ON A QUESTION OF ERDŐS AND MOSER

B. SUDAKOV, E. SZEMERÉDI, and V. H. VU

Abstract

For two finite sets of real numbers A and B, one says that B is sum-free with respect to A if the sum set $\{b + b' \mid b, b' \in B, b \neq b'\}$ is disjoint from A. Forty years ago, Erdős and Moser posed the following question. Let A be a set of n real numbers. What is the size of the largest subset B of A which is sum-free with respect to A?

In this paper, we show that any set A of n real numbers contains a set B of cardinality at least $g(n) \ln n$ which is sum-free with respect to A, where $g(n)$ tends to infinity with n. This improves earlier bounds of Klarner, Choi, and Ruzsa and is the first superlogarithmic bound for this problem.

Our proof combines tools from graph theory together with several fundamental results in additive number theory such as Freiman's inverse theorem, the Balog-Szemerédi theorem, and Szemerédi's result on long arithmetic progressions. In fact, in order to obtain an explicit bound on $g(n)$, we use the recent versions of these results, obtained by Chang and by Gowers, where significant quantitative improvements have been achieved.

1. Introduction

For two finite sets of numbers A and B, we denote by $A + B$ the sum set $\{a + b \mid a \in A, b \in B\}$. A closely related notion is $A \overset{*}{+} B$, which denotes the set $\{a + b \mid a \in A, b \in B, a \neq b\}$. For convenience, we write $2A$ for $A + A$. In general, $lA = (l-1)A + A$. Similarly, we write 2^*A for $A \overset{*}{+} A$ and l^*A for the collection of sums of l-distinct elements of A.

Forty years ago, Erdős and Moser [5] (see also [10, Problem C14]) posed the following question. Let A be a set of n real numbers. What is the size of the largest subset B of A such that 2^*B is disjoint from A, that is, no element of A can be represented as a sum of two distinct elements of B? We call such a set B *sum-free*

DUKE MATHEMATICAL JOURNAL
Vol. 129, No. 1, © 2005
Received 12 March 2004. Revision received 28 September 2004.
2000 *Mathematics Subject Classification*. Primary 11P70; Secondary 11B75.
Sudakov's research supported in part by National Science Foundation grant numbers DMS-0355497 and DMS-0106589 and by an Alfred P. Sloan fellowship.
Vu's research supported in part by National Science Foundation grant number DMS-0200357, a National Science Foundation CAREER award, and an Alfred P. Sloan fellowship.

with respect to A. Note that in order to obtain a nontrivial result, it is necessary to consider sums of distinct elements of B. Indeed, if A is a set of consecutive powers of two, then the largest subset B of A such that $2B$ is disjoint from A has exactly one element.

Denote by $\phi(A)$ the maximum cardinality of a subset of A which is sum-free with respect to A. Let $\phi(n)$ be the minimum of $\phi(A)$ over all sets A of n real numbers. Erdős and Moser [5] showed that $\phi(n) \le n/3$ and suggested that it probably has order $o(n)$. The first improvement over the Erdős and Moser result was due to Selfridge (see [5]), who showed that $\phi(n) \le n/4$. Choi [4], using sieve methods, proved that $\phi(n) \le O(n^{2/5+\epsilon})$, where ϵ is an arbitrarily small positive constant. He also noted that in this problem, it suffices to consider the special case when A is a set of integers. Choi's result was slightly improved by Baltz, Schoen, and Srivastav [2], who showed that $\phi(n) \le O(n^{2/5} \ln^{2/5} n)$. A huge improvement of the upper bound was very recently obtained by Ruzsa [13], who proved that

$$\phi(n) = e^{O(\sqrt{\ln n})}.$$

In the following, we describe Ruzsa's construction, which, besides being very clever, is short and instructive.

Let $d = \Theta(\sqrt{\ln n})$. It is enough to construct a set $A \subset \mathbb{Z}^d$ such that $|A| > n$ and $\phi(A) \le e^{O(\sqrt{\ln n})}$. Then A can be mapped into $\mathbb{Z}$ using a projection $p(x_1, \dots, x_d) = x_1 + mx_2 + \cdots + m^{d-1}x_d$ with suitably large m. Let

$$B_r = \left\{ (x_1, \dots, x_d) \in \mathbb{Z}^d \;\middle|\; \sum x_i^2 \le r \right\},$$

where B_r is the set of integral lattice points in the ball of radius r centered at the origin. Let $r = e^{O(\sqrt{\ln n})}$, and let $A = \bigcup_{i=0}^{r-1} 2^i \times B_{r-i}$, where $k \times X = \{kx \mid x \in X\}$. For an appropriate choice of d and r, one can make $|A| > n$, and we claim that $\phi(A) \le 2^d r = e^{O(\sqrt{\ln n})}$. Indeed, let $S \subset A$ be of size $> 2^d r$. Then there is an index i such that the set $S_i = S \cap 2^i \times B_{r-i}$ has size $> 2^d$. Note that $i < r - 1$ since $|B_1| = 2d + 1 < 2^d$. By the pigeonhole principle, one can find two vectors $b', b'' \in B_{r-i}$ whose coordinates are congruent modulo 2 such that $2^i b', 2^i b'' \in S_i$. Then the point $b = (b' + b'')/2$ has integer coordinates, and one can check that $\|b\|^2 \le r - i - 1$. Hence $2^i b' + 2^i b'' = 2^{i+1} b$ is an element of the set $2^{i+1} \times B_{r-i-1} \subset A$, a contradiction.

Let us now turn our attention to the lower bound. It was remarked by Klarner (unpublished) and mentioned by Erdős in [5] that $\phi(n) = \Omega(\ln n)$. The first published proof of this bound appeared about ten years later in Choi's paper [4]. Choi proved that $\phi(n) \ge \log_2 n$. Recently, Ruzsa [13] improved this result slightly by showing that $\phi(n) > 2 \log_3 n - 1$.

A natural way to prove a lower bound is to use a greedy-type argument. Technically speaking, one would try to construct the desired set B element by element via

a greedy procedure. This was the approach behind both Choi's and Ruzsa's proofs. Quite interestingly, Ruzsa also showed that $\ln n$ is the limit of this natural method.

Taking into account the above results and the fact that the upper bound on $\phi(n)$ is $n^{o(1)}$, one might suspect that the true order of magnitude of $\phi(n)$ is indeed $\Theta(\ln n)$. However, we show that this is not the case by proving the first superlogarithmic bound for $\phi(n)$.

THEOREM 1.1
There is a function $g(n)$ tending to infinity with n such that the following holds. Any set A of n integers contains a subset B with cardinality $g(n)\ln n$ such that B is sum-free with respect to A.

Our proof of Theorem 1.1 shows that one can take $g(n)$ to be of order $(\ln^{(5)} n)^{1-o(1)}$, where $\ln^{(i)} x$ denotes iterative logarithm that is defined by

$$\ln^{(1)} x = \ln x, \qquad \ln^{(i+1)} x = \ln(\ln^{(i)} x).$$

It is simpler to describe $g(n)$ as the inverse of an iterative exponential function. To do so, we use the notation $p \uparrow q$ for p^q with the obvious convention for bracketing so that $p \uparrow q \uparrow r$ stands for $p \uparrow (q \uparrow r)$. Let

$$F(h) = \exp\left(h^{182} \cdot (2 \uparrow (e^{h^{32770}}) \uparrow 2 \uparrow 2 \uparrow (2h+9))\right).$$

We can set $g(n)$ in Theorem 1.1 to equal $c(m/\ln m)$, where $m = F^{-1}(n^{1/2})$ and c is a sufficiently small positive constant. The main ingredient in the proof of Theorem 1.1 is the following result.

THEOREM 1.2
Let X and Y be two finite sets of positive integers with

$$\frac{1}{h^{29}}|Y| \geq |X| \geq F(h),$$

where $F(h)$ is defined as above and h is a sufficiently large integer. Then Y contains a subset Z of size h which is disjoint from X and is sum-free with respect to the union $X \cup Y$.

In Section 2 we deduce Theorem 1.1 from Theorem 1.2. Most of the remaining part of the paper is devoted to the proof of Theorem 1.2. This proof combines several fundamental results in additive number theory such as Freiman's inverse theorem (see [7]), the Balog-Szemerédi result on a statistical version of Freiman's theorem (see [1]), and Szemerédi's theorem on long arithmetic progressions (see [14]). In fact, in order to

obtain the claimed explicit bound on $g(n)$, we use the most recent versions of these results, where significant quantitative improvements have been achieved. In particular, we use Chang's version of Freiman's inverse theorem (see [3]), Gowers's version of the Balog-Szemerédi result (see [8]), and Gowers's quantitative bound on long arithmetic progressions (see [9]). After Gowers's recent and deep work on long arithmetic progressions (see [8], [9]), it has become known that these results are highly connected. On the other hand, it is interesting to see all of them used in the same proof. The crucial additional tool, which links the above results together, is a multiplicative property of generalized arithmetic progressions, described in Lemma 5.2. Finally, we also invoke a few standard facts and arguments from graph theory.

Both Theorems 1.1 and 1.2 can be generalized as follows. A set B is called *k-sum-free* with respect to a set A if, for any $2 \le l \le k$, the sum of any l different elements of B is not in A. In other words, the union $\bigcup_{l=2}^{k} l^* B$ is disjoint from A.

THEOREM 1.3

For every fixed positive integer $k \ge 2$, there exists a function $g_k(n)$ tending to infinity with n such that the following holds. Any set A of n integers contains a subset B with cardinality $g_k(n) \ln n$ such that B is k-sum-free with respect to A.

Our proof shows that $g_k(n)$ is roughly the same as $g(n)$ for all fixed k. Let a_k be a sufficiently large constant, and let $M(k)$ be the least common multiple of $2, \dots, k$. Set

$$F_k(h) = \exp\left(h^{a_k} \cdot (2 \uparrow (e^{h^{a_k}}) \uparrow 2 \uparrow 2 \uparrow (M(k)h + 9))\right).$$

We can set $g_k(n)$ in Theorem 1.3 to equal $c(m/\ln m)$, where $m = F_k^{-1}(n^{1/2})$ and c is a sufficiently small positive constant.

THEOREM 1.4

For any positive integer k, there is a positive constant c_k such that the following statement holds for every sufficiently large integer h. Let X and Y be two finite sets of integers where

$$\frac{1}{h^{c_k}}|Y| \ge |X| \ge F_k(h),$$

where $F_k(h)$ is defined as above. Then Y contains a subset Z of size h which is disjoint from X and is k-sum-free with respect to the union $X \cup Y$.

One can deduce Theorem 1.3 from Theorem 1.4 in exactly the same way that one deduces Theorem 1.1 from Theorem 1.2. However, to prove Theorem 1.4, one needs to generalize several tools that were used in the proof of Theorem 1.2. Some of these generalizations are not entirely straightforward and might be interesting in their own right.

The rest of the paper is organized as follows. In Section 2 we deduce Theorem 1.1 from Theorem 1.2. In Sections 3, 4, and 5 we present the necessary tools for proving Theorem 1.2. Section 3 contains a tool from graph theory, Chang's version of Freiman's inverse theorem, and Gowers's result on long arithmetic progressions. In Section 4 we discuss several statistical versions of Freiman's theorem, including the original result of Balog and Szemerédi, a recent refinement of Gowers, and an extension of these results. Section 5 is devoted to a multiplicative property of sets with additive structure. The main result of this section states that if a set X (statistically) resembles a generalized arithmetic progression, then it contains many elements x such that $2x$ does not belong to X. The proof of Theorem 1.2 comes in Section 6. In Section 7 we generalize several tools to prove Theorem 1.4. The last section, Section 8, is devoted to concluding remarks.

2. Proof of Theorem 1.1

In this section, we show how to deduce Theorem 1.1 from Theorem 1.2. Throughout the proof, the asymptotic notation is used under the assumption that $h \to \infty$. We omit unnecessary floor and ceiling signs for the sake of clarity.

Without loss of generality, we can assume that all elements of A are positive. Indeed, A contains at least $n/2$ elements with the same sign, and we do not care about constant factors. Let $a_1 > a_2 > \cdots > a_n > 0$ be the elements of A in decreasing order, and let h be the largest integer such that $F(h) \le n^{1/2}$. Denote by A_0 the set of the first $F(h)$ elements of A; that is, $A_0 = \{a_1, \dots, a_{F(h)}\}$, A_1 is the set of the first $2h^{30}F(h)$ elements of A, A_2 is the set of the first $(2h^{30})^2F(h)$ elements of A, and so on. In general, A_i, $0 \le i \le T$, is the set of the first $(2h^{30})^i F(h)$ elements of A, where T is the largest integer satisfying

$$(2h^{30})^T F(h) \le n.$$

As we choose h such that $F(h) \le n^{1/2}$, we have

$$T = \Omega\Big(\frac{\ln n}{\ln h}\Big).$$

Applying Theorem 1.2 to $X = A_0$ and $Y = A_1$, we obtain a subset B_1 of A_1 of cardinality h which is sum-free with respect to $A_0 \cup A_1 = A_1$. It is important to notice that since the elements in A are in decreasing order, B_1 is sum-free with respect to A.

Let A_2' be the subset of A_2 obtained by deleting from A_2 the set A_1 and also all elements that can be represented as the difference of an element from A_1 and an element from B_1. Using the facts that A_2 has $(2h^{30})^2F(h)$ elements, A_1 has $(2h^{30})F(h)$ elements, and B_1 has h elements, we conclude that A_2' has at least

$$(2h^{30})^2F(h) - (2h^{30})hF(h) - (2h^{30})F(h) \ge 3h^{60}F(h)$$

elements. To apply Theorem 1.2 to $X = A_2 \backslash A_2'$ and $Y = A_2'$, we first need to check that $(1/h^{29})|Y| \geq |X|$. This follows from the fact that X has at most

$$(2h^{30})hF(h) + (2h^{30})F(h) \leq 3h^{31}F(h)$$

elements. The other condition, $|X| \geq F(h)$, holds trivially as X contains A_1. Thus, by Theorem 1.2, there exists a subset B_2 of A_2' of cardinality h which is sum-free with respect to $X \cup Y = A_2$. We next claim that the union $B_1 \cup B_2$ is sum-free with respect to A_2.

To verify this claim, notice that, by the definition of A_2', the sum of an element of B_1 and an element of B_2 is not in A_1. By the properties of the ordering and the fact that all elements are positive, any element in B_1 is larger than any element in $A_2 \backslash A_1$. Therefore the sum of an element in A_1 and an element in A_2 cannot be in $A_2 \backslash A_1$. Furthermore, the sum of two elements in B_1 is not in A_1, as B_1 is sum-free with respect to A_1. This sum is not in $A_2 \backslash A_1$ either, thanks to the ordering. So we can conclude that the union $B_1 \cup B_2$ is sum-free with respect to A_2. Again by the ordering, we can have a stronger statement that $B_1 \cup B_2$ is sum-free with respect to A.

By iterating this argument, we obtain disjoint subsets $B_1, \ldots, B_T$; each has cardinality h such that the union $\bigcup_{i=1}^{T} B_i$ is sum-free with respect to A. Since

$$T = \Omega\Big(\frac{\ln n}{\ln h}\Big),$$

this union has

$$h \cdot T \geq c\frac{\ln n}{\ln h}\, h$$

elements, where c is a positive constant. Set $g(n) = c(h/\ln h)$. The definition of h implies that h is approximately $F^{-1}(n^{1/2})$, which tends to infinity with n. Thus $g(n)$ tends to infinity with n. The union $B = \bigcup_{i=1}^{T} B_i$ has at least $g(n) \ln n$ elements and is sum-free with respect to A.

The only technical point in the above iteration is the verification of the condition (of Theorem 1.2) that $(1/h^{29})|Y| \geq |X|$ at each step. We discuss this point in more detail. Assume that $B_1, \ldots, B_i$ have been found. We define A_{i+1}' as the set obtained from A_{i+1} by deleting A_i and all elements that can be represented as the difference of an element of A_i and an element in $\bigcup_{j=1}^{i} B_j$. Notice that if $a \in A_i$ and $b \in B_j$ satisfy $a - b \in A_{i+1}$, then $a > b$; and so a must be an element of A_j. Thus the number of elements which can be represented as the difference of an element of A_i and an element in $\bigcup_{j=1}^{i} B_j$ is upper bounded by

$$\begin{aligned}\sum_{j=1}^{i} |A_j||B_j| &= \sum_{j=1}^{i} h(2h^{30})^j F(h) \\ &< 1.1h(2h^{30})^i F(h).\end{aligned}$$

Setting $X = A_{i+1} \setminus A'_{i+1}$ and $Y = A'_{i+1}$, we have

$$\begin{aligned} |X| &\le 1.1h(2h^{30})^i F(h) + |A_i| \\ &= 1.1h(2h^{30})^i F(h) + (2h^{30})^i F(h) \\ &\le 1.2h(2h^{30})^i F(h) \end{aligned}$$

and

$$\begin{aligned} |Y| &= |A_{i+1}| - |X| \\ &\ge (2h^{30})^{i+1} F(h) - 1.2h(2h^{30})^i F(h) \\ &> 1.9h^{30}(2h^{30})^i F(h). \end{aligned}$$

The above two inequalities guarantee that

$$\frac{1}{h^{29}}|Y| \ge |X|,$$

as required. This concludes the proof. □

3. Some tools

The goal of this section is to provide various results that we need for the proof of Theorem 1.2. The first is a statement about graphs. The second is Gowers's result on long arithmetic progression. (In fact, we are going to use a corollary of this result.) The third is Chang's quantitative version of Freiman's theorem.

3.1. Independent sets in graphs

A graph $G = G(V, E)$ consists of a vertex set V and an edge set E, where each edge is an unordered pair of vertices. We write (a, b) for the edge between a and b and say that a is *adjacent* to b. The *neighborhood* $N(a)$ of a vertex a is a set of all vertices adjacent to a, and $|N(a)|$ is called the *degree* of a. A subset I of V is an *independent* set if it does not contain an edge; that is, there is no $a, b \in I$ such that $(a, b) \in E$. We need the following lemma.

LEMMA 3.1

Let h be a positive integer, and let $G(V, E)$ be a graph such that $|E| \le |V|^2/(3h)$ and $|V| \ge 3h$. Then G contains an independent set of size h.

Proof

This lemma is an easy corollary of a well-known theorem in graph theory which asserts that the size of the maximum independent set is at least $|V|/(d + 1)$, where d is the average degree of the graph. The sum of the degrees is twice the number of

edges, so the first assumption of the lemma implies that the average degree is at most $(2|V|^2/(3h))/|V| = 2|V|/(3h)$. Therefore the size of the largest independent set is at least

$$\frac{|V|}{2|V|/(3h)+1} \geq \frac{|V|}{|V|/h} = h,$$

due to the assumption $|V| \geq 3h$. □

The following notions are needed later. A graph G is *bipartite* if its vertex set V can be partitioned into two sets V_1 and V_2 such that every edge has one end in V_1 and the other end in V_2. We call V_1 and V_2 the *color classes* of G.

A hypergraph H consists of a vertex V and a family of subsets of V, which we call *edges*. We say that H is *k-uniform* if every edge of H has size k. A k-uniform hypergraph H is *k-partite* if we can partition the vertex set V into k sets $V_1, \ldots, V_k$ such that every edge of H contains exactly one vertex from V_i for all $1 \leq i \leq k$. The sets $V_1, \ldots, V_k$ are the color classes of H.

3.2. Arithmetic progressions in dense sets

Another result that we need in our proof is a corollary of a well-known theorem of Szemerédi [14] which asserts that every dense subset of $\{1, \ldots, n\}$ must contain long arithmetic progression. In order to obtain the best possible bound in our main result, we use the following remarkable quantitative version of Szemerédi's theorem proved by Gowers [9].

THEOREM 3.2

Let $0 < \delta \leq 1/2$, and let k be a positive integer. Let

$$n \geq 2 \uparrow 2 \uparrow (\delta)^{-1} \uparrow 2 \uparrow 2 \uparrow (k+9),$$

and let A be a subset of an arithmetic progression of length n with cardinality at least δn. Then A contains an arithmetic progression of length k.

COROLLARY 3.3

Let A be a subset of density δ of an arithmetic progression of length n, where $0 < \delta \leq 1/2$ and

$$n \geq 2 \uparrow 2 \uparrow (\delta)^{-1} \uparrow 2 \uparrow 2 \uparrow (2k+9).$$

Then A contains a subset A' of k elements such that for any two elements x, y of A', there is an element z of A satisfying $x + y = 2z$.

Proof

By Theorem 3.2, A contains an arithmetic progression P of length $2k$. Take A' to be

the set consisting of the elements with even indices of P. □

Remark. In Corollary 3.3, one can have $2k + 8$ instead of $2k + 9$ by taking the set of elements with odd indices. We prefer to have the current proof, as it is consistent with the proof of a more general statement presented later in the paper.

3.3. Sumsets and Freiman's theorem

Let A be a finite set of integers with the property that the size of $A + A$ is not much larger than the size of A. What information does this give about A? This problem is called an *inverse* problem in additive number theory since one wants to deduce some information about the structure of A knowing something about the sumset $A+A$. This is in contrast with *direct* problems, where properties of A are used to get information about $A + A$.

Let $x_0, \ldots, x_d$ and $m_1 \geq m_2 \geq \cdots \geq m_d > 0$ be integers. The set

$$P = \left\{ x_0 + \sum_{j=1}^{d} \lambda_j x_j \;\middle|\; 0 \leq \lambda_j \leq m_j - 1 \right\}$$

is called a *generalized* arithmetic progression of dimension d. The product $m_1 m_2 \cdots m_d$ is called the *volume* of P. It is clear that the volume of P is always at least its cardinality. If the two quantities are equal, then P is *proper*. It is important that when we consider a generalized arithmetic progression, we talk about a specific representation.

If P is a generalized arithmetic progression of dimension d, then $2P$ is a generalized arithmetic progression of dimension d whose volume is 2^d times the volume of P. It follows that if A is a subset of a d-dimensional arithmetic progression P with volume $C|A|$, then $|A + A| \leq 2^d C|A|$. A fundamental result in additive number theory, proved by Freiman [7], tells us that this is basically the only example of sets with small sumsets.

We use the following quantitative version of Freiman's theorem obtained by Chang [3] (using an earlier approach of Ruzsa [12]).

THEOREM 3.4

Let A be a finite set of integers such that $|A + A| \leq C|A|$ for some positive number C. Then there exists a proper generalized arithmetic progression P of dimension at most C and volume at most $e^{O(C^2 \ln^3 C)}|A|$ which contains A.

Here C may be a function of $|A|$. The constant in O does not depend on C or on $|A|$.

4. Statistical versions of Freiman's theorem

In many applications, instead of the sumset $A + A$, one has access only to a dense subset of $A + A$. Even in this case, one is able to draw a useful conclusion thanks to a result of Balog and Szemerédi [1]. In order to describe this result, we first need a definition.

Let A and B be two sets of integers. Let $G = G(A, B, E)$ be a bipartite graph whose color classes are A and B and whose edge set is E (an *edge* is a pair (a, b), where $a \in A$ and $b \in B$). We denote by $A +_G B$ the collection of the sums $a + b$, where $a \in A$, $b \in B$, and $(a, b) \in E$.

Balog and Szemerédi [1] proved that if A and B are two sets of cardinality n and $|E| \geq n^2/K$ and $|A +_G B| \leq Cn$, where K and C are positive constants not depending on n, then one can find $A' \subset A$ and $B' \subset B$ such that $|A'| \geq n/K'$, $|B'| \geq n/K'$, and $|A' + B'| \leq C'n$, where K' and C' are constants depending on K and C but not on n.

With a new proof, Gowers [8] has recently strengthened this statement by allowing K and C to be functions of n. He shows that both K' and C' can be bounded by polynomials with fixed degrees in K and C. The polynomials in Gowers's proof were implicit, but by following his ideas, one can work out the explicit version below.

THEOREM 4.1

Let n, C, K be positive numbers, and let A and B be two sets of n integers. Suppose that there is a bipartite graph $G(A, B, E)$ with at least n^2/K edges and $|A +_G B| \leq Cn$. Then one can find a subset $A' \subset A$ and a subset $B' \subset B$ such that $|A'| \geq n/(16K^2)$, $|B'| \geq n/(4K)$, and $|A' + B'| \leq 2^{12}C^3K^5n$.

As the proof is relatively short, we include it. This proof is slightly different from Gowers's original proof. The heart of the proof is the following graph-theoretical lemma, which is of independent interest.

LEMMA 4.2

Let n and K be positive numbers, and let $G = G(A, B, E)$ be a bipartite graph, where $|B| \leq |A| = n$ and $|E| = n^2/K$. Then one can find $A' \subset A$ and $B' \subset B$ such that $|A'| \geq n/(16K^2)$, $|B'| \geq n/(4K)$, and for each $a \in A'$ and $b \in B'$, there are $n^2/(2^{12}K^5)$ paths of length 3 whose two end points are a and b.

Using this lemma, we can generalize Theorem 4.1 to sums of more than two sets. In the rest of this section, we first use Lemma 4.2 to prove Theorem 4.1. Next, we prove Lemma 4.2. Finally, we present and prove the generalization.

Proof of Theorem 4.1
Assume that $A' \subset A$ and $B' \subset B$ satisfy the assertion of Lemma 4.2. For $a \in A'$, $b \in B'$, consider a path (a, b', a', b). Clearly,

$$y = a + b = (a + b') - (a' + b') + (a' + b) = x - x' + x'',$$

where $x = a + b', x' = a' + b', x'' = a' + b$ are elements of $X = A +_G B$. (Here we make critical use of the fact that (a, b'), (b', a'), (a', b) are edges of G.) Thus every element $y \in A' + B'$ equals $x - x' + x''$ for at least $n^2/(2^{12}K^5)$ ordered triples (x, x', x''). On the other hand, X has cardinality at most Cn, so there are at most C^3n^3 such triples. This implies that the number of y's is at most

$$\frac{C^3n^3}{n^2/(2^{12}K^5)} = 2^{12}C^3K^5n;$$

that is, $|A' + B'| \leq 2^{12}C^3K^5n$, as claimed. □

Proof of Lemma 4.2
First, we delete from B all vertices with degree less than $n/(2K)$. This way, at most $n^2/(2K)$ edges get deleted, so the remaining graph has at least $n^2/(2K)$ edges. For convenience, let us keep the same notation. We are now working with a graph $G(A, B, E)$, where $|A| = n$, $|B| \leq n$, and $|E| \geq n^2/(2K)$.

Choose a point $v \in A$ uniformly at random, and consider the neighborhood $N(v)$ of v. The set B' will be a subset of $N(v)$; but first we need to prove a few properties of $N(v)$. Set $X = |N(v)|$. Clearly,

$$\mathbf{E}(X) = \frac{\sum_{v \in A} |N(v)|}{|A|} = \frac{|E|}{|A|} \geq \frac{n^2/(2K)}{n} = \frac{n}{2K}. \tag{1}$$

Next, we call a pair (u, w), $u, w \in B$, *bad* if the number of common neighbors of u and w is less than $n/(128K^3)$, that is, if $|N(u, w)| \leq n/(128K^3)$. We are going to estimate Y, the number of bad pairs that belong to $N(v)$. Consider a bad pair (u, w); both u and w belong to $N(v)$ if v is chosen from $N(u, w)$. The probability that this happens is

$$\frac{|N(u, w)|}{n} \leq \frac{n/(128K^3)}{n} = \frac{1}{128K^3}. \tag{2}$$

Since $|B| \leq n$, there are at most $\binom{n}{2}$ bad pairs. Thus, by the linearity of expectations, the expectation of Y is at most

$$\mathbf{E}(Y) \leq \frac{1}{128K^3}\binom{n}{2} \leq \frac{n^2}{256K^3}.$$

Let S be the set of elements $u \in N(v)$ which form a bad pair with at least $n/(32K^2)$ other elements $w \in N(v)$. Letting Z be the cardinality of S, we have

$$Z\frac{n}{32K^2} \le 2Y,$$

where the factor 2 is due to the fact that a bad pair (u, w) may be counted twice, once from u and once from w. It follows that

$$\mathbf{E}(Z) \le \frac{32K^2}{n} \cdot 2\mathbf{E}(Y) \le \frac{64K^2}{n}\frac{n^2}{256K^3} = \frac{n}{4K}. \tag{3}$$

Inequalities (1) and (3) imply that $\mathbf{E}(X - Z) \ge n/(4K)$. Thus there is a choice of v such that $X - Z \ge n/(4K)$, namely, $|N(v)\backslash S| \ge n/(4K)$. Pick such a vertex v, and denote by B' the set $N(v)\backslash S$. Then $|B'| \ge n/(4K)$, and for every $u \in B'$, there are at most $n/(32K^2)$ elements $w \in B'$ such that (u, w) is bad.

Next, we define A' to be the set of those $a \in A$ where a has at least $n/(16K^2)$ neighbors in B'. We first estimate $|A'|$ by counting the number of edges between B' and A. Since each vertex in B has degree at least $n/(2K)$, this number is at least $n/(2K)|B'| \ge n^2/(8K^2)$. On the other hand, it is at most

$$n|A'| + \frac{n}{16K^2}|A\backslash A'| < n|A'| + \frac{n}{16K^2}n.$$

It follows that $n|A'| \ge n^2/(16K^2)$, so $|A'| \ge n/(16K^2)$, as claimed.

Now we show that for any $a \in A'$ and $b \in B'$, there are many paths of length 3 connecting a and b. Indeed, $a \in A$ has at least $n/(16K^2)$ neighbors in B'. At most $n/(32K^2)$ of these neighbors can form a bad pair with b, so there are $n/(32K^2)$ neighbors b' of a such that $|N(b, b')| \ge n/(128K^3)$. For any $a' \in N(b, b')$, the four points a, b', a', b (in that order) form a path of length 3. Thus the number of paths is at least

$$\frac{n}{32K^2} \cdot \frac{n}{128K^3} \ge \frac{n^2}{2^{12}K^5},$$

completing the proof. □

Now we use the above approach to obtain a hypergraph version of Theorem 4.1. Let $A_1, \ldots, A_k$ be k sets of integers, and let E be some family of ordered k-tuples $(a_1, \ldots, a_k)$ such that $a_i \in A_i$, $1 \le i \le k$. The sets $A_1, \ldots, A_k$ together with E define a k-uniform, k-partite hypergraph H, where E is the edge set of H. (Notice that a bipartite graph is a special case when $k = 2$.) We denote by $\bigoplus_{Hi=1}^{k} A_i$ the collection of the sums $a_1 + \cdots + a_k$, where $(a_1, \ldots, a_k) \in E$.

THEOREM 4.3

For any positive integer k, there are polynomials $f_k(x, y)$ and $g_k(x, y)$ with degrees

and coefficients depending only on k such that the following holds. Let n, C, K be positive numbers. If $A_1, \dots, A_k$ are sets of n positive integers, $H(A_1, \dots, A_k, E)$ is a k-partite, k-uniform hypergraph with at least n^k/K edges, and $\left|\bigoplus_{Hi=1}^{k} A_i\right| \le Cn$, then one can find subsets $A_i' \subset A_i$ such that

- $|A_i'| \ge n/(f_k(C, K))$ *for all* $1 \le i \le k$;
- $|A_1' + \cdots + A_k'| \le g_k(C, K) \cdot n$.

Proof

The heart of the proof is the following claim.

CLAIM 4.4

Let $A_1, \dots, A_k$ and n, C, K be as in Theorem 4.3. Set $X = \bigoplus_{Hi=1}^{k} A_i$. There are subsets $A_i' \subset A_i, i = 1, \dots, k$, sets of integers $Y_j, 1 \le j \le 2k-2$, and polynomials $\alpha_k(C, K), \beta_k(C, K), \gamma_k(C, K)$ with coefficients and degrees depending only on k such that the following properties hold.

- $|Y_j| \le \alpha_k(C, K)n$.
- $|A_i'| \ge n/(\beta_k(C, K))$.
- *Every element in $A_1' + \cdots + A_k'$ can be written in the form $x + \sum_{j=1}^{2k-2} y_j$, where $x \in X$, $y_j \in Y_j$ in at least $n^{2k-2}/(\gamma_k(C, K))$ ways.*

It is easy to deduce Theorem 4.3 from this claim via the same counting argument used in the proof of Theorem 4.1. For the sets $A_1', \dots, A_k'$ as in the claim, we have

$$\begin{aligned} |A_1' + \cdots + A_k'| &\le \frac{|X| \prod_{j=1}^{2k-2} |Y_j|}{n^{2k-2}/(\gamma_k(C, K))} \\ &\le C\alpha_k^{2k-2}(C, K)\gamma_k(C, K) \cdot n. \end{aligned}$$

Thus one can set $f_k = \beta_k(C, K)$ and $g_k = C\alpha_k^{2k-2}(C, K)\gamma_k(C, K)$ to conclude the proof.

Proof of Claim 4.4

We prove the claim by induction on k. The base case $k = 2$ was treated in the proof of Theorem 4.1. Let us now consider $k \ge 3$. Denote by E' the set of all $(k-1)$-tuples $(a_2, \dots, a_k)$ with $a_i \in A_i, 2 \le i \le k$, such that $(a_1, a_2, \dots, a_k) \in E$ for at least $n/(2K)$ elements $a_1 \in A_1$. By definition,

$$n|E'| + \frac{n}{2K} n^{k-1} \ge |E| \ge \frac{n^k}{K};$$

hence

$$|E'| \ge \frac{n^{k-1}}{2K}.$$

Let $H' = H'(A_2, \dots, A_k, E')$ be the corresponding $(k-1)$-uniform, $(k-1)$-partite hypergraph whose edge set is E', and let $Z = \bigoplus_{H'} {}_{i=2}^{k} A_i$. To bound the size of Z, note that for every $z \in Z$ there are at least $n/(2k)$ elements $a \in A_1$ for which $a + z \in X$. A simple double counting argument shows that there is an element $a_1 \in A$ such that $a_1 + z \in X$ for at least $(|Z|n/(2K))/|A_1| = |Z|/(2k)$ elements z. This implies that $|Z|/(2k) \le |X| \le Cn$, and hence $|Z| \le 2KCn$.

Next, we show that there are at least $n/(4K)$ elements $z \in Z$ such that $z = a_2 + \cdots + a_k$ for at least $n^{k-2}/(8K^2C)$ edges $(a_2, \dots, a_k) \in E'$. If it were not the case, we would get a contradiction since

$$\begin{aligned}\frac{n^{k-1}}{2K} \le |E'| &< \frac{n^{k-2}}{8K^2C}|Z| + n^{k-2}\frac{n}{4K} \\ &\le \frac{n^{k-2}}{8K^2C}(2KCn) + \frac{n^{k-1}}{4K} \\ &= \frac{n^{k-1}}{2K}.\end{aligned}$$

Let $Z' \subset Z$ be a set of size $n/(4K)$ such that for every $z \in Z'$, $z = a_2 + \cdots + a_k$ for at least $n^{k-2}/(8K^2C)$ edges $(a_2, \dots, a_k) \in E'$.

Consider the bipartite graph with color classes A_1 and Z' and edge set $\mathscr{E} = \{(a, z) \mid a + z \in X\}$. By the definition of Z', every element $z \in Z' \subset Z$ has degree at least $n/(2K)$ in A_1, so this graph has at least $|Z'|(n/(2K)) = n^2/(8K^2)$ edges. Applying Lemma 4.2 as in the proof of Theorem 4.1, we can find subsets $A_1' \subset A_1, |A_1'| \ge n/(2^{10}K^4)$ and $Z'' \subset Z', |Z''| \ge n/(32K^2)$ such that every element $A_1' + Z''$ can be written as $x_1 - x_2 + x_3$ for at least $n^2/(2^{27}K^{10})$ ordered triples $(x_1, x_2, x_3), x_i \in X$.

Denote by E'' the set of all $(k-1)$-tuples $(a_2, \dots, a_k)$ with $a_i \in A_i, 2 \le i \le k$, and $a_2 + \cdots + a_k \in Z''$. Since by our assumption every element of Z'' equals $a_2 + \cdots + a_k$ for at least $n^{k-2}/(8K^2C)$ different $(k-1)$-tuples, we obtain

$$|E''| \ge \frac{n^{k-2}}{8K^2C}|Z''| = \frac{n^{k-1}}{2^8K^4C}.$$

Let $H'' = H''(A_2, \dots, A_k, E'')$ be the $(k-1)$-uniform, $(k-1)$-partite hypergraph whose edge set is E''. This hypergraph has at least $n^{k-1}/(2^8K^4C)$ edges and

$$\bigoplus_{H''}{}_{i=2}^{k} A_i = Z''.$$

Thus, by the induction hypothesis, there are $2k-4$ sets of integers Y_j, $|Y_j| \le \alpha(C, K)n$, $3 \le j \le 2k-2$, and subsets $A_i' \subset A_i, |A_i'| \ge n/(\beta(C, K)), 2 \le i \le k$, such that every element in $A_2' + \cdots + A_k'$ can be written as $z + \sum_{j=3}^{2k-2} y_j$ for at least

$n^{2k-4}/(\gamma(C,K))$ sequences $(z, y_3, \ldots, y_{2k-3})$, where $z \in Z''$, $y_j \in Y_j$ and $\alpha(C,K)$, $\beta(C,K)$, $\gamma(C,K)$ are polynomials of fixed degrees and coefficients in K and C.

Now consider an element $a_1 + a_2 + \cdots + a_k \in A_1' + \cdots + A_k'$. We have the fact that $a_2 + \cdots + a_k$ can be written in at least $n^{2k-4}/(\gamma(C,K))$ ways as $z + \sum_{j=3}^{2k-2} y_j$ with $z \in Z''$ and $y_j \in Y_j$. We also have the fact that for every z in the above sum, $a_1 + z$ can be written in at least $n^2/(2^{27}K^{10})$ ways as $x_1 - x_2 + x_3$ with $x_i \in X$. Therefore $a_1 + a_2 + \cdots + a_k$ can be written as $x_1 - x_2 + x_3 + \sum_{j=1}^{2k-4} y_j$ in at least

$$\frac{n^2}{2^{27}K^{10}} \cdot \frac{n^{2k-4}}{\gamma(C,K)} = \frac{n^{2k-2}}{2^{27}K^{10}\gamma(C,K)}$$

different ways. Define $Y_1 = -X$ and $Y_2 = X$. This completes the proof of the induction step and the proof of Claim 4.4. □

Thus Theorem 4.3 is proven. □

The following statement is a special case of the result which was proved by Ruzsa using Plunnecke's theorem (see, e.g., Nathanson [11]). This lemma gives us control on the cardinality of the sumset $B + B$, given that we know something about the cardinality of the sumset $A + B$ for some set A.

LEMMA 4.5
Let A and B be two finite sets of integers; then

$$|B + B| \le \frac{|A + B|^2}{|A|}.$$

Using this lemma, one can obtain the following corollary of Theorem 4.1.

COROLLARY 4.6
Let n, C, K be positive numbers, and let A be a set of n integers. Suppose that there is a graph H with A as its vertex set with at least $n^2/(2K)$ edges and $|A +_H A| \le Cn$. Then one can find a subset $B \subset A$ such that $|B| \ge n/(4K)$ and $|B + B| \le 2^{28}C^6K^{13}|B|$.

Proof
Define a bipartite graph G as follows. The two color classes are two copies of A, and a vertex u in the first color class is connected to vertex v in the second color class if and only if (u, v) was an edge in H. Since every edge in H contributes two edges of G, it has n^2/K edges. By definition, $|A +_H A| = |A +_G A|$. Using Theorem 4.1, we can find two subsets A' and B in A such that $|A'| \ge n/(16K^2)$, $|B| \ge n/(4K)$, and

$|A' + B| \le 2^{12}C^3K^5n$. Then by Lemma 4.5,

$$|B + B| \le \frac{|A' + B|^2}{|A'|} \le 2^{26}C^6K^{12}n \le 2^{28}C^6K^{13}|B|,$$

completing the proof. □

5. A multiplicative property of (α, β)-sets

Our last tool is a multiplicative property of generalized arithmetic progressions. Let us start with a definition that, in a slightly different form, first appeared in [15]. It plays an important role in our proof.

Definition 5.1

A finite set X of integers is called an (α, β)-set if it contains a subset X' with the following two properties: $|X'| \ge \alpha|X|$ and $|X' + X'| \le \beta|X'|$.

For a set X, we use the notation $2 \times X$ to denote the set obtained by doubling every element of X; that is, $2 \times X = \{2x | x \in X\}$. We are concerned with the difference set $(2 \times X)\backslash X$. It is clear that if there is no restriction on X, then the above difference set can be very small. For instance, if X consists of consecutive powers of two, then the difference set has exactly one element. On the other hand, we are going to show that the (α, β)-property forces this difference set to be large.

LEMMA 5.2

For any positive constants $\beta > 1 > \alpha$, the following holds. If X is an (α, β)-set, then X contains a subset Y with density at least $\alpha^2/(400\beta^2)$ such that $2 \times Y$ is disjoint from X; that is,

$$|(2 \times X)\backslash X| \ge \frac{\alpha^2}{400\beta^2}|X|.$$

Notice that the (α, β)-property is an additive property. Thus Lemma 5.2 can be viewed, intuitively, as evidence of the general phenomenon that additive and multiplicative properties cannot hold together. (A famous problem of this type is the Erdős-Szemerédi sum-product problem; see [6].)

Proof

Set $\delta = \alpha^2/(400\beta^2)$. Assume, by contradiction, that X is an (α, β)-set but

$$|(2 \times X)\backslash X| < \delta|X|.$$

Set $\epsilon = \alpha/(9\beta)$. Let X_i, $i = 0, 1, \ldots$, be the set of those elements of X which are divisible by 2^i but not divisible by 2^{i+1}. We consider two cases.

(a) There is some index i such that $|X_i| \geq \epsilon|X|$. In this case, notice that the difference set $(2 \times X_i)\backslash X_{i+1}$ is a subset of $(2 \times X)\backslash X$. Since $|(2 \times X)\backslash X| < \delta|X|$, it follows that $|(2 \times X_i)\backslash X_{i+1}| < \delta|X|$. But $|X_i| \geq \epsilon|X|$, so

$$|X_{i+1}| \geq |X_i| - \delta|X| \geq (\epsilon - \delta)|X|.$$

By induction, we have the fact that $|X_{i+j}| \geq (\epsilon - j\delta)|X|$ for all $\epsilon/\delta \geq j \geq 0$. On the other hand, by the definitions of δ and ϵ, it is routine to show that

$$\sum_{j=0}^{\lfloor \epsilon/\delta \rfloor} (\epsilon - j\delta) > 1.$$

It thus follows that

$$|X| \geq \sum_{j=0}^{\lfloor \epsilon/\delta \rfloor} |X_{i+j}| \geq \sum_{j=0}^{\lfloor \epsilon/\delta \rfloor} (\epsilon - j\delta)|X| > |X|, \tag{4}$$

which is a contradiction.

(b) We have $|X_i| < \epsilon|X|$ for every i. We exploit the fact that X is an (α, β)-set. Let X'_i denote the intersection of X_i and X', where X' is as in Definition 5.1. Since $\epsilon < \alpha/3$, there is an index l such that the union $Y = \bigcup_{i=0}^{l} X'_i$ has cardinality between $(1/3)|X'|$ and $(2/3)|X'|$. Let L be the number of $X'_i, i \leq l$, which are not empty. Since

$$|X'_i| \leq |X_i| < \epsilon|X| \leq \frac{\epsilon}{\alpha}|X'|,$$

it follows that

$$L > \frac{\alpha}{3\epsilon}. \tag{5}$$

Setting $Z = X'\backslash Y$, we have $(1/3)|X'| \leq |Z| \leq (2/3)|X'|$. We reach a contradiction by showing that the sumset $X' + X'$ is too large. Let i and j be two indices such that $0 \leq i < j \leq l$ and X'_i and X'_j are not empty. Let x_i be an element of X'_i, and let x_j be an element of X'_j. For any two elements $u, v \in Z$, we claim that

$$x_i + u \neq x_j + v. \tag{6}$$

To verify this claim, notice that by the definition of X_i, x_i is divisible by 2^i but not divisible by 2^{i+1}. Furthermore, u is divisible by 2^l and thus is divisible by 2^i. It follows that the sum $x_i + u$ is not divisible by 2^{i+1}. On the other hand, both elements of the right-hand side, x_j and v, are divisible by 2^j and thus are divisible by 2^{i+1} as $j > i$.

Thus (6) implies that $Y + Z$ has at least $L|Z|$ elements. Since $Y + Z$ is a subset of $X' + X'$, it follows, via (5), that

$$|X' + X'| \geq |Y + Z| \geq L|Z| > \frac{\alpha}{3\epsilon}\frac{|X'|}{3}.$$

Substituting $\epsilon = \alpha/(9\beta)$, we obtain $|X' + X'| > \beta|X'|$. This contradicts the assumption that X is an (α, β)-set and completes our proof. □

6. Proof of Theorem 1.2

Let us recall the setting of the theorem. We have two finite sets X and Y, where $1/(h^{29})|Y| \geq |X| \geq F(h)$. We want to show that Y contains a subset Z of cardinality h such that Z is sum-free with respect to the union $U = X \cup Y$. Throughout the proof, we use the asymptotic notation under the assumption that $h \to \infty$.

Define a graph H on Y using the rule that (w, v), $w, v \in Y$, is an edge if and only if $w + v$ belongs to U. It is clear that an independent set of this graph is sum-free with respect to U. If H has less than $|Y|^2/(3h)$ edges, then by Lemma 3.1, there is an independent set Z of size larger than h, and we are done. Thus we can assume that the number of edges in H is at least $|Y|^2/(3h)$. By definition, $|Y +_H Y| \leq |U| \leq 2|Y|$, so we can apply Corollary 4.6 with $n = |Y|$, $K = 3h/2$, and $C = 2$. By this corollary, there is a set $Y_1 \subset Y$ such that $|Y_1| = \Theta(|Y|/h)$ and $|Y_1 + Y_1| = O(h^{13}|Y_1|)$. (One can have explicit values for the hidden constants in Θ and O from Corollary 4.6, but we are not interested in these numbers.)

By the above discussion, $U = X \cup Y$ is a $(c/h, c'h^{13})$-set for some positive constants c and c'. (Here Y_1 plays the role of X' in Definition 5.1.) We are now in position to invoke Lemma 5.2. It implies that there is a subset U' of U with cardinality $\Omega(|Y|/h^{28})$ such that $2 \times U'$ is disjoint from U. Since $X \leq |Y|/h^{29}$, it follows that for sufficiently large h, $|U'| \geq 3|X|$. Thus $Y_2 = U' \backslash X$ still has cardinality $\Omega(|Y|/h^{28})$. Moreover, $Y_2 \subset Y$, and $2 \times Y_2$ is disjoint from U.

Next, we repeat the arguments from the beginning of the proof with Y_2 now playing the role of Y. Let H' be the graph on Y_2 whose edges are the pairs $w, v \in Y_2$ such that $w + v \in U$. By definition, $|Y_2 +_{H'} Y_2| \leq |U| \leq 2|Y| \leq \alpha h^{28}|Y_2|$ for some positive constant α. As before, we can show that H' has at least $|Y_2|^2/(3h)$ edges, and so we can apply Corollary 4.6 with $n = |Y_2|$, $K = 3h/2$, and $C = \alpha h^{28}$. By this corollary, there is set $Y_3 \subset Y_2$ such that $|Y_3| = \Theta(|Y_2|/h) = \Omega(|Y|/h^{29})$ and $|Y_3 + Y_3| = O(h^{181}|Y_3|)$. The crucial extra information we have about Y_3, compared to what we knew about Y_1, is that $2 \times Y_3$ is disjoint from U.

Now we can apply Theorem 3.4 to Y_3. We conclude that there is a proper generalized arithmetic progression $P = \left\{x_0 + \sum_{j=1}^{d} \lambda_j x_j \,\middle|\, 0 \leq \lambda_j \leq m_j - 1\right\}$ of dimension $d = O(h^{181})$ and volume at most $\exp(h^{32770})|Y_3|$ which contains Y_3. Without loss of generality, we can assume that $d = \Theta(h^{181})$. The volume of P is upper bounded by m_1^d. Therefore

$$m_1 \geq |Y_3|^{1/d} \geq \left(\frac{|Y|}{h^{29}}\right)^{1/d} = \Omega(|Y|^{1/d}),$$

where in the last equality we use the fact that $h^{1/d}$ is a constant, which follows from the assumption $d = \Theta(h^{181})$. Note also that P is a disjoint union of one-dimensional

arithmetic progressions of length m_1 and that Y_3 has density at least $\exp(-h^{32770})$ in P. Thus, by averaging, we conclude that there is an arithmetic progression Q of length m_1 such that the intersection $Y_4 = Q \cap Y_3$ has density at least $\delta = \exp(-h^{32770})$ in Q.

Next, we focus on Y_4. We construct the desired set Z as a subset of Y_4 by applying Corollary 3.3. In the current setting, h plays the role of k, and m_1 (the length of Q) plays the role of n. If the corollary can be applied, then we can conclude that Y_4 contains a subset Z with h elements such that the sum of any two elements of Z equals twice some element of Y_4. Since $2 \times Y_4$ is a subset of $2 \times Y_3$, it is disjoint from U. Thus Z is sum-free with respect to U.

The only thing we need to verify is that h, δ, and m_1 satisfy the condition of Corollary 3.3. By taking a logarithm (with natural base) of both sides, we can see that this condition is equivalent to

$$2 \uparrow (e^{h^{32770}}) \uparrow 2 \uparrow 2 \uparrow (2h+9) \le \log_2 e \ln m_1.$$

Recall that $d = \Theta(h^{181})$ and $m_1 = \Omega(|Y|^{1/d})$. Hence, given that h is sufficiently large, we have

$$\ln m_1 = \Omega\Big(\frac{1}{h^{181}} \ln |Y|\Big) \ge \frac{\ln |Y|}{h^{182}}.$$

Now it is enough to show that

$$\ln |Y| \ge h^{182} \times 2 \uparrow (e^{h^{32770}}) \uparrow 2 \uparrow 2 \uparrow (2h+9).$$

Since the right-hand side is $\ln F(h)$, the last inequality follows directly from the assumption of the theorem that $|Y| \ge |X| \ge F(h)$. This completes our proof. □

7. Proof of Theorems 1.3 and 1.4

One can modify the proof of Theorem 1.1 to show how to deduce Theorem 1.3 from Theorem 1.4. Here we briefly sketch the parts in the proof which need slight adjustments. The most notable difference is the definition of sets X and Y at every step of the construction.

As before, let A_i denote the set of the first $((h^{\alpha_k})^i F(h))$-elements of A, where α_k is a sufficiently large constant. At step i, we already found sets $B_1, \dots, B_i$ such that every B_j is a subset of $A_j \setminus A_{j-1}$, has size h, and $\bigcup_{j=1}^{i} B_j$ is k-sumset-free with respect to A. Let X' be the set of all integers which for some index $1 \le j \le i$ can be represented as the difference between an element from A_j and the sum of at most $k-1$ elements from $\bigcup_{l=j}^{i} B_l$. For an appropriate choice of α_k, we have

$$|X'| \le \sum_{j=1}^{i} \big((i+1-j)h\big)^{k-1} |A_j| = |A_i| \sum_{j=1}^{i} \frac{((i+1-j)h)^{k-1}}{(h^{\alpha_k})^{i-j}} \le O(h^{k-1}|A_i|).$$

Let $Y = A_{i+1} \setminus A_i$, and let $X = A_i \cup X'$; then

$$|X| \le |X'| + |A_i| \le O(h^k |A_i|) \le |Y|/h^{c_k}.$$

Therefore, as before, we can apply Theorem 1.4 to find a new set $B_{i+1} \subset A_{i+1} \setminus A_i$ of size h such that the set $\bigcup_{l=1}^{i+1} B_l$ is still sumset-free with respect to A.

To prove Theorem 1.4, we need to generalize several of our tools. These generalizations are presented in Sections 7.1 – 7.4.

7.1. Independent sets in hypergraphs

First of all, we need a general version of Lemma 3.1. In the current situation, graphs no longer suffice, and we need to consider hypergraphs.

Consider k hypergraphs $H_1, \dots, H_k$ on the same vertex set V. We say that a subset I of V is independent with respect to all H_l if I does not contain any edge from any H_l, $1 \le l \le k$. We need the following generalization of Lemma 3.1.

LEMMA 7.1

For every positive integer k, there is a constant c_k such that the following holds. Let H_l, $l = 2, \dots, k$, be an l-uniform hypergraph on the same vertex set V. If the number of edges of H_l is at most $|V|^l/(c_k h^{l-1})$ for all $2 \le l \le k$ and $|V| \ge 2h$, then V contains a subset I of size h which is independent with respect to all H_l.

Our proof shows that one can set $c_k = 2^{k+1}$. The optimal value of c_k might be much smaller, but it is not our concern at the moment.

Proof

This proof is different from the proof of Lemma 3.1. We create I by the following procedure. First, select each vertex of V, randomly and independently, with a probability p (the value of p is optimized later). Let S denote the set of selected vertices. Delete one vertex from every edge $e \in \bigcup H_l$, which is entirely contained in S. Denote the remaining set as I. It is clear that I is an independent set with respect to all H_l. We show that with an appropriate choice of p, the expectation of $|I|$ is at least h, and this proves the claim of the lemma.

The expectation of $|S|$ is, trivially, $p|V|$. Let E_l denote the set of edges of H_l. For any $2 \le l \le k$, the expectation of the number of edges of H_l falling entirely into S is

$$p^l |E_l| \le \frac{p^l |V|^l}{c_k h^{l-1}}.$$

Therefore the expectation of I is at least

$$p|V| - \sum_{l=2}^{k} \frac{p^l |V|^l}{c_k h^{l-1}}.$$

Setting $p = 2h/|V|$ and $c_k = 2^{k+1}$, we have

$$|I| \geq p|V| - \sum_{l=2}^{k} \frac{p^l |V|^l}{c_k h^{l-1}} = h\Big(2 - \sum_{l=2}^{k} \frac{1}{2^{k+1-l}}\Big) > h,$$

completing the proof. □

7.2. Difference sets with simultaneous multipliers

Another tool we need is the following generalization of Lemma 5.2.

LEMMA 7.2

For any positive constants $\beta > 1 > \alpha$ and a set of positive integers $a_1, \ldots, a_k \geq 2$, there is a constant δ such that the following holds. If X is an (α, β)-set, then it contains a subset Y of density at least δ such that $a_i \times Y$ is disjoint from X for all $1 \leq i \leq k$. Furthermore, δ is bounded from below by a polynomial in α and β^{-1} whose degree and coefficients depend only on $a_1, \ldots, a_k$.

Proof

Let $p_1, \ldots, p_m$ be the prime divisors of the product $a_1 \cdots a_k$. For each vector $\mathbf{e} = (e_1, \ldots, e_m)$, where $0 \leq e_i$, let $X_{\mathbf{e}}$ be the collection of all elements x of X, where x is divisible by $\prod_{i=1}^{m} p_i^{e_i}$ but $x/\big(\prod_{i=1}^{m} p_i^{e_i}\big)$ is not divisible by any of the p_i's. The sets $X_{\mathbf{e}}$ form a partition of X. We use the following simple fact, whose proof is left to the reader.

Fact 7.3

Let x and y be two different elements of some $X_{\mathbf{e}}$. Let a and b be two integers whose prime divisors are all among $p_1, \ldots, p_m$. Then $ax \neq by$.

The structure of the proof of Lemma 7.2 is similar to that of the proof of Lemma 5.2. We consider two cases. The analysis in each case is a little bit more involved than in the proof of Lemma 5.2.

First, we exploit the fact that X is an (α, β)-set. Let X' be a subset of X such that $|X'| \geq \alpha|X|$ and $|X' + X'| \leq \beta|X'|$, and let $X'_{\mathbf{e}}$ be the intersection of $X_{\mathbf{e}}$ and X'. We claim that there is a vector $\mathbf{e} = (e_1, \ldots, e_m)$ such that $X'_{\mathbf{e}} \geq |X'|/(g_m(\beta))$, where $g_m(x)$ is defined as $2^m \cdot \prod_{i=0}^{m-1}(2^{i+1}x + 1)$ if $m \geq 1$ and $g_0(x) = 1$. Note that $g_m(\beta)$ is polynomial in β and is bounded by $2^m(2^{m+1}\beta)^m = 2^{m^2+2m}\beta^m$ for all $\beta > 1$. We prove our claim by induction on m. For $m = 0$, it is obvious.

Let X'_i, $i = 0, 1, \ldots$, be the set of those elements of X' which are divisible by p_1^i but not divisible by p_1^{i+1}. Denote by l the largest index such that the union $\bigcup_{i=0}^{l} X'_i$ still has cardinality less than $|X'|/2$. Let $Z = \bigcup_{j>l} X'_j$, and let L be the number

of $X'_i, i \le l$, which are not empty. As we already show in the proof of Lemma 5.2, Z has the property $|X' \setminus Z + Z| \ge L|Z|$. Since, by definition, $|Z| \ge |X'|/2$ and $|X' + X'| \le \beta|X'|$, we conclude that $L \le 2\beta$.

Let $X'' = \bigcup_{i=0}^{l+1} X'_i$. By our assumptions on l, it has size at least $|X'|/2$ and thus has the property

$$|X'' + X''| \le |X' + X'| \le \beta|X'| \le 2\beta|X''|.$$

Now, by the induction hypothesis, there is a vector $\mathbf{e}' = (e_2, \dots, e_m)$ such that $X''_{\mathbf{e}'} \ge |X''|/(g_{m-1}(2\beta))$. Let $e_1 \le l+1$ be the index such that $|X''_{\mathbf{e}'} \cap X'_{e_1}|$ is maximal, and let $\mathbf{e} = (e_1, e_2, \dots, e_m)$. Then, by definition, $X''_{\mathbf{e}'} \cap X'_{e_1}$ is a subset of $X'_{\mathbf{e}}$ and has size at least $|X''_{\mathbf{e}'}|/(L+1) \ge |X''_{\mathbf{e}'}|/(2\beta+1)$. This implies that

$$|X'_{\mathbf{e}}| \ge \frac{|X''_{\mathbf{e}'}|}{2\beta+1} \ge \frac{|X''|}{(2\beta+1)g_{m-1}(2\beta)} \ge \frac{|X'|}{2(2\beta+1)g_{m-1}(2\beta)} = \frac{|X'|}{g_m(\beta)}$$

and completes the proof of the claim.

Let $\epsilon = \alpha/(g_m(\beta))$, and let $\delta = \epsilon^2/3$. By the above discussion, we obtain the fact that there is a vector $\mathbf{e} = (e_1, \dots, e_m)$ such that

$$|X_{\mathbf{e}}| \ge |X'_{\mathbf{e}}| \ge \frac{|X'|}{g_m(\beta)} \ge \frac{\alpha|X|}{g_m(\beta)} = \epsilon|X|.$$

One can also easily check that these ϵ and δ satisfy

$$\sum_{i=0}^{\lfloor \epsilon/\delta \rfloor} (\epsilon - i\delta) > 1.$$

Next, we show that $|X_{\mathbf{e}}| \ge \epsilon|X|$ implies the existence of the set Y, which satisfies the assertion of the lemma. Suppose, by contradiction, that such a Y does not exist. We construct a set sequence $X_0, X_1, \dots, X_{\lfloor \epsilon/\delta \rfloor}$ with the following properties:

- $X_0 = X_{\mathbf{e}}$;
- $|X_i| \ge (\epsilon - i\delta)|X|$;
- the X_i's are mutually disjoint subsets of X;
- for any $x \in X_i, i \ge 1$, there is $y \in X_{i-1}$ such that $x = ya_l$ for some $1 \le l \le k$.

Then, by the definition of δ, the cardinality of the union of the X_i's would be at least

$$|X| \sum_{i=0}^{\lfloor \epsilon/\delta \rfloor} (\epsilon - i\delta) > |X|,$$

which provides the desired contradiction.

Given the (false) assumption that Y does not exist, the above set sequence is constructed as follows. By definition, $X_0 = X_{\mathbf{e}}$. Assume that $X_0, \dots, X_i$ has been

constructed for some $i \geq 0$. An element y of X_i is *perfect* if the product $a_l y$ does not belong to X for every $1 \leq l \leq k$. By the assumption, the set Y of perfect elements has cardinality less than $\delta|X|$. Thus the set $Z = X_i \backslash Y$ has cardinality at least

$$|X_i| - \delta|X| \geq (\epsilon - i\delta - \delta)|X| = \big(\epsilon - (i+1)\delta\big)|X|.$$

For any $z \in Z$, there is some a_l $(1 \leq l \leq k)$ such that $a_l Z$ belongs to X. Thus we can partition Z into k mutually disjoint sets $Z_1, Z_2, \ldots, Z_k$ such that for any $z \in Z_l$ $(1 \leq l \leq k)$, the product $a_l z$ belongs to X. Define $X_{i+1} = \bigcup_{l=1}^{k} a_l \times Z_l$. Consider an element u in $a_l \times Z_l$ and an element v in $a_{l'} \times Z_{l'}$ for some $l \neq l'$. There are two different elements x and y in $X_{\mathbf{e}}$ such that u is the product of x with $i+1$ (not necessarily different) numbers from the set $\{a_1, \ldots, a_k\}$ and v is the product of x with $i+1$ (not necessarily different) numbers from the set $\{a_1, \ldots, a_k\}$. By Fact 7.3, u and v are different. Thus the sets $a_l \times Z_l$ are disjoint from each other, and

$$|X_{i+1}| \geq \big(\epsilon - (i+1)\delta\big)|X|.$$

To show that X_{i+1} is disjoint from the union $\bigcup_{j=0}^{i} X_j$, consider an element u in X_{i+1} and an element $v \in X_j$, $0 \leq j \leq i$. By the above argument, u is the product of an element x in X_0 with $i+1$ (not necessarily different) numbers from the set $\{a_1, \ldots, a_k\}$, and v is the product of an element x in X_0 with j (not necessarily different) numbers from the set $\{a_1, \ldots, a_k\}$. Note that, given u, x is uniquely determined. If $x \neq y$, then, again by Fact 7.3, u and v are different. If $x = y$, then u is the product of v with $i+1-j$ numbers from the set $\{a_1, \ldots, a_k\}$, and thus $u > v$. This concludes the proof. □

7.3. A general version of Corollary 3.3

Let $M(k)$ denote the least common divisors of $2, \ldots, k$. The following statement is an immediate corollary of Theorem 3.2.

COROLLARY 7.4

Let A be a subset of density δ of an arithmetic progression of length n, where $0 < \delta \leq 1/2$ and

$$n \geq 2 \uparrow 2 \uparrow (\delta)^{-1} \uparrow 2 \uparrow 2 \uparrow \big(M(k)h + 9\big).$$

Then A contains a subset A' of h elements such that for any $2 \leq l \leq k$ and any l different elements $x_1, \ldots, x_l$ in A', there is an element z in A satisfying

$$x_1 + \cdots + x_l = lz.$$

Proof

By Theorem 3.2, A contains an arithmetic progression P of length $M(k)h$. The el-

ements of P with indices divisible by $M(k)$ form a set of size h with the desired property. □

7.4. Proof of Theorem 1.4

With all the necessary tools in hand, this proof is merely a formality. One can repeat all arguments from the proof of Theorem 1.2 (appropriately generalized) without much difficulty.

Consider two sets X and Y as given in Theorem 1.4. In the first argument, the obvious modification that we need is to consider many hypergraphs instead of a single graph. Given $2 \le l \le k$, we define a hypergraph H_l on Y as follows. A set of i different elements of Y forms an edge of H_l if its sum belongs to the union $U = X \cup Y$. A subset I of Y is k-sum-free with respect to U if and only if it is independent with respect to all H_l. By Lemma 7.1, either I has at least h elements (and we are done) or there is some $2 \le l \le k$ such that the hypergraph H_l has at least $|Y|^l/(2^{k+1}h^{l-1})$ edges.

There is a subset $S = \{s_1, \dots, s_{l-2}\}$ of Y of size $l-2$ which is contained in at least

$$\frac{|E(H_l)|}{|Y|^{l-2}} \ge \frac{|Y|^2}{2^{k+1}h^{l-1}}$$

edges of H_l. This implies that there are at least $|Y|^2/(2^{k+1}h^{l-1})$ pairs of vertices $v, w \in Y$ such that $v + w + \sum_i s_i \in U$. Define graph G on the vertex set Y whose edges are the above-mentioned pairs. Then G has at least $|Y|^2/(2^{k+1}h^{l-1})$ edges, and $|Y +_G Y| \le |U| \le 2|Y|$. Now we can apply Corollary 4.6 to find a subset $Y_1 \subset Y$ such that $|Y_1| = \Theta(|Y|/h^\alpha)$ and $|Y_1 + Y_1| = O(h^\alpha |Y_1|)$ for some constant α. It implies that U is an $(h^{-\alpha}, h^\alpha)$-set.

Next, we apply Lemma 7.2 with multipliers $2, 3, \dots, k$ to obtain a set $U' \subset U$, where $|U'| = \Omega(|Y|/h^\beta)$ for some constant β such that $l \times U'$ is disjoint from U for all $2 \le l \le k$. Using the condition that $|Y|/|X|$ is bounded from below by a sufficiently large power of h, we can delete from U' all elements belonging to X while only slightly changing the cardinality of U'. This way, we obtain a subset Y_2 of Y, where $|Y_2| = \Omega(|Y|/h^\beta)$ and $l \times Y_2$ is disjoint from U for all $2 \le l \le k$. We proceed the same way as in the proof of Theorem 1.2 to obtain a subset Y_3 of Y and some constant γ such that $|Y_3| = \Omega(|Y|/h^\gamma)$, $|Y_3 + Y_3| = O(h^\gamma |Y_3|)$, and $l \times Y_3$ is disjoint from U for all $2 \le l \le k$.

Applying Theorem 3.4 together with the averaging argument from the proof of Theorem 1.2, we can show that there is an arithmetic progression Q of length $|Y|^{1/h^\eta}$ and a subset $Y_4 \subset Y_3$ such that a set $Y_4 \cap Q$ has density at least $\delta = e^{-h^\eta}$ in Q for some positive constant η. Finally, we apply Corollary 7.4 to complete the proof. One

can easily see that the condition on $\ln|Y|$ becomes

$$\ln|Y| \geq h^{\eta} \times 2 \uparrow (e^{h^{\eta}}) \uparrow 2 \uparrow 2 \uparrow \big(M(k)h + 9\big),$$

where $M(k)$ denotes the least common divisor of $2, \ldots, k$. □

8. Concluding remarks

• Given a set B, so far we have considered the collection of sums of two different elements of B and also the collection of sums of at most k different elements of B. It is natural to ask what happens if we consider all possible partial sums of B. Let us denote by S_B the collection of all partial sums of B

$$S_B = \Big\{ \sum_{x \in C} x \,\Big|\, C \subset B, |C| \geq 2 \Big\}.$$

We say that a set B is *absolutely sum-free* with respect to A if S_B is disjoint from A.

It is fairly simple to prove that any set A of n integers contains a subset B of cardinality $\Omega(\ln n)$ which is absolutely sum-free with respect to A. Indeed, we can construct B as follows. Let $A_i, i = 0, 1 \ldots$, be the set of the 4^i largest elements of A. We claim that we can choose elements $b_0 > b_2 > \cdots > b_r$, $r = \lfloor \log_4 n \rfloor$, such that $b_i \in A_i \setminus A_{i-1}$ and $\{b_1, \ldots, b_r\}$ is absolutely sum-free with respect to A. Let b_0 be the largest element of A. If $b_0, \ldots, b_i$ are defined, let C be the set of integers which for some index $1 \leq j \leq i$ can be represented as the difference of an element from A_j and a partial sum of elements $b_j, \ldots, b_i$ (including the empty sum). Note that C contains all elements of A_i and has size at most

$$\begin{aligned} |C| \leq \sum_{j=0}^{i} |A_j| 2^{i-j+1} &= \sum_{j=0}^{i} 4^j 2^{i-j+1} \\ &= 4^i \sum_{j=0}^{i} 2^{j-i+1} \\ &< 4^{i+1} = |A_{i+1}|. \end{aligned}$$

Let b_{i+1} be any element of $A_{i+1} \setminus C$. By definition, $\{b_1, \ldots, b_i, b_{i+1}\}$ is absolutely sum-free with respect to A_{i+1} and hence also with respect to A.

It would be interesting to decide if this bound can be improved by a multiplicative factor tending to infinity. Our approach for k-sum-free sets does not work in this case. It is easy to obtain a variant of Theorem 1.4, but the condition $|Y|/h^{O(1)} \geq |X|$ would be replaced by a stronger condition that $|Y|/(G(h)) \geq |X|$, where $G(h)$ is exponential in h. Therefore iterating this result gives no improvement over the $\Omega(\ln n)$ bound.

• Any improvement on Gowers's bound on long arithmetic progression automatically leads to an improvement on the order of magnitude of $g(n)$. However, it seems

that even if one has the optimal estimate on long arithmetic progressions, the bound on $g(n)$ is still sublogarithmic in n. The same applies for Chang's bound on Freiman's theorem.

References

[1] A. BALOG and E. SZEMERÉDI, *A statistical theorem of set addition*, Combinatorica **14** (1994), 263 – 268. MR 1305895 131, 138

[2] A. BALTZ, T. SCHOEN, and A. SRIVASTAV, *Probabilistic construction of small strongly sum-free sets via large Sidon sets*, Colloq. Math. **86** (2000), 171 – 176. MR 1808673 130

[3] M.-C. CHANG, *A polynomial bound in Freiman's theorem*, Duke Math. J. **113** (2002), 399 – 419. MR 1909605 132, 137

[4] S. L. G. CHOI, *On a combinatorial problem in number theory*, Proc. London Math. Soc. (3) **23** (1971), 629 – 642. MR 0292785 130

[5] P. ERDŐS, "Extremal problems in number theory" in *Proceedings of Symposia in Pure Mathematics, Vol. VIII*, Amer. Math. Soc., Providence, 1965, 181 – 189. MR 0174539 129, 130

[6] P. ERDŐS and E. SZEMERÉDI, "On sums and products of integers" in *Studies in Pure Mathematics*, Birkhäuser, Basel, 1983, 213 – 218. MR 0820223 144

[7] G. A. FREIMAN, *Foundations of a Structural Theory of Set Addition*, Transl. Math. Monogr. **37**, Amer. Math. Soc., Providence, 1973. MR 0360496 131, 137

[8] W. T. GOWERS, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Funct. Anal. **8** (1998), 529 – 551. MR 1631259 132, 138

[9] ———, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465 – 588; *Erratum*, Geom Funct. Anal. **11** (2001), 869. MR 1844079; MR 1866805 132, 136

[10] R. K. GUY, *Unsolved Problems in Number Theory*, 2nd ed., Problem Books in Math., Unsolved Probl. in Intuitive Math. **1**, Springer, New York, 1994. MR 1299330 129

[11] M. B. NATHANSON, *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, Grad. Texts in Math. **165**, Springer, New York, 1996. MR 1477155 143

[12] I. Z. RUZSA, *Generalized arithmetical progressions and sumsets*, Acta Math. Hungar. **65** (1994), 379 – 388. MR 1281447 137

[13] ———, *Sum-avoiding subsets*, to appear in Ramanujan J. 130

[14] E. SZEMERÉDI, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199 – 245. MR 0369312 131, 136

[15] E. SZEMERÉDI and V. H. VU, *Finite and infinite arithmetic progressions in sumsets*, to appear in Ann. of Math. (2), preprint, 2003, http://math.ucsd.edu/~vanvu/papers.html 144

Sudakov
Department of Mathematics, Princeton University, Princeton, New Jersey 08544, USA; bsudakov@math.princeton.edu

Szemerédi
Computer Science Department, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08854, USA; szemered@cs.rutgers.edu

Vu
Department of Mathematics, University of California, San Diego, La Jolla, California 92093, USA; vanvu@ucsd.edu