+\mathcal{B}$ in the set of t-th powers, i.e., when $\mathcal{A}$ and $\mathcal{B}$ are not of about the same size (see Lemma 5.3 below). Secondly, the nontrivial sumset growth estimates in progression-free sets from Theorem 1.1 are essential. This can be extended to the following results:

Theorem 3.1. *Let N be sufficiently large. Assume that $\mathcal{A},\mathcal{B},\mathcal{C} \subset [1,N]$ such that $\mathcal{A}+\mathcal{B}+\mathcal{C} \subset \mathcal{S}$, where $\mathcal{S}$ denotes the set of squares. Assume that $|\mathcal{A}| \le |\mathcal{B}| \le |\mathcal{C}|$. Then the following holds:*

1. *$|\mathcal{A}| = O((\log N)^{\frac{4}{5}})$, i.e., in the symmetric case $|\mathcal{A}| = |\mathcal{B}| = |\mathcal{C}|$ we also have that $|\mathcal{A}| = |\mathcal{B}| = |\mathcal{C}| = O((\log N)^{4/5})$.*
2. *Assume that $|\mathcal{A}| \ge C \log\log N$, for sufficiently large C. Then also the second smallest set can be bounded from above:*
$$|\mathcal{B}| = O((\log N)^{\frac{4}{5}+o(1)}),$$
while for the largest set, we have the bound
$$|\mathcal{C}| = O\left(\log N \left(\frac{\log\log N}{\log\log\log N}\right)^2\right).$$
3. *Assume that $|\mathcal{A}| \gg (\log N)^{\delta}$ for any $\delta \in (0, \frac{4}{5}]$, then*
$$|\mathcal{B}| = O_{\delta}((\log N)^{\frac{4}{5}})$$
and $|\mathcal{C}| \ll_{\delta} \min(\log N, ((\log N)/|\mathcal{B}|^{1/2})^{4/3})$.

Remark 3.2. Further bounds can be proved for iterated sumsets. We give another example here, and discuss this idea in more detail in section 6.

Let $\mathcal{A}_1 + \cdots + \mathcal{A}_5 \subset \mathcal{S}$ and let $|\mathcal{A}_i| \le |\mathcal{A}_j|$ when $i < j$. Then $|\mathcal{A}_1| = O((\log N)^{\frac{64}{101}})$ and $|\mathcal{A}_2| = O((\log N)^{\frac{16}{23}})$. Furthermore, there is a constant $C > 0$ such that if $|\mathcal{A}_1| \ge C \log\log N$, then $|\mathcal{A}_2| = O((\log N)^{\frac{64}{101}+o(1)})$ and $|\mathcal{A}_3| = O((\log N)^{\frac{16}{23}+o(1)})$.

Theorem 3.3. *Let $t \ge 3$ and let $\mathcal{S}_t$ denote the set of positive t-th powers. Moreover, assume that $\mathcal{A}_1 + \mathcal{A}_2 + \cdots + \mathcal{A}_d \subset \mathcal{S}_t$ and that $|\mathcal{A}_i| \le |\mathcal{A}_j|$ when $i < j$. Then we have for $i = 1, \ldots, d-1$ the following estimate:*
$$|\mathcal{A}_i| \ll_t (\log N)^{\frac{1}{d-i}+o(1)}.$$

Remark 3.4. Furthermore, in the situation of Theorem 3.3 it holds by Gyarmati's result cited above that
$$Q = |\mathcal{A}_1||\mathcal{A}_2| \cdots |\mathcal{A}_{\lfloor d/2 \rfloor}| \le 4t \log N.$$
This implies in particular $d \le 2(\log\log N/\log \mathcal{A}_{\text{g.m.}}) + O_t(1)$, where $\mathcal{A}_{\text{g.m.}}$ denotes the geometric mean of the $\lfloor d/2 \rfloor$ factors. This extends the special case of Hilbert cubes [10] with $|\mathcal{A}_i| = 2$.

Similarly one can achieve good upper bounds on the product of the larger summands if we have a lower bound on Q. For instance, if $Q \geq c(t)\log\log N$, then there is a constant $c_2(t) > 0$ such that

$$|\mathcal{A}_{\lfloor d/2\rfloor+1}|\cdots|\mathcal{A}_d| \leq c_2(t)\log N(\log\log N)^2$$

by Lemma 5.4 below.

4. Proof of Theorem 1.1

We first prove the easier second part of Theorem 1.1. Suppose that $|\mathcal{A}+\mathcal{B}| < |\mathcal{A}||\mathcal{B}|$, then there exist four elements $a, a' \in \mathcal{A}$, and $b, b' \in \mathcal{B}$ with $a + b = a' + b'$, and then the following equations hold for the differences: (Note that e.g., $-(a+b) = -b-a$ in a nonabelian group, written additively.)

$$\begin{aligned} b' - b &= -a' + a \\ a + b' - b - a &= a - a' \\ (a + b') - (a + b) &= (a + b) - (a' + b) = d \quad \text{(say)}. \end{aligned}$$

Therefore the following arithmetic progression of length 3 is contained in $\mathcal{A}+\mathcal{B}$:

$$\begin{aligned} -d + (a + b) &= (a' + b) - (a + b) + (a + b) = a' + b \in \mathcal{A}+\mathcal{B} \\ (a + b) &\in \mathcal{A}+\mathcal{B} \\ d + (a + b) &= a + b' \in \mathcal{A}+\mathcal{B} \end{aligned}$$

because $d+(a+b) = a+b-b-a'+(a+b) = a+(-a'+a)+b = a+(b'-b)+b = a+b'$. The abelian case is even easier to prove: If $d = b' - b = a - a'$, then $a + b - d = a' + b, a + b, a + b + d = a + b'$ is a 3-progression in $\mathcal{A}+\mathcal{B}$.

Remark 4.1. In the non-abelian case there is another type of an arithmetic progression, $x + z = y + y$. We give an example with matrices and thus use multiplicative notation. We show that it is possible that $|A \circ B| < |A||B|$, but that there is no solution of $y^2 = xz$ with distinct matrices $x, y, z \in A\circ B$. Let $A = \{a_1, a_2\}, B = \{b_1, b_2\}$ with

$$a_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, a_2 = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}, b_1 = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, b_2 = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Observe that $A \circ B = \left\{ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \right\}$. Let $y \in A \circ B$, Observe that y^2 is a symmetric matrix, whereas for any two distinct $x, z \in A \circ B$ the matrix xz is not symmetric.

In order to prove part (1) of Theorem 1.1, we first prove the following lemma. Here $k\mathcal{C}$ is defined as

$$k\mathcal{C} = \mathcal{C} + \cdots + \mathcal{C} \ (k \text{ times}).$$

Lemma 4.2. *Let G be a commutative group where every non-zero element has order at least $2k+1$, and assume that $\mathcal{C} \subset G$ is a subset such that*

$$|\mathcal{C}|^k > |\mathcal{C} - 2\mathcal{C}|^{k-1}.$$

Then the set $\mathcal{C} - \mathcal{C}$ contains an arithmetic progression of length $2k+1$ which is symmetric about zero.

Proof of Lemma 4.2. Consider the following map from $\mathcal{C}^k$ to $(\mathcal{C} - 2\mathcal{C})^{k-1}$:

$$f(x_1, x_2, \ldots, x_k) = (x_2 - 2x_1, x_3 - x_2 - x_1, \ldots, x_k - x_{k-1} - x_1).$$

By assumption of the lemma, two values must coincide, and if $f(x_1, \ldots, x_k) = f(y_1, \ldots, y_k)$, then

$$x_{i+1} - y_{i+1} = x_i - y_i + x_1 - y_1, \text{ for } i = 1, \ldots, k,$$

so $x_i - y_i = i(x_1 - y_1)$. (Note that $x_1 \neq y_1$, because otherwise, we would get that $(x_1, \ldots, x_k) = (y_1, \ldots, y_k)$.)

The arithmetic progression of length $2k+1$ consists of these k numbers (i.e., $x_i - y_i, i = 1, \ldots, k$), their negatives and 0. These $2k+1$ elements are all distinct because $x_1 - y_1$ is a non-zero element of G and thus has order at least $2k+1$ by assumption. □

Moreover, we need the following lemma [26, Corollary 1.7.5, p. 105].

Lemma 4.3. *Let $\mathcal{A}, \mathcal{B}$ be sets in a commutative group and write $|\mathcal{A}| = m, |\mathcal{A} + \mathcal{B}| = s$. Let t be a real number, $0 \le t < m$. There is a set $\mathcal{X} \subset \mathcal{A}, |\mathcal{X}| > t$ such that*

$$|\mathcal{X} + 2\mathcal{B}| \le \frac{s^2}{(m-t)^2} \left(|\mathcal{X}| - \frac{t(t+m)}{2m} \right).$$

Proof of Theorem 1.1. Write $|\mathcal{A}| = m$, $|\mathcal{B}| = n$, $|\mathcal{A} + \mathcal{B}| = s$. By Lemma 4.3, with $t = m/2$, there is an $\mathcal{X} \subset A$ such that $|\mathcal{X}| > m/2$ and

$$|\mathcal{X} + 2\mathcal{B}| < 4\frac{s^2}{m^2}|\mathcal{X}| < 4\frac{s^2}{m}.$$

By averaging over all possible values $h \in \mathcal{X} + \mathcal{B}$ there exists h such that

$$|\mathcal{X} \cap (h - \mathcal{B})| \ge \frac{|\mathcal{X}||\mathcal{B}|}{|\mathcal{X} + \mathcal{B}|} > \frac{mn}{2s}.$$

Let $\mathcal{C} = \mathcal{X} \cap (h - \mathcal{B})$. We have $\mathcal{C} - \mathcal{C} \subset \mathcal{A} + \mathcal{B} - h$, so if $\mathcal{A} + \mathcal{B}$ does not contain an arithmetic progression of length $2k+1$, neither does $\mathcal{C} - \mathcal{C}$. By Lemma 4.2 we have

$$|\mathcal{C}|^k \le |\mathcal{C} - 2\mathcal{C}|^{k-1}. \tag{4.1}$$

Also

$$\mathcal{C} - 2\mathcal{C} \subset (\mathcal{X} + 2\mathcal{B}) - 2h,$$

so

$$|\mathcal{C} - 2\mathcal{C}| < 4\frac{s^2}{m}.$$

Substituting this into (4.1) we get

$$s^{3k-2} > 2^{2-3k} m^{2k-1} n^k.$$

The powers of 2 can be removed, by an application of the tensor power trick: Suppose that such an inequality holds for some constant, then for suitable higher products, it would hold with the same constant. This implies that the constant can be chosen as 1. □

5. Proofs of Theorems 3.1 and 3.3

We collect those results that we need for the proof:

Lemma 5.1 (Fermat, see Dickson [8]). *The set of squares does not contain arithmetic progressions of length 4.*

Lemma 5.2 (Darmon and Merel [7]). *Let $t \ge 3$. The set $\mathcal{S}_t$ of positive t-th powers does not contain arithmetic progressions of length 3.*

In the asymmetric case, that is, when the sets in the sumset differ significantly in size, stronger results than those following from Gyarmati's work [17] were recently established:

Lemma 5.3 ([13]). *There exist constants $C_1, C_2 > 0$ such that the following holds. Let $\mathcal{A}, \mathcal{B} \subset [1, N]$ with $\mathcal{A} + \mathcal{B} \subset \mathcal{S}_t$, and let m be a positive integer. If*

$$|\mathcal{B}| \ge C_1 m^2 t \log N,$$

then

$$|\mathcal{A}| \le C_2 m t^{1/m} (\log N)^{1/m}.$$

A natural threshold in Theorem 5.3 for the size of m is $\log\log N$, as for larger m both bounds on $|\mathcal{A}|, |\mathcal{B}|$ get worse. In particular, with $m = \lfloor \log\log N \rfloor$ or $m > \frac{1}{\delta}$ this implies the following special cases:

Lemma 5.4 ([13]). *Let $\mathcal{A}, \mathcal{B} \subset [1, N]$ with $\mathcal{A} + \mathcal{B} \subset \mathcal{S}_t$, where $\mathcal{S}_t$ denotes the set of t-th powers. Then there are positive constants c_1, c_2 such that the following holds. Suppose that*

$$|\mathcal{A}| > c_1 t^{(1/\lfloor \log\log N \rfloor)} \log\log N,$$

then we have a bound on the size of $\mathcal{B}$, namely:

$$|\mathcal{B}| \le c_2 t (\log N)(\log\log N)^2.$$

Moreover, for $\delta > 0$, if

$$|\mathcal{A}| \gg (\log N)^{\delta},$$

then

$$|\mathcal{B}| \ll_{\delta,t} \log N.$$

Proof of Theorem 3.1. (1) Let $\mathcal{A} + \mathcal{B} + \mathcal{C} \subset \mathcal{S}$. Suppose that $|\mathcal{A}| \gg (\log N)^{\delta}$, with $\delta \ge \frac{4}{5}$, then by Theorem 1.1 (applied to $|\mathcal{B} + \mathcal{C}|$ with $k = 2$) and Lemma 5.4

$$|\mathcal{A}|^5 \le |\mathcal{B}|^5 \le |\mathcal{B}|^2 |\mathcal{C}|^3 \le |\mathcal{B} + \mathcal{C}|^4 \ll (\log N)^4,$$

which implies that $|\mathcal{A}| \ll (\log N)^{4/5}$.

(2) If $|\mathcal{A}| \ge C \log\log N$, for sufficiently large N and C, then by Lemma 5.4: $|\mathcal{B} + \mathcal{C}| \ll \log N (\log\log N)^2$. By Theorem 1.1, we have

$$|\mathcal{B}|^5 \le |\mathcal{B}|^2 |\mathcal{C}|^3 \le |\mathcal{B} + \mathcal{C}|^4 \ll ((\log N)(\log\log N)^2)^4,$$

i.e., $|\mathcal{B}| \ll (\log N)^{\frac{4}{5}+o(1)}$.

Furthermore, again by Theorem 1.1,

$$|\mathcal{A} + \mathcal{B}| \ge |\mathcal{A}|^{5/4} \gg (\log\log N)^{5/4},$$

so

$$|\mathcal{A} + \mathcal{B}| \gg m(\log N)^{1/m} \gg \frac{(\log\log N)^{5/4}}{\log\log\log N},$$

where $m = \lfloor \frac{4 \log\log N}{\log\log\log N} \rfloor$. By Lemma 5.3, we get

$$|\mathcal{C}| \ll \left(\frac{\log\log N}{\log\log\log N} \right)^2 \log N.$$

(3) If $|\mathcal{A}| \gg (\log N)^{\delta}$, for sufficiently large N, then by Lemma 5.4: $|\mathcal{B} + \mathcal{C}| \ll_{\delta} \log N$. As before, this gives:

$$|\mathcal{B}|^5 \le |\mathcal{B}|^2 |\mathcal{C}|^3 \le |\mathcal{B} + \mathcal{C}|^4 \ll_{\delta} (\log N)^4,$$

i.e., $|\mathcal{B}| \ll_{\delta} (\log N)^{\frac{4}{5}}$. Moreover, by Lemma 5.4 and Theorem 1.1,

$$|\mathcal{B}|^2 |\mathcal{C}|^3 \le |\mathcal{B} + \mathcal{C}|^4 \ll_{\delta} (\log N)^4,$$

so

$$|\mathcal{C}| \ll_{\delta} (\log N)^{4/3} / |\mathcal{B}|^{2/3}.$$

and by Lemma 5.4 also $|\mathcal{C}| \ll_\delta \log N$.

□

Proof of Theorem 3.3. If $|\mathcal{A}_d| \leq c_1 t \log\log N$ (with the constant $c_1 > 0$ as in Lemma 5.4), then the claim follows by the assumption that $|\mathcal{A}_i| \leq |\mathcal{A}_d|$. Thus we may assume that $|\mathcal{A}_d| > c_1 t \log\log N$, and therefore, by Lemma 5.4,

$$|\mathcal{A}_i + \cdots + \mathcal{A}_{d-1}| \ll_t \log N (\log\log N)^2.$$

By Lemma 5.2, the set $\mathcal{A}_i + \cdots + \mathcal{A}_{d-1}$ does not contain any arithmetic progressions of length three. By Theorem 1.1,

$$|\mathcal{A}_i|^{d-i} \leq |\mathcal{A}_i| \cdots |\mathcal{A}_{d-1}| = |\mathcal{A}_i + \cdots + \mathcal{A}_{d-1}|.$$

Combining these estimates yields

$$|\mathcal{A}_i|^{d-i} \ll_t \log N (\log\log N)^2,$$

which implies the claimed result. □

6. More details on iterated sumsets

If an iterated sumset $\mathcal{A}_1 + \cdots + \mathcal{A}_r$ is $(2k+1)$-progression-free, this property also holds for any shorter sum of a subset of these sets. Consequently, the new bounds established for binary sumsets in Theorem 1.1 induce bounds on iterated sumsets. As the bounds are asymmetric, multiple sumsets can be formed in various ways, depending on the sizes of the sets involved.

6.1. **Proof of Remark 1.** Let $\mathcal{A}_1 + \cdots + \mathcal{A}_5 \subset \mathcal{S}$. We may assume that $|\mathcal{A}_5| \gg (\log N)^{\frac{16}{23}}$ (as otherwise, since $|\mathcal{A}_1|, |\mathcal{A}_2| \leq |\mathcal{A}_5|$, we are done). By Lemma 5.4, we have

$$|\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4| \ll \log N;$$

on the other hand, by repeated application of Theorem 1.1,

$$\begin{aligned}|\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4| &\geq |\mathcal{A}_1|^{\frac{1}{2}} |\mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4|^{\frac{3}{4}} \\ &\geq |\mathcal{A}_1|^{\frac{1}{2}} |\mathcal{A}_2|^{\frac{3}{8}} |\mathcal{A}_3 + \mathcal{A}_4|^{\frac{9}{16}} \\ &\geq |\mathcal{A}_1|^{\frac{1}{2}} |\mathcal{A}_2|^{\frac{3}{8}} |\mathcal{A}_3|^{\frac{9}{32}} |\mathcal{A}_4|^{\frac{27}{64}} \geq |\mathcal{A}_1|^{\frac{101}{64}},\end{aligned}$$

so $|\mathcal{A}_1| \ll (\log N)^{\frac{64}{101}}$. Similarly, we get

$$\begin{aligned}|\mathcal{A}_2 + \mathcal{A}_3 + \mathcal{A}_4| &\geq |\mathcal{A}_2|^{\frac{1}{2}} |\mathcal{A}_3 + \mathcal{A}_4|^{\frac{3}{4}} \\ &\geq |\mathcal{A}_2|^{\frac{1}{2}} |\mathcal{A}_3|^{\frac{3}{8}} |\mathcal{A}_4|^{\frac{9}{16}} \geq |\mathcal{A}_2|^{\frac{23}{16}},\end{aligned}$$

so $|\mathcal{A}_2| \ll (\log N)^{\frac{16}{23}}$.

The second claim follows (as before) by Lemma 5.4 applied to $\mathcal{A}_2+\mathcal{A}_3+\mathcal{A}_4+\mathcal{A}_5$ and $\mathcal{A}_3+\mathcal{A}_4+\mathcal{A}_5$, respectively. □

6.2. **High iterated sumsets for 5-progression-free cases.** For sumsets being 3-progression-free the size of the sumset is a product of the sizes of the summands. Therefore, let us first study the ternary case avoiding progressions of length $2k+1=5$ in detail.

Let $|\mathcal{A}|=T^\alpha, |\mathcal{B}|=T^\beta, |\mathcal{C}|=T^\gamma$ where α,β,γ are positive real constants. Suppose that $|\mathcal{A}|\le|\mathcal{B}|\le|\mathcal{C}|$.

$$|\mathcal{A}+(\mathcal{B}+\mathcal{C})|\ge T^{\alpha/2}T^{(3/4)(3\gamma/4+\beta/2)}=T^{\alpha/2+3\beta/8+9\gamma/16}. \tag{1}$$

$$|\mathcal{B}+(\mathcal{A}+\mathcal{C})|\ge T^{\beta/2}T^{(3/4)(3\gamma/4+\alpha/2)}=T^{3\alpha/8+\beta/2+9\gamma/16}. \tag{2}$$

$$|\mathcal{C}+(\mathcal{A}+\mathcal{B})|\ge T^{\gamma/2}T^{(3/4)(3\beta/4+\alpha/2)}=T^{3\alpha/8+9\beta/16+\gamma/2}. \tag{3}$$

$$|\mathcal{C}+(\mathcal{A}+\mathcal{B})|\ge T^{3\gamma/4}T^{(1/2)(3\beta/4+\alpha/2)}=T^{\alpha/4+3\beta/8+3\gamma/4}. \tag{4}$$

(2) is clearly superior to (1), as the larger parameter β gets the larger weight. And (2) is also superior to (3). Hence one has

$$|\mathcal{A}+\mathcal{B}+\mathcal{C}|\ge T^{\max(3\alpha/8+\beta/2+9\gamma/16,\alpha/4+3\beta/8+3\gamma/4)}.$$

If γ dominates, then equation (4) is the strongest inequality. But let us also compare this when all three sets are of equal size $|\mathcal{A}|=|\mathcal{B}|=|\mathcal{C}|=T$. Then equations (1), (2) and (3) give $T^{23/16}$, whereas equation (4) only gives $T^{11/8}$.

A similar study can be done for quaternary or higher sumsets. As before, let $|\mathcal{A}|=T^\alpha, |\mathcal{B}|=T^\beta, |\mathcal{C}|=T^\gamma$ and let $|\mathcal{D}|=T^\delta$. In the case where all 4 sets are of equal size T one achieves

$$|\mathcal{D}+(\mathcal{C}+(\mathcal{A}+\mathcal{B}))|\ge T^{3(3(\alpha/2+3\beta/4)/4+\gamma/2)/4+\delta/2}=T^{9\alpha/32+27\beta/64+3\gamma/8+\delta/2}=T^{101/64},$$

whereas

$$|(\mathcal{D}+\mathcal{C})+(\mathcal{A}+\mathcal{B}))|\ge T^{3(3\delta/4+\gamma/2)/4+(3\beta/4+\alpha/2)/2}=T^{\alpha/4+3\beta/8+3\gamma/8+9\delta/16}=T^{25/16}.$$

If the sets are not of equal size and if δ dominates, one can, for example, achieve:

$$|\mathcal{D}+(\mathcal{C}+(\mathcal{A}+\mathcal{B}))|\ge T^{\alpha/8+3\beta/16+3\gamma/8+3\delta/4}.$$

For higher iterations of 5-progression-free sums of sets of equal sizes, it is not obvious in which order to bracket and add to achieve the best possible result. Bracketing according to the pattern $(((\mathcal{A}_1+\mathcal{A}_2)+\mathcal{A}_3)+A_4)\dots$ gives an exponent $1/2+3/4\,(1/2+3/4\,(1/2+3/4))\dots$ tending to the fixed point of $\alpha = 1/2 + 3\alpha/4$, i.e., $\alpha = 2$; this is far from optimal. On the other hand, bracketing in sizes of 2,4,8 etc.

$$((\mathcal{A}_1 + \mathcal{A}_2) + (\mathcal{A}_3 + A_4)) + ((\mathcal{A}_5 + \mathcal{A}_6) + (\mathcal{A}_7 + \mathcal{A}_8)) + (\dots(\dots + \mathcal{A}_r))$$

gives unbounded exponents like $(5/4)^n$, if $r = 2^n$. Let $e_k(r)$ denote the largest exponent from these different bracketing of r sumsets of equal size, avoiding progressions of length $2k+1$. Hence $e_2(2) = \frac{5}{4}$ and $e_2(3) = \frac{23}{16}$. This implies $e_2(2^n) \geq 1.25^n$ and $e_2(3^n) \geq (23/16)^n$. This last case of summing very many sets of equal size in brackets of three of brackets of three sets etc. is asymptotically slightly better than the case of powers of 2 as $(23/16)^{1/\log 3} \approx 1.39143 \ > \ (5/4)^{1/\log 2} \approx 1.37979$.

A measure to compare asymptotically different bracketing methods is based on comparing $e_2(r)^{1/\log(r)}$. With a computer program we listed the best exponents up to 150,000 summands, and a few individual ones. This shows that for higher exponents of 2 or 3 non-negligible improvements over the estimates by lifting from small values (as above) can be achieved:
$e_2(3^8) \geq 19.0587$, $(e_2(3^4))^2 \geq 18.8707$, $e_2(3^2)^4 \geq 18.5454$, $(e_2(3))^8 \geq 18.2332$, $e_2(3^{11})/(e_2(3))^{11} \approx 1.06641$, and similarly
$e_2(2^{16}) \geq 41.3428$,$(e_2(2^8))^2 \geq 40.9146$, $(e_2(2^4))^4 \geq 39.9085$, $(e_2(2^2))^8 \geq 38.4708$, $(e_2(2))^{16} \geq 35.5271$, $e_2(2^{17})/e_2(2)^{17} \approx 1.17576$.

Hence one may wonder if there is an efficient method to achieve close to the best possible values of $e_2(r)$ for large values of r. One might try to take the asymmetric weighting of the two summands into account. Empirically, it seems close to the best possible if the smaller set collects about $0.35r$ of the summands, and the larger set the remaining about $0.65r$ summands. (Note that the maximum of $\frac{1}{2}\log r_1 + \frac{3}{4}\log(r - r_1)$ is at $r_1 = 0.4r$, and near the maximum there is a rather wide plateau). The Fibonacci numbers are an easy sequence which correspond to a weighting of about $0.382n$ and $0.618n$. The first values up to $r = 21$ generated by a Fibonacci-bracketing (with sizes of number of summands being (2,1,2,3,5,8 ...)

$$(((((\mathcal{A}_1{+}\mathcal{A}_2){+}\mathcal{A}_3){+}(\mathcal{A}_4{+}\mathcal{A}_5)){+}(\mathcal{A}_6{+}\mathcal{A}_7{+}\mathcal{A}_8)){+}(\mathcal{A}_9{+}\mathcal{A}_{10}{+}\cdots{+}\mathcal{A}_{13})){+}\cdots,$$

are indeed the best possible, and later they are still very close to the best possible. One gets, for example,

$$e_2(2) = \frac{5}{4}, e_2(3) = \frac{23}{16}, e_2(5) = \frac{109}{64}, e_2(8) = \frac{511}{256}, e_2(13) = \frac{2405}{1024}.$$

Note, when $r = 8$, the bracketing in powers of 2 would only give an exponent $(5/4)^3 = 500/256 < 511/256$. The quotient of the optimal values of $e_2(F_i)$ and the simple approximation by a Fibonacci bracketing (which only needs $O(\log F_i)$ additions and multiplictions) is up to $F_i = 196418$ at most 1.0058, i.e., much smaller than the corresponding quotients for the bracketing with powers of 2 or 3 (see above).

Numerically, it seems possible that $(e_2(r))^{1/\log r}$ tends to a limit near 1.4.

7. Acknowledgements

The authors would like to thank the referee and Ilya Shkredov for comments on the manuscript.

C. E. is supported by a joint FWF-ANR project ArithRand, grant numbers FWF I 4945-N and ANR-20-CE91-0006.

References

1. N. Alon, R. Alweiss, Y. P. Liu, A. Martinsson, and S. Narayanan, *Arithmetic progressions in sumsets of sparse sets*, Integers **21A** (2021), Paper No. A3, 7. MR 4304359
2. L. Becker, P. Ivanisvili, D. Krachun, and J. Madrid, *Discrete Brunn-Minkowski inequality for subsets of the cube*, 2024.
3. T. F. Bloom and O. Sisask, *An improvement to the Kelley-Meka bounds on three-term arithmetic progressions*, 2023.
4. J. Bourgain, *On arithmetic progressions in sums of sets of integers*, A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 105–109. MR 1117007
5. J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova, *Explicit constructions of RIP matrices and related problems*, Duke Math. J. **159** (2011), no. 1, 145–185. MR 2817651
6. E. Croot, I. Z. Ruzsa, and T. Schoen, *Arithmetic progressions in sparse sumsets*, Combinatorial Number Theory (Walter de Gruyter, Berlin, 2007) (2007), 157–164.
7. H. Darmon and L. Merel, *Winding quotients and some variants of Fermat's last theorem*, J. Reine Angew. Math. **490** (1997), 81–100. MR 1468926
8. L. E. Dickson, *History of the theory of numbers. Vol. II: Diophantine analysis*, Chelsea Publishing Co., New York, 1966. MR 245500
9. R. Dietmann and C. Elsholtz, *Hilbert cubes in progression-free sets and in the set of squares*, Israel J. Math. **192** (2012), no. 1, 59–66. MR 3004074
10. ______, *Hilbert cubes in arithmetic sets*, Rev. Mat. Iberoam. **31** (2015), no. 4, 1477–1498. MR 3438398
11. G. Elekes, M. B. Nathanson, and I. Z. Ruzsa, *Convexity and sumsets*, J. Number Theory **83** (2000), no. 2, 194–201. MR 1772612
12. C. Elsholtz, Z. Hunter, L. Proske, and L. Sauermann, *Improving Behrend's construction: Sets without arithmetic progressions in integers and over finite fields*, (2024).
13. C. Elsholtz and L. Wurzinger, *Sumsets in the set of squares*, The Quarterly Journal of Mathematics **75** (2024), 1243–1254.
14. P. Erdős and P. Turán, *On some sequences of integers.*, J. Lond. Math. Soc. **11** (1936), 261–264.
15. G. A. Freiman, H. Halberstam, and I. Z. Ruzsa, *Integer sum sets containing long arithmetic progressions*, J. London Math. Soc. (2) **46** (1992), no. 2, 193–201. MR 1182477
16. B. Green, *Arithmetic progressions in sumsets*, Geom. Funct. Anal. **12** (2002), no. 3, 584–597. MR 1924373
17. K. Gyarmati, *On a problem of Diophantus*, Acta Arithmetica **97** (2001), 53–65.
18. D Hajela and P Seymour, *Counting points in hypercubes and convolution measure algebras*, Combinatorica **5** (1985), no. 3, 205–214.

19. K. Henriot, *On arithmetic progressions in $A+B+C$*, Int. Math. Res. Not. IMRN (2014), no. 18, 5134–5164. MR 3264678
20. ———, *Arithmetic progressions in sets of small doubling*, Mathematika **62** (2016), no. 2, 587–613. MR 3521343
21. Z. Kelley and R. Meka, *Strong bounds for 3-progressions*, 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS) (2023), 933–973.
22. D. Matolcsi, I. Z. Ruzsa, G. Shakan, and D. Zhelezov, *An analytic approach to cardinalities of sumsets*, Combinatorica **42** (2022), no. 2, 203–236. MR 4426299
23. I. Z. Ruzsa, *Arithmetic progressions in sumsets*, Acta Arith. **60** (1991), no. 2, 191–202. MR 1139055
24. ———, *Arithmetical progressions and the number of sums*, Period. Math. Hungar. **25** (1992), no. 1, 105–111. MR 1200845
25. ———, *Erdős and the integers*, J. Number Theory **79** (1999), no. 1, 115–163. MR 1724257
26. ———, *Sumsets and structure*, Combinatorial number theory and additive group theory, Adv. Courses Math. CRM Barcelona, Birkhäuser Verlag, Basel, 2009, pp. 87–210. MR 2522038
27. I. Z. Ruzsa and J. Solymosi, *Sumsets of semiconvex sets*, Canad. Math. Bull. **65** (2022), no. 1, 84–94. MR 4396625
28. R. Salem and D. C. Spencer, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U.S.A. **28** (1942), 561–563. MR 7405
29. T. Sanders, *Three-term arithmetic progressions and sumsets*, Proc. Edinb. Math. Soc. (2) **52** (2009), no. 1, 211–233. MR 2475890
30. T. Schoen, *A note on sumsets of progression-free sets, manuscript.*
31. Y. V. Stanchescu, *Planar sets containing no three collinear points and non-averaging sets of integers*, Discrete Math. **256** (2002), no. 1-2, 387–395. MR 1927560
32. E. Szemerédi and V. H. Vu, *Finite and infinite arithmetic progressions in sumsets*, Ann. of Math. (2) **163** (2006), no. 1, 1–35. MR 2195131
33. D. R. Woodall, *A theorem on cubes*, Mathematika **24** (1977), no. 1, 60–62.

Institut für Analysis und Zahlentheorie, Technische Universität Graz, Kopernikusgasse 24, A-8010 Graz, Austria
Email address: elsholtz@math.tugraz.at

Alfréd Rényi Institute of Mathematics, Budapest, Pf. 127, H-1364 Hungary
Email address: ruzsa@renyi.hu

Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria
Email address: lena.wurzinger@ist.ac.at