Annals of Mathematics, **163** (2006), 1–35

Finite and infinite arithmetic progressions in sumsets

By E. Szemerédi and V. H. Vu*

Abstract

We prove that if A is a subset of at least $cn^{1/2}$ elements of $\{1, \dots, n\}$, where c is a sufficiently large constant, then the collection of subset sums of A contains an arithmetic progression of length n. As an application, we confirm a long standing conjecture of Erdős and Folkman on complete sequences.

1. Introduction

For a (finite or infinite) set A of positive integers, S_A denotes the collection of finite subset sums of A

$$S_A = \left\{ \sum_{x \in B} x | B \subset A, |B| < \infty \right\}.$$

Two closely related notions are that of lA and l^*A: lA denotes the set of numbers which can be represented as a sum of l elements of A and l^*A denotes the set of numbers which can be represented as a sum of l different elements of A, respectively. (If $l > |A|$, then l^*A is the empty set.) It is clear that

$$S_A = \cup_{l=1}^{\infty} l^* A.$$

One of the fundamental problems in additive number theory is to estimate the length of the longest arithmetic progression in S_A, l_A and l^*A, respectively.

The purpose of this paper is multi-fold. We shall prove a sharp result concerning the length of the longest arithmetic progression in S_A. Via the proof, we would like to introduce a new method which can be used to handle many other problems. Finally, the result has an interesting application, as we can use it to settle a forty-year old conjecture of Erdős and Folkman concerning complete sequences.

*Research supported in part by NSF grant DMS-0200357, by an NSF CAREER Grant and by an A. Sloan Fellowship.

Theorem 1.1. *There is a positive constant c such that the following holds. For any positive integer n, if A is a subset of [n] with at least $cn^{1/2}$ elements, then S_A contains an arithmetic progression of length n.*

Here and later $[n]$ denotes the set of positive integers between 1 and n.

The proof Theorem 1.1 introduces a new and useful method to prove the existence of long arithmetic progressions in sumsets. Our method relies on inverse and geometrical arguments, rather than on Fourier analysis like most papers on this topic. This method opens a way to attack problems which previously have seemed very hard. Let us, for instance, address the problem of estimating the length of the longest arithmetic progression in lA (where A is a subset of $[n]$), as a function of l, n and $|A|$. In special cases sharp results have been obtained, thanks to the works of several researchers, including Bourgain, Freiman, Halberstam, Ruzsa and Sárközy [2], [6], [8], [17]. Our method, combined with additional arguments, allows us to derive a sharp bound for this length for a wide range of l and $|A|$. For instance, we can obtain a sharp bound whenever $l = n^\alpha$ and $|A| = n^\beta$, where α and β are arbitrary positive constants at most 1. Details will appear in a subsequent paper [19].

An even harder problem is to estimate the length of the longest arithmetic progression in l^*A. The distinction that the summands must be different frequently poses a great challenge. (A representative example is Erdős-Heilbronn vs Cauchy-Danveport [15].) On the other hand, one of our arguments (the tiling technique discussed in §5) seems to provide an effective tool to overcome this challenge. Although there are still many details to be verified, we believe that with this tool, we could handle l^*A as successfully as lA. As a consequence, one can prove a sharp bound for the length of the longest arithmetic progression in S_A even when the cardinality of A is much smaller than $n^{1/2}$, extending Theorem 1.1. Our method also works for multi-sets (where an element may appear many times). A result concerning multi-sets will be mentioned in Section 7.

Let us now make a few comments on the content of Theorem 1.1. The bound in this theorem is sharp up to the constant factor c. In fact, it is sharp from two different points of view. First, it is clear that if A is the interval $[cn^{1/2}]$, then the length of the longest arithmetic progression in S_A is $O(n)$. Second, and more interesting, there is a positive constant α such that the following holds: For all sufficiently large n there is a set $A \subset [n]$ with cardinality $\alpha n^{1/2}$ such that the longest arithmetic progression in S_A has length $O(n^{3/4})$. We provide a concrete construction at the end of Section 5.

We next discuss an application of Theorem 1.1. We can use this theorem to confirm a well-known and long standing conjecture of Erdős, dating back to 1962. In fact, the study of Theorem 1.1 was partially motivated by this conjecture.

An infinite set A is *complete* if S_A contains every sufficiently large positive integer. The notion of complete sequences was introduced by Erdős in the early sixties and has since then been studied extensively by various researchers (see §6 of [5] or §4.3 of [15] for surveys).

The central question concerning complete sequences is to find sufficient conditions for completeness. In 1962, Erdős [4] made the following conjecture

CONJECTURE 1.2. *There is a constant c such that the following holds. Any increasing sequence $A = \{a_1 < a_2 < a_3 < \dots\}$ satisfying*

(a) $A(n) \geq cn^{1/2}$

(b) S_A *contains an element of every infinite arithmetic progression,*

is complete.

Here and later $A(n)$ denotes the number of elements of A not exceeding n. The bound on $A(n)$ is best possible, up to the constant factor c, as shown by Cassels [3] (see also below for a simple construction). The second assumption (b) is about modularity and is necessary as shown by the example of the sequence of even numbers. So Erdős's conjecture basically says that a sequence is complete if it is sufficiently dense and satisfies a trivially necessary modular condition.

Erdős [4] proved that the statement of the conjecture holds if one replaces (a) by a stronger condition that $A(n) \geq cn^{(\sqrt{5}-1)/2}$. A few years later, in 1966, Folkman [9] improved Erdős' result by showing that $A(n) \geq cn^{1/2+\varepsilon}$ is sufficient, for any positive constant ε. The first and simpler step in Folkman's proof is to show that any sequence satisfying (b) can be partitioned into two subsequences with the same density, one of which still satisfies (b). In the next and critical step, Folkman shows that if A is a sequence with density at least $n^{1/2+\varepsilon}$ then S_A contains an infinite arithmetic progression. His result follows immediately from these two steps. In the following we say that A is *subcomplete* if S_A contains an infinite arithmetic progression. Folkman's proof, quite naturally, led him to the following conjecture, which (if true) would imply Conjecture 1.2.

CONJECTURE 1.3. *There is a constant c such that the following holds. Any increasing sequence $A = \{a_1 < a_2 < a_3 < \dots\}$ satisfying $A(n) \geq cn^{1/2}$ is subcomplete.*

Here is an example which shows that the density $n^{1/2}$ is best possible (up to a constant factor) in both conjectures. Let m be a large integer divisible by 8 (say, 10^4) and A be the sequence consisting of the union of the intervals $[m^{2^i}/4, m^{2^i}/2]$ $(i = 0, 1, 2 \dots)$. It is clear that this sequence has density $\Omega(n^{1/2})$ and satisfies (b). On the other hand, the difference between $m^{2^i}/4$ and the sum of all elements preceding it tends to infinity as i tends to infinity. Thus S_A cannot contain an infinite arithmetic progression. (The constants 1/4 and 1/2 might be improved to slightly increase the density of A.)

Folkman's result has further been strengthened recently by Hegyvári [11] and Łuczak and Schoen [13], who (independently) reduced the density $n^{1/2+\varepsilon}$ to $cn^{1/2}\log^{1/2} n$, using a result of Freiman and Sárközy (see §7). Together with Conjecture 1.3, Folkman also made a conjecture about nondecreasing sequences (where the same number may appear many times). We address this conjecture in the concluding remarks (§7).

An elementary application of Theorem 1.1 helps us to confirm Conjecture 1.3. Conjecture 1.2 follows immediately via Folkman's partition argument. In fact, as we shall point out in Section 7, the statement we need in order to confirm Conjecture 1.3 is weaker than Theorem 1.1.

Corollary 1.4. *There is a positive constant c such that the following holds. Any increasing sequence of density at least $cn^{1/2}$ is subcomplete.*

Corollary 1.5. *There is a positive constant c such that the following holds. Any increasing sequence $A = \{a_1 < a_2 < a_3 < \dots\}$ satisfying*

(a) $A(n) \geq cn^{1/2}$

(b) S_A *contains an element of every infinite arithmetic progression*,

is complete.

Let us conclude this section with a remark regarding notation. Through the paper, we assume that n is sufficiently large, whenever needed. The asymptotic notation is used under the assumption that n tends to infinity. Greek letters ε, γ, δ etc. denote positive constants, which are usually small (much smaller than 1). Lower case letters d, h, g, l, m, n, s denote positive integers. In most cases, we use d, h and g to denote constant positive integers. The logarithms have base two, if not otherwise specified. For the sake of a better presentation, we omit unnecessary floors and ceilings. For a positive integer m, $[m]$ denotes the set of positive integers in the interval from 1 to m, namely, $[m] = \{1, 2, \dots, m\}$.

The notion of sumsets is central in the proofs. If A and B are two sets of integers, $A + B$ denotes the set of integers which can be represented as a sum of one element from A and one element from B: $A+B = \{a+b | a \in A, b \in B\}$. We write $2A$ for $A + A$; in general, $lA = (l-1)A + A$.

A graph G consists of a (finite) vertex set V and an edge set E, where an element of E (an edge) is a (unordered) pair (a, b), where $a \neq b \in V$. The degree of a vertex a is the number of edges containing a. A subset I of $V(G)$ is called an *independent* set if I does not contain any edge. A graph is bipartite if its vertex set can be partitioned into two sets V_1 and V_2 such that every edge has one end point in V_1 and one end point in V_2 (V_1 and V_2 are referred to as the color classes of V).

2. Main lemmas and ideas

Let us start by presenting a few lemmas. After the reader gets himself/herself acquainted with these lemmas, we shall describe our approach to the main theorem (Theorem 1.1).

As mentioned earlier, our method relies on inverse arguments and so we shall make frequent use of Freiman type inverse theorems. In order to state these theorems, we first need to define generalized arithmetic progressions. A generalized arithmetic progression of rank d is a subset Q of $\mathbb{Z}$ of the form $\{a + \sum_{i=1}^{d} x_i a_i | 0 \leq x_i \leq n_i\}$; the product $\prod_{i=1}^{d} n_i$ is its volume, which we denote by $\mathrm{Vol}(Q)$. The a_i's are the differences of Q. In fact, as two different generalized arithmetic progressions might represent the same set, we always consider generalized arithmetic progressions together with their structures. Let $A = \{a + \sum_{i=1}^{d} x_i a_i | 0 \leq x_i \leq n_i\}$ and $B = \{b + \sum_{i=1}^{d} x_i a_i | 0 \leq x_i \leq m_i\}$ be two generalized arithmetic progressions with the same set of differences. Then their sum $A + B$ is the generalized arithmetic progression $\{(a + b) + \sum_{i=0}^{d} z_i a_i | 0 \leq z_i \leq n_i + m_i\}$.

Freiman's famous inverse theorem asserts that if $|A + A| \leq c|A|$, where c is a constant, then A is a dense subset of a generalized arithmetic progression of constant rank. In fact, the statement still holds in a slightly more general situation, when one considers $A+B$ instead of $A+A$, as shown by Ruzsa [16], who gave a very nice proof which is quite different from the original proof of Freiman. The following result is a simple consequence of Freimain's theorem and Plünnecke's theorem (see [18, Th. 2.1], for a proof). The book [14] of Nathanson contains a detailed discussion on both Plünnecke's and Ruzsa's results.

THEOREM 2.1. *For every positive constant c there is a positive integer d and a positive constant k such that the following holds. If A and B are two subsets of $\mathbb{Z}$ with the same cardinality and $|A + B| \leq c|A|$, then $A + B$ is a subset of a generalized arithmetic progression P of rank d with volume at most $k|A|$.*

In the case $A = B$, it has turned out that P has only $\lfloor \log_2 c \rfloor$ *essential* dimensions. The following is a direct corollary of Theorem 1.3 from a paper of Bilu [1]. One can also see that it is a direct consequence of Freiman's cube lemma and Freiman's homomorphism theorem [7].

THEOREM 2.2. *For any positive constant $c \geq 2$ there are positive constants δ and c' such that the following holds. If $A \subset \mathbb{Z}$ satisfies $|A| \geq c^2$*

and $|2A| \le c|A|$, *then there is a generalized arithmetic progression* P *of rank* $\lfloor \log_2 c \rfloor$ *such that* $\mathrm{Vol}(P) \le c'|A|$ *and* $|P \cap A| \ge \delta|A| \ge \frac{\delta}{c'}\mathrm{Vol}(P)$.

Next, we take a closer look at generalized arithmetic progressions of rank 2. The following two lemmas show that under certain circumstances, a generalized arithmetic progression P of rank 2 contains a long arithmetic progression whose length is proportional to the cardinality of P.

LEMMA 2.3. *Let* $P = \{x_1a_1 + x_2a_2 | 0 \le x_i \le l_i\}$ *be a generalized arithmetic progression of rank* 2 *where* $l_i \ge 5a_i > 0$ *for* $i = 1, 2$. *Then* P *contains an arithmetic progression of length* $\frac{3}{5}|P|$ *and difference* $\gcd(a_1, a_2)$.

This lemma was proved in an earlier paper [18]; we sketch the proof for the sake of completeness.

Proof of Lemma 2.3. We shall prove that P contains an arithmetic progression of length $\frac{3}{5\gcd(a_1,a_2)}(l_1a_1+l_2a_2)$ and difference $\gcd(a_1, a_2)$. A simple argument shows that

$$\frac{3}{5\gcd(a_1, a_2)}(l_1a_1 + l_2a_2) \ge \frac{3}{5}|P|.$$

It suffices to consider the case when a_1 and a_2 are co-prime. In this case we shall actually show that P contains an interval of length $\frac{3}{5}(l_1a_1 + l_2a_2)$.

In the following we identify P with the cube $Q = \{(x_1, x_2) | 0 \le x_i \le l_i\}$ of integer points in $\mathbb{Z}^2$ together with the canonical map

$$f : \mathbb{Z}^2 \to \mathbb{Z} : \ f((x_1, x_2)) = x_1a_1 + x_2a_2.$$

The desired progression will be provided by a walk in this cube, following a specific rule. Once the walk terminates, its two endpoints will be far apart, showing that the progression has large length.

As a_1 and a_2 are co-prime, there are positive integers l_1', l_1'', l_2' and l_2'' such that $l_1', l_1'' < a_2$, $l_2', l_2'' < a_1$ and

$$l_1'a_1 - l_2'a_2 = l_2''a_2 - l_1''a_1 = 1. \tag{1}$$

We show that P contains the interval $[\frac{1}{5}(l_1a_1 + l_2a_2), \frac{4}{5}(l_1a_1 + l_2a_2)]$. Let u_1 and u_2 denote the vectors $(l_1', -l_2')$ and $(-l_1'', l_2'')$, respectively. Set $v_0 = (l_1/5, l_2/5)$. We construct a sequence $v_0, v_1, \ldots,$ such that $f(v_{j+1}) = f(v_j) + 1$ as follows. Once v_j is constructed, set $v_{j+1} = v_j + u_i$ given that one can find $1 \le i \le 2$ such that $v_j + u_i \in Q$ (if both i satisfy this condition then choose any of them). If there is no such i, then stop. Let $v_t = (y_t, z_t)$ be the last point of this sequence. As neither $v_t + u_1$ nor $v_t + u_2$ belong to Q, both of the following two conditions $(*)$ and $(**)$ must hold:

$(*)$ $y_t + l_1' > l_1$ or $z_t - l_2' \le 0$.

$(**)$ $y_t - l_1'' \le 0$ or $z_t + l_2'' > l_2$.

Since $l_1' < a_2 \le l_1/2$, $y_t + l_1' > l_1$ and $y_t - l_1'' \le 0$ cannot occur simultaneously. The same holds for $z_t - l_2'' \le 0$ and $z_t + l_2' > l_2$. Moreover, since $f(v_j)$ is increasing and $y_0 = l_1/5 \ge a_2 > l_1''$ and $z_0 = l_2/5 \ge a_1 > l_2'$, we can conclude that $z_t - l_2' \le 0$ and $y_t - l_1'' \le 0$ cannot occur simultaneously, either. Thus, the only possibility left is $y_t + l_1' > l_1$ and $z_t + l_2'' > l_2$. This implies that $y_t > l_1 - l_1' \ge \frac{4}{5}l_1$ and $z_t > l_2 - l_1'' \ge \frac{4}{5}l_2$. Thus

$$f(v_t) > \frac{4}{5}(l_1a_1 + l_2a_2), \tag{2}$$

concluding the proof. □

Lemma 2.4. *If $U \subset [m]$ is a generalized arithmetic progression of rank 2 and $l|U| \ge 20m$, where both m and $|U|$ are sufficiently large, then lU contains an arithmetic progression of length m.*

Proof of Lemma 2.4. Assume that $U = \{a + x_1a_1 + x_2a_2 | 0 \le x_i \le u_i\}$. We can assume that $u_1, u_2 > 10$ (if u_1 is small, then it is easy to check that lU' contains a long arithmetic progression, where $U' = \{a + x_2a_2 | 0 \le x_2 \le u_2\}$). Now let us consider

$$lU = \{la + x_1a_1 + x_2a_2 | 0 \le x_i \le lu_i\}. \tag{3}$$

By the assumption $l|U| \ge 20m$, we have $l(u_1 + 1)(u_2 + 1) \ge 20m$. As $u_1, u_2 \ge 10$, it follows that $lu_1u_2 \ge 10m$. On the other hand, U is a subset of $[m]$ so the difference of any two elements of U has absolute value at most m. It follows that $u_1a_1 \le m$. This implies

$$u_1a_1 \le m \le lu_1u_2/10.$$

So it follows that $10a_1 \le lu_2$. Similarly $10a_2 \le lu_1$. Thus lU satisfies the assumption of Lemma 2.3 and this lemma implies that lU contains an arithmetic progression of length at least

$$\frac{3}{5}|lU| \ge \frac{3}{5}2m > m,$$

concluding the proof. In the inequality $\frac{3}{5}|lU| \ge \frac{3}{5}2m$ we used the fact that $|lU| \ge 2m$. This fact follows immediately (and with room to spare) from the assumption $l|U| \ge 20m$ and the well-known fact that $|A + B| \ge |A| + |B|$, unless both A and B are arithmetic progressions of the same difference. (We leave the easy proof as an exercise.) □

Despite its simplicity, Lemma 2.4 plays an important role in our proof. It shows that in order to obtain a long arithmetic progression, it suffices to obtain a large multiple of a generalized arithmetic progression of rank 2. As the reader will see, generalized arithmetic progressions of rank 2 are actually the main objects of study in this paper.

The next lemma asserts that by adding several subsets of positive density of a certain generalized arithmetic progression of constant rank, one can fill an entire generalized arithmetic progression of the same rank and comparable cardinality. This is one of our main technical tools and we shall refer to it as the "filling" lemma.

LEMMA 2.5. *For any positive constant* γ *and positive integer* d, *there is a positive constant* γ' *and a positive integer* g *such that the following holds. If* $X_1, \ldots, X_g$ *are subsets of a generalized arithmetic progression* P *of rank* d *and* $|X_i| \geq \gamma \operatorname{Vol}(P)$ *then* $X_1 + \cdots + X_g$ *contains a generalized arithmetic progression* Q *of rank* d *and cardinality at least* $\gamma' \operatorname{Vol}(P)$. *Moreover*, *the distances of* Q *are multiples of the distances of* P.

Remark. The conditions of this lemma imply that the ratio between the cardinality and the volume of P is bounded from below by a positive constant. The quantities $\operatorname{Vol}(P)$, $|P|$, $\operatorname{Vol}(Q)$, $|Q|$, $|X_i|$'s differ from each other by constant factors only.

Let us now give a sketchy description of our plan. In view of Lemma 2.4, it suffices to show that S_A contains a (sufficiently large) multiple of a (sufficiently large) generalized arithmetic progression of rank 2. We shall carry out this task in two steps. The first step is to produce one relatively large generalized arithmetic progression. In the second step, we put many copies of this generalized arithmetic progression together to obtain a large multiple of it. This multiple will be sufficiently large so that we can invoke Lemma 2.4. These two steps are not independent, as both of them rely on the following structural property of A: Either S_A contains an arithmetic progression of length n (and we are done), or a large portion of A is trapped in a small generalized arithmetic progression of rank 2. This is the content of the main structural lemma of our proof.

LEMMA 2.6. *There are positive constants* β_1 *and* β_2 *such that the following holds. For any positive integer* n, *if* A *is a subset of* $[n]$ *with at least* $n^{1/2}$ *elements then either* S_A *contains an arithmetic progression of length* n, *or there is a subset* A' *of* A *such that* $|A'| \geq \beta_1 |A|$ *and* A' *is contained in generalized arithmetic progression* W *of rank* 2 *with volume at most* $n^{1/2} \log^{\beta_2} n$.

The reader might feel that the above description of our plan is somewhat vague. However, at this stage, that is the best we could do without involving too much technicality. The plan will be updated gradually and become more and more concrete as our proof evolves.

There are two technical ingredients of the proof which deserve mentioning. The first is what we call a *tree* argument. This argument, in spirit, works as follows. Assume that we want to add several sets $A_1, \ldots, A_m$. We shall add them in a special way following an algorithm which assigns sets to the vertices of a tree. A set of any vertex contains the sum of the sets of its children. If the set at the root of the tree is not too large, then there is a level where the sizes of the sets do not increase (compared to the sizes of their children) too much. Thus, we can apply Freiman's inverse theorems at this level to deduce useful information. The creative part of this argument is to come up with a proper algorithm which suits our need.

The second important ingredient is the so-called *tiling* argument, which helps us to create a large generalized arithmetic progression by tiling many small generalized arithmetic progressions together. (In fact, it would be more precise to call it *wasteful tiling* as the small generalized arithmetic progressions may overlap.) This technique will be discussed in detail in Section 5.

The rest of the paper is organized as follows. In the next section, we prove Lemma 2.5. In Section 4, we prove Lemma 2.6. Both of these proofs make use of the tree argument mentioned above, but in different ways. The proof of Theorem 1.1 comes in Section 5, which contains the tiling argument. In Section 6, we prove the Erdős-Folkman conjectures. The final section, Section 7, is devoted to concluding remarks.

3. Proof of Lemma 2.5

We shall need the following lemma which is a corollary of a result of Lev and Smelianski (Theorem 6 of [12]). This lemma is relatively easy and the reader might want to consider it an exercise.

LEMMA 3.1. *The following holds for all sufficiently large m. If A and B are two sets of integers of cardinality m and $|A+B| \leq 2.1m$, then A is a subset of an arithmetic progression of length $1.1m$.*

We also need the following two simple lemmas.

LEMMA 3.2. *For any positive constant ε there is a positive integer h_0 such that the following holds. If $h \geq h_0$ and $A_1, \dots, A_h$ are arithmetic progressions of length at least εn of an interval I of length n, then there is a number $h' \geq .09\varepsilon^2 h$ and an arithmetic progression B of length $.9\varepsilon n$ such that at least h' among the A_i's contain B.*

Proof of Lemma 3.2. Consider the following bipartite graph. The first color class consists of $A_1, \dots, A_h$. The other color class consists of the arithmetic progressions of length $.9\varepsilon n$ in I. Since the difference of an arithmetic progression of length $.9\varepsilon n$ in I is at most $1/(.9\varepsilon)$, the second color class has at most $n/(.9\varepsilon)$ vertices. Moreover, an arithmetic progression of length εn contains at least $.1\varepsilon n$ arithmetic progression of length $.9\varepsilon n$. Thus, each vertex in the first class has degree at least $.1\varepsilon n$ and so the number of edges is at least

$.1\varepsilon nh$. It follows that there is a vertex in the second color class with degree at least $\frac{.1\varepsilon nh}{n/(.9\varepsilon)} = .09\varepsilon^2 h$. The progression corresponding to this vertex satisfies the claim of the lemma. □

LEMMA 3.3. *Let B be an interval of cardinality n and B' be a subset of B containing at least $.8n$ elements. Then $B' + B'$ contains an interval of length $1.2n + 2$.*

Proof of Lemma 3.3. Without loss of generality we can assume that $B = [n]$. If an integer m can be represented as a sum of two elements in B in more than $.2n$ ways (we do not count permutations) than $m \in B' + B'$. To conclude, notice that every m in the interval $[.4n + 1, 1.6n - 1]$ has more than $.2n$ representations. □

To prove Lemma 2.5, we use induction on d. The harder part of the proof is to handle the base case $d = 1$. To handle this case we apply the tree method mentioned in the introduction.

Without loss of generality we can assume that g is a power of 4, $|X_i| = n_1$ and $0 \in X_i$ for all $1 \le i \le g$. Let m be the cardinality of P; we can also assume that P is the interval $[m]$. Set $X_i^1 = X_i$ for $i = 1, \dots, g$ and $g_1 = g$. Here is the description of the algorithm we would like to study.

The algorithm. At the t^{th} step, the input is a sequence $X_1^t, \dots, X_{g_t}^t$ of sets of the same cardinality n_t where g_t is an even number. Choose a pair $1 \le i < j \le g_t$ which maximizes $|X_i^t + X_j^t|$ (if there are many such pairs choose an arbitrary one). Denote the sum $X_i^t + X_j^t$ by X_1'. Remove i and j from the index set and repeat the operation to obtain X_2' and so on. After $g_t/2$ operations we obtain a set sequence $X_1', \dots, X_{g_t/2}'$ which has decreasing cardinalities. Define $g_{t+1} = g_t/4$. Consider the sequence $X_1', \dots, X_{g_{t+1}}'$ and truncate all but the last set so that all of them have the same cardinality (which is $|X_{g_{t+1}}'|$). The truncated sets will be named $X_1^{t+1}, \dots, X_{g_{t+1}}^{t+1}$ and they form the input of the next step. The algorithm halts when the input sequence has only one element. A simple calculation shows that $g_t = \frac{1}{4^{t-1}} g_1$ for all possible t's.

Notice that $X_{g_t}^t$ is a subset of $2^t P$ and thus $n_t = |X_{g_t}^t|$ is at most $2^t m$. On the other hand,

$$n_1 = |X_{g_1}^1|/m \ge \gamma.$$

So, for some $t \le \log_{1.05} \frac{1}{\gamma}$, $n_{t+1} \le 2.1 n_t$. By the description of the algorithm, there are $g_t/2$ sets among the X_i^t such that every pair of them have cardinality at most $n_{t+1} \le 2.1 n_t$. To simplify the notations, call these sets $Y_1, \dots, Y_h$. We have that

$$h = g_t/2 \ge \frac{1}{4^t} g_1.$$

So, by increasing g_1 we can assume that h is sufficiently large, whenever needed.

We have that $|Y_i| = n_t$ and $|Y_i + Y_j| \leq 2.1 n_t$ for all $1 \leq i < j \leq h$. We are now in position to invoke an inverse statement and at this stage all we need is Lemma 3.1 (which is much simpler than Freiman's general theorem). By this lemma, every Y_i is a subset of an arithmetic progression A_i of length at most $1.1 n_t$. Moreover, A_i is a subset of $2^t P$. Also observe that by the definition of t, $n_t/|2^t P| \geq \gamma$.

We can extend the A_i's obtained prior to Lemma 3.2 so that each of them has length exactly $1.1 n_t$. By Lemma 3.2, provided that g_t is sufficiently large, there are A_i and A_j such that $B = A_i \cap A_j$ is an arithmetic progression of length at least n_t. Now consider Y_i and Y_j which are subsets of A_i and A_j, respectively. Since Y_i and Y_j both have n_t elements, $B' = Y_i \cap Y_j \cap B$ has at least $.8 n_t$ elements.

The set $B' + B'$ is a subset of $Y_i + Y_j$, which, in turn, is a subset of $X_1 + \cdots + X_g$ (recall that we assume $0 \in X_i$ for every i). This and Lemma 3.3 complete the proof for the base case $d = 1$.

Now assume that the hypothesis holds for all $d \leq r$; we are going to prove it for $d = r + 1$. This proof uses a combinatorial counting argument and is independent of the previous proof. In particular, we do not need the tree method here.

Consider a generalized arithmetic progression P of rank $r + 1$ and its canonical decomposition $P = P_1 + P_2$, where P_1 is an arithmetic progression and P_2 is a generalized arithmetic progression of rank r (P_1 is the first "edge" of P). For every $x \in P_2$, denote by $P_1^i(x)$ the set of those elements y of P_1 where $x + y \in X_i$. We say that x is *i-normal* if $P_1^i(x)$ has density at least $\gamma/2$ in P_1. Since $|X_i| \geq \gamma \mathrm{Vol}(P)$, the set N_i of i-normal elements has density at least $\gamma/2$ in P_2, for all possible i.

Let $g = g' g''$ where g' and g'' are large constants satisfying $g'' \gg g' \gg 1/\gamma$. Partition $X_1, \ldots, X_g$ into g'' groups with cardinality g' each. Consider the first group. Without loss of generality, we can assume that its members are $X_1, \ldots, X_{g'}$ and also that $|N_1| = \cdots = |N_{g'}| = \gamma |P_2|/2$. Order the elements in each N_i increasingly. For each $1 \leq k \leq |N_1|$, let $x_1^k, \ldots, x_{g'}^k$ be the k^{th} elements in $N_1, \ldots, N_{g'}$, respectively. Consider the sets $P_1(x_i^k), P_1(x_2^k), \ldots, P_1(x_{g'}^k)$. Given that g' is sufficiently large, we can apply the statement for the base case $d = 1$ to obtain an arithmetic progression A_k of length $\gamma_1 |P_1|$, for some positive constant γ_1 depending on γ. Each of the A_k, $k = 1, 2, \ldots, |N_1|$, is a subset of $g' P_1$ which has length $g' |P_1|$ (to be exact, the length of $g' P_1$ is $g' |P_1| + O(1)$; but since the error term $O(1)$ plays no role, we omit it here and later to simplify the presentation), so the density of each A_k in $g' P_1$ is γ_1 / g'. Applying Lemma 3.2 with $n = g' |P_1|$ and $\varepsilon = \gamma_1 / g'$, a $.09(\gamma_1/g')^2$ fraction of the A_k's contain the same arithmetic progression B of length $.9 \gamma_1 |P_1|$. Without loss of generality, we can assume $A_1, \ldots, A_L$, where

$$L = .09(\gamma_1/g')^2 |N_1| = .09(\gamma_1/g')^2 \gamma |P_2|/2,$$

all contain B. Let Y_1 be the collection of the sums $x_k = x_1^k + \cdots + x_{g'}^k$, $1 \le k \le L$. By the ordering, all x_k's are different so $|Y_1| = L$ and thus Y_1 has density

$$L/g'|P_2| = .09(\gamma_1/g')^2(\gamma/2g')$$

in $g'P_2$. Moreover, the set $Y_1 + B_1$ is a subset of $X_1 + \cdots + X_{g'}$.

Next, by considering the second group, we obtain $Y_2 + B_2$ and so on. Now we focus on the sets $Y_1 + B_1, \ldots, Y_{g''} + B_{g''}$. Each B_j is an arithmetic progression in $g'P_1$ with density

$$.9\gamma_1|P_1|/g'|P_1| = .9\gamma_1/g'.$$

By Lemma 3.2, at least a

$$.09(.9\gamma_1/g')^2 \ge .07(\gamma_1/g')^2$$

fraction of the B_j's contain the same arithmetic progression C of length

$$.9(.9\gamma_1|P_1|) \ge .8\gamma_1|P_1|.$$

Without loss of generality, we can assume that $B_1, \ldots, B_{g'''}$ contain C, where

$$g^{'''} = .08(\gamma_1/g')^2 g^{''}.$$

By setting $g^{''}$ sufficiently large compared to $g^{'}$, we can assume that $g^{'''}$ is sufficiently large.

Now we are in position to conclude the proof. As $Y_1, \ldots, Y_{g'''}$ have density at least $.09(\gamma_1/g')^2(\gamma/2g')$ in $g'P_2$, for a sufficiently large $g^{'''}$, $Y_1 + \cdots + Y_{g'''}$ contains a generalized arithmetic progression D of rank r of constant density in $g^{'''}(g'P_2)$, due to the induction hypothesis. The set $C + D$ is a generalized arithmetic progression of rank $r+1$ with positive constant density in $g^{'''}(g'P)$. On the other hand, this generalized arithmetic progression is a subset of $(Y_1 + C) + \cdots + (Y_{g'''} + C)$. As we assumed $0 \in X_i$ for $1 \le i \le g$, the sum $(Y_1 + C) + \cdots + (Y_{g'''} + C)$ is a subset of $X_1 + \cdots + X_g$, completing the proof. $\square$

4. Proof of Lemma 2.6

This proof is relatively long and we break it into several parts. In the first subsection, we present two lemmas. The next subsection contains the description of an algorithm (again we use the tree method), which is somewhat more involved than the one used in the proof of Lemma 2.5. In the third subsection, we analyze this algorithm and construct the desired sets A' and W. The fourth and final subsection is devoted to the verification of a technical statement which we need in order to show that W has the properties claimed by the lemma.

4.1. *Two simple lemmas.* The first lemma is a simple result from graph theory.

LEMMA 4.1. *Let G be a graph with vertex set V. If $|V| \geq K^2 - K$ and G does not contain an independent set of size K then there is a vertex with degree at least $|V|/K$.*

Proof of Lemma 4.1. Let I be an independent set with maximum cardinality. By the assumption of the lemma $|I| \leq K - 1$. Since I has maximum cardinality, for any vertex $a \in V \backslash I$ there must be a vertex $b \in I$ such that (a, b) is an edge (otherwise $I \cup \{a\}$ would be a larger independent set). Thus, there must be at least $|V| - |I|$ edges with one end point in I and the other end point in $V \backslash I$. Therefore, there is a vertex in I with degree at least

$$\frac{|V| - |I|}{|I|} \geq \frac{|V| - (K-1)}{K-1} \geq \frac{|V|}{K},$$

where in the last inequality we used the assumption $|V| \geq K^2 - K$. □

LEMMA 4.2. *Any set A with $\Omega(n^{1/2})$ elements has a subset A' with $O(\log n)$ elements such that $|S_{A'}| = \Omega(n^{1/2})$.*

Proof of Lemma 4.2. We find A' by the greedy algorithm. We choose the first element x_1 of A' arbitrarily. Assume that $x_1, \ldots, x_i$ have been chosen. We denote by S_i the sumset $S_{\{x_1,\ldots,x_i\}}$ and s_i its cardinality. We choose x_{i+1} from $A \backslash \{x_1, \ldots, x_i\}$ to maximize $s_{i+1} = |S_{\{x_1,\ldots,x_{i+1}\}}|$ (ties are broken arbitrarily). If $s_{i+1} \leq 1.1 s_i$ then $x_{i+1} + S_i$ and S_i should have at least $.9 s_i$ elements in common. Since x_{i+1} was chosen optimally, we have that

$$|S_i - S_i| \geq .9 s_i |A \backslash \{x_1, \ldots, x_i\}|.$$

Since $|S_i - S_i| \leq s_i^2$, $s_i \geq .9|A \backslash \{x_1, \ldots, x_i\}|$. Let $A' = \{x_1, \ldots, x_i\}$, where i is the first index satisfying either $s_{i+1} \leq 1.1 s_i$ or $|S_{A'}| \geq n^{1/2}$. The definition of i and the above calculation show that A' satisfies the claim of Lemma 4.2. □

Remark 4.3. With a small modification, we can have A' such that $|A'| = O(\log n)$ and $|l^*A'| = \Omega(n^{1/2})$, where $l = |A'|/2$ and l^*A' denotes the collection of sum of l different elements from A'.

Fix a small positive constant ε (say 1/100) and let T be the first integer such that $(1/2 - \varepsilon)^T \leq \frac{\log n}{n^{1/2}}$. One can find a positive constant K (depending on ε) such that

$$K^{3T/4} \geq n^{11/10}. \tag{4}$$

Using Lemma 4.2 iteratively one can produce mutually disjoint subsets $A'_1, \ldots, A'_m$ of A with the following properties: $|A'_i| = O(\log n)$, $m = \Omega(n/\log n)$, $|S_{A'_i}| = \Omega(n^{1/2})$ and $|\cup_{i=1}^m A'_i| \leq |A|/2$. We denote by A_1, A_2, and B_i the sets $\cup_{i=1}^m A'_i$, $A \backslash A_1$, and $S_{A'_i}$, respectively.

In what follows, we assume that S_A does not contain an arithmetic progression of length n. Our proof has two main steps. In the first step, we create a generalized arithmetic progression P with constant rank and small volume which contains a positive constant fraction of A_2. In the second step, we use P to construct the required generalized arithmetic progression W.

4.2. *The algorithm.* We are going to apply the tree method and this subsection is devoted to the description of the algorithm. To start, set $m_0 = m$. Truncate the B_i's so each of them has exactly $b_0 = \alpha n^{1/2}$ elements, for some positive constant α. Denote by B_i^0 the truncation of B_i. We start with the sequence $B_1^0, \dots, B_{m_0}^0$, each element of which has exactly b_0 elements. Without loss of generality, we may assume that m_0 is even. At the beginning, the elements in A_2 are called *available*.

A general step of the algorithm functions as follows. The input is a sequence $B_1^t, \dots, B_{m_t}^t$ of sets of the same cardinality b_t. Consider the sets $B_i^t + B_j^t$ for all possible pairs i and j. Choose i and j where the sum has maximum cardinality (if there are many pairs, order them lexicographically and choose the first one — the order is not important at all, our only goal is to make the operation well-defined). Next, choose $x_1, \dots, x_K$ from the set of available elements so that

$$B_1' = (B_i^t + B_j^t) \cup \Big(\cup_{i=1}^K (B_i^t + B_j^t + x_i) \Big)$$

has maximum cardinality (we break ties as above). Remove i and j from the index set and the x_i's from the available set and repeat the operation to obtain B_2' and so on. We end up with a set sequence $B_1', \dots, B_{m_t/2}'$ where $|B_1'| \geq \dots \geq |B_{m_t/2}'|$.

Let m_{t+1} be the largest even integer not exceeding $(1-\varepsilon)m_t/2$ and set $b_{t+1} = |B_{m_{t+1}}'|$. Truncate the B_i''s ($i < m_{l+1}$) so that the remaining sets have exactly b_{t+1} elements each. Denote by B_i^{t+1} the remaining subset of B_i'. The sequence $B_1^{t+1}, \dots, B_{m_{t+1}}^{t+1}$ is the output of the step.

If $m_{t+1} \geq 3$, then we continue with the next step. Otherwise, the algorithm terminates.

We would like to say a few words about how to exploit this algorithm to our advantage. By the description of the algorithm

$$B_{m_k}^k = (B_i^{k-1} + B_j^{k-1}) \cup \Big(\cup_{h=1}^K (B_i^{k-1} + B_j^{k-1} + x_h) \Big) \tag{5}$$

for some i, j and x_h's. We are going to show that there is some step k where $|B_{m_k}^k|$ is bounded by $a|B_i^{k-1}|$, for some constant a. This enables us to apply Freiman's theorem to get information about B_i^{k-1} and $B_{m_k}^k$. Furthermore, we can show that there is some overlap among the sets $(B_i^{k-1} + B_j^{k-1} + x_h)$ $(h = 1, \dots, K)$, since otherwise their union would be too large. Thanks to this information and also the fact that we choose the x_h in an optimal way, we can derive some properties of the set of available elements. The desired sets A' and W will be constructed from the set of available elements using this property.

Before starting the analysis of the algorithm, let us pause for a moment and make some simple observations:

- B_i^t is a subset of S_A (more precisely a subset of S_{A_1}) for any possible t and i.
- The maximum element in B_i^t is at most $(2^{t+1}-1)n$ (induction).
- For any possible t, $b_{t+1} \geq 2b_t$.
- At each step, the length of the sequence shrinks by a factor $1/2-\varepsilon$, so the algorithm terminates after $T' = (1-o(1))T$ steps.
- The number of elements x_i used in the algorithm is $O(n^{1/2}/\log n)$, so at any step, there are always $(1-o(1))|A_2|$ available elements.

Now comes an important observation

Fact 4.4. *There is an index $k \leq \frac{3}{4}T$ such that $b_k \leq K^k b_0$.*

Proof of Fact 4.4. As S_A is a subset of $[cn^{3/2}]$ for some constant c, $b_k = O(n^{3/2})$. On the other hand, the definition of K implies

$$K^{3T/4}b_0 = \Omega(K^{3T/4}n^{1/2}) \gg n^{3/2},$$

proving the claim. □

4.3. *Finding A' and W.* Let k be the first index where $b_k \leq K^k b_0$. This means $|B^k_{m_k}| \leq K^k b_0$. By the description of the algorithm

$$B^k_{m_k} = (B_i^{k-1} + B_j^{k-1}) \cup \left(\cup_{h=1}^{K} (B_i^{k-1} + B_j^{k-1} + x_h)\right) \tag{6}$$

for some i, j and x_h's. This implies that

$$|B^k_{m_k}| \geq |B_i^{k-1} + B_j^{k-1}| \tag{7}$$

where $1 \leq i < j \leq m_{k-1}$ and both B_i^{k-1} and B_j^{k-1} have cardinality $b_{k-1} \geq K^{k-1}b_0$. The definition of k then implies that $|B^k_{m_k}| \leq K b_{k-1}$, so

$$|B_i^{k-1} + B_j^{k-1}| \leq K|B_i^{k-1}|. \tag{8}$$

Applying Freiman's theorem to (8), we can deduce that there is a generalized arithmetic progression R with constant rank containing B_i^{k-1} and $\mathrm{Vol}(R) = O(|B_i^{k-1}|) = O(b_{k-1})$.

We say that two elements u and v of B_j^{k-1} are equivalent if their difference belongs to $R - R$. If u and v are not equivalent then the sets $u + B_i^{k-1}$ and $v + B_i^{k-1}$ are disjoint, since B_i^{k-1} is a subset of R. By (8), the number of equivalence classes is at most K. Let us denote these classes by $C_1, \dots, C_K$, where some of the C_s's might be empty. We have $B_i^{k-1} \subset R$ and $B_j^{k-1} \subset \cup_{s=1}^K C_s$.

Let us now take a close look at (6). The assumption $|B_{m_k}^k| \le K|B_i^{k-1}|$ and (6) imply that there must be a pair s_1, s_2 such that the intersection

$$(B_i^{k-1} + B_j^{k-1} + x_{s_1}) \cap (B_i^{k-1} + B_j^{k-1} + x_{s_2})$$

is not empty. Moreover, the set $\{x_1, \dots, x_K\}$ in (6) was chosen optimally. Thus, for any set of K available elements, there are two elements x and y such that the intersection $(B_i^{k-1} + B_j^{k-1} + x) \cap (B_i^{k-1} + B_j^{k-1} + y)$ is not empty. This implies

(9)
$$x - y \in (B_i^{k-1} + B_j^{k-1}) - (B_i^{k-1} + B_j^{k-1}) \subset \cup_{1 \le g,h \le K}\Big((R + C_g) - (R + C_h)\Big).$$

Define a graph G on the set of available elements as follows: x and y are adjacent if and only if

$$x - y \in (B_i^{k-1} + B_j^{k-1}) - (B_i^{k-1} + B_j^{k-1}).$$

By the argument above, G does not contain an independent set of size K, so by Lemma 4.1 there should be a vertex x with degree at least $|V(G)|/K$. (Here K is a constant so the condition $|V| \ge K^2 - K$ holds trivially.) By (9) and the pigeon hole principle, there is a pair (g, h) such that there are at least $|V(G)|/K^3$ elements y satisfying

$$x - y \in (R + C_g) - (R + C_h). \tag{10}$$

Both C_g and C_h are subsets of translates of R; so the set Y of the elements y satisfying (10) is a subset of a translate of $P = (R + R) - (R + R)$. Recall that at any step, the number of available elements is $(1 - o(1))|A_2|$, we have

$$|Y| \ge (1 - o(1))|A_2|/K^3 = \Omega(|A_2|). \tag{11}$$

Claim 4.5. *There is a generalized arithmetic progression U of rank two such that $|U \cap P| = \Omega(\mathrm{Vol}(P))$ and $\mathrm{Vol}(U) = O(n^{1/2} \log^\beta n)$, for some positive constant β.*

Assuming Claim 4.5, we conclude the proof of Lemma 2.6 as follows. We say that two elements in P are equivalent if their difference belongs to $U - U$. If x and y are not equivalent, then $x + (U \cap P)$ and $y + (U \cap P)$ are disjoint subsets of $P + P$. Since $\mathrm{Vol}(P + P) = O(\mathrm{Vol}(P))$, the condition $|U \cap P| = \Omega(\mathrm{Vol}(P))$ implies that the number of equivalence classes is bounded by a constant. So, there is an equivalence class whose intersection with A_2 has cardinality $\Omega(|A_2|)$. On the other hand, there is a translate W of $U - U$ containing this class. As

$\mathrm{Vol}(U) = O(n^{1/2}\log^{\beta} n)$ and U has rank two, W is also a generalized arithmetic progression of ranks 2 and volume $O(n^{1/2}\log^{\beta} n)$, as required by Lemma 2.6.

4.4. *Proof of Claim* 4.5. Let us go back to the definition of $B^k_{m_k}$ (see (6)). When we define $B^k_{m_k}$, we choose i and j to maximize the cardinality of $B^{k-1}_i + B^{k-1}_j$. On the other hand, as $m_k \le (1/2-\varepsilon)m_{k-1}$, for any remaining index i, we have at least $l = 2\varepsilon m_{k-1}$ choices for j. This means that there are l sets $B^{k-1}_{j_1}, \dots, B^{k-1}_{j_l}$, all of the same cardinality b_{k-1}, such that

$$|B^{k-1}_i + B^{k-1}_{j_r}| \le |B^{k-1}_i + B^{k-1}_j| \le K b_{k-1} \tag{12}$$

for all $1 \le r \le l$.

From now on, we work with the sets $B^{k-1}_{j_r}$, $1 \le r \le l$. By considering equivalence classes (as in the paragraph following (8)), we can show that for each r, $B^{k-1}_{j_r}$ contains a subset D_r which is a subset of a translate of R and $|D_r| \ge |B^{k-1}_{j_r}|/K = \Omega(\mathrm{Vol}(R))$. The sum of all D_r's is a subset of S_A.

By Lemma 2.5, there is a constant g such that $D_1 + \dots + D_g$ contains a generalized arithmetic progression Q_1 with cardinality at least $\gamma\mathrm{Vol}(R)$ for some positive constant γ. Using the next g D_i's, we can create Q_2 and so on. At the end, we have $l_1 = \lfloor l/g \rfloor$ generalized arithmetic progression $Q_1, \dots, Q_{l_1}$. Each of these has rank $d = rank(R)$ and cardinality at least $\gamma\mathrm{Vol}(R)$. Moreover, they are subsets of translates of the generalized arithmetic progression $R' = gR$ which also has volume $O(\mathrm{Vol}(R))$.

There are only $O(1)$ possibilities for the difference sets of the Q_i. Thus, there is a positive constant γ_1 such that at least a γ_1 fraction of the Q_i's has the same difference set. Consequently, there is a generalized arithmetic progression Q (of rank d and cardinality at least $\gamma\mathrm{Vol}(R)$) and an integer $l_2 = \Omega(l_1)$ so that there are least l_2 translates of Q among the Q_i's. (To be more precise, there are l_2 among the Q_i's which contains a translate of Q. We can truncate these Q_i's so that they equal a translate of Q.) Without loss of generality, we can assume that $Q_1, \dots, Q_{l_2}$ are translates of Q.

Next, we investigate the sets $Q_1, \dots, Q_{l_2}$. Their sum is clearly a translate of $l_2 Q$. Moreover, this sum is a subset of S_A. Thus, S_A contains a translate of $l_2 Q$.

Define a sequence $T_0 = Q$, $T_{i+1} = 2T_i$. Let i_0 be the first i such that $|T_{i+1}| \le 7|T_i|$. (The argument below shows that i_0 exists.) A combination of Lemma 2.2 and Lemma 2.5 implies that there is a constant h such that hT_i contains a generalized arithmetic progression U_0 of rank 2 where

$$|U_0| = \Omega(|T_i|) = \Omega(7^{i_0}|Q|).$$

Using the equivalence class argument, we can show that there is a translate U of $U_0 - U_0$ such that

$$|U \cap T_0| = |U \cap Q| = \Omega(|Q|).$$

Now, let us take a close look at l_2 and Q. Following the calculation, we see that

(13)
$$l_2 = \Omega(l_1) = \Omega(l) = \Omega(m_{k-1}) \geq \Omega((1/2-\varepsilon)^{k-1} m_0) \geq \varepsilon_1 (1/2-\varepsilon)^{k-1} \frac{n^{1/2}}{\log n},$$

for some positive constant ε_1. Furthermore,

(14) $$|Q| = \Omega(\mathrm{Vol}(R)) = \Omega(b_{k-1}) = \Omega(K^{k-1} b_0) = \Omega(K^{k-1} n^{1/2}).$$

Equation (14) implies that

(15) $$|U_0| = \Omega(7^{i_0}\mathrm{Vol}(Q)) \geq \varepsilon_2 K^{k-1} 7^{i_0} n^{1/2},$$

for some positive constant ε_2.

Observe that Q can be viewed (after a proper translation) as a subset of $[g2^{k+1}n]$ for some constant g. Indeed, Q is contained in the sum $D_1 + \cdots + D_g$ and each D_j is a subset of some $B^{k-1}_{i_j}$, which, in turn, is a subset of $[2^{k+1}n]$. Thus U_0 is a subset of the interval $[2^{i_0}hg2^{k+1}n]$. Moreover, as S_A contains a translate of $l_2 Q$, S_A contains a translate of $\frac{l_2}{2^{i_0}h} U_0 = l_3 U_0$, where

(16) $$l_3 = \frac{l_2}{2^{i_0}h} \geq \frac{\varepsilon_1}{2^{i_0}h}(1/2-\varepsilon)^{k-1}\frac{n^{1/2}}{\log n}.$$

Let us consider two cases:

(i) The product of the right-most formulae in (16) and (15) is at least $20(2^{i_0}hg2^{k+1}n)$.

In this case $l_3|U_0|$ satisfies the condition of Lemma 2.4 with $m = 2^{i_0}hg2^{k+1}n$. Therefore $l_3 U_0$ contains a arithmetic progression of length $m > n$. As a translate of $l_3 U_0$ is a subset of S_A, it follows that S_A contains an arithmetic progression of length n, a contradiction.

(ii) The product is less than $20(2^{i_0}hg2^{k+1}n)$. This implies that

$$\frac{\varepsilon_1\varepsilon_2}{h}(\frac{7}{4})^{i_0}(\frac{K}{2}(\frac{1}{2}-\varepsilon))^{k-1}\frac{n}{\log n} \leq 80hgn.$$

It follows that $\frac{1}{\log n}(\frac{7}{4})^{i_0}(\frac{K}{2}(\frac{1}{2}-\varepsilon))^{k-1}$ is upper bounded by the constant $\frac{80h^2g}{\varepsilon_1\varepsilon_2}$. We choose K sufficiently large so that $\frac{K}{2}(\frac{1}{2}-\varepsilon) > 1$; this implies that $(\frac{7}{4})^{i_0} = O(\log n)$. Thus there is a positive constant β such that $(2^d)^{i_0} \leq \log^\beta n$, where d is the rank of P. Now let us bound $\mathrm{Vol}(U_0)$. It is clear that

$$\mathrm{Vol}(U_0) \leq (2^d)^h \mathrm{Vol}(T_{i_0}) \leq (2^d)^h (2^d)^{i_0}\mathrm{Vol}(P) = \Theta((2^d)^{i_0}\mathrm{Vol}(P)).$$

Taking (14) into account, we deduce that

(17) $$\mathrm{Vol}(U_0) = \Theta((2^d)^{i_0}\mathrm{Vol}(P)) = O(n^{1/2}\log^\beta n).$$

As $\mathrm{Vol}(U) = O(\mathrm{Vol}(U_0))$, the proof of Claim 4.5 is complete. □

5. Proof of Theorem 1.1

A rough description of our plan is the following. We first use Lemma 2.6 to find a large set B whose elements can be represented as a sum of two elements of A in many ways. In the second step, we use the elements of B to construct a large generalized arithmetic progression of rank 2. (See the paragraph following Lemma 2.4 for an explanation why a large generalized arithmetic of rank 2 is all we need.)

The following definition plays an important role in the proof.

Definition 5.1. A number x has multiplicity m with respect to a set A if x can be represented as a sum of two different elements of A in at least m ways. A set B has multiplicity m with respect to A if every element of B has multiplicity m with respect to A.

The reader might wonder why a set B with high multiplicity is useful. In the next few sentences we try to give a quick explanation. Consider a set B with multiplicity m and a sum $s = b_1 + \cdots + b_l$, where $b_i \in B$ and $l \leq m/2$. We claim that one can write s as a sum of different elements of A. We show this by induction on l. Trivially there are two different elements a_1 and $a_{1'}$ of A such that $b_1 = a_1 + a_{1'}$. Assume that

$$b_1 + \dots b_r = (a_1 + a_{1'}) + \cdots + (a_r + a_{r'}),$$

where the elements on the right-hand side are all different and $r + 1 \leq m/2$. Consider $b_1 + \cdots + b_r + b_{r+1}$. Notice that for any $i \leq r$, each of the two numbers a_i and $a_{i'}$ appear in at most one representation of b_{r+1}. Thus, there are at most $2r$ representations of b_{r+1} which we cannot use. Since $2r < m$, there is a good representation left.

The above argument allows us to consider the sumset lB and not have to worry about using the same element in a sum many times. As we pointed out in the introduction, it is much more convenient when one allows repetitions in the sum.

Let A be a subset of $[n]$ with at least $cn^{1/2}$ elements, where c is a sufficiently large constant. We assume (for a contradiction) that S_A does not contain an arithmetic progression of length n. By Lemma 2.6, there is a generalized arithmetic progression P with constant rank 2 such that $A_1 = P \cap A$ has constant density α in A and P has volume at most $n^{1/2} \log^{\beta} n$, for some constant β. Here neither α nor β depends on c, so by increasing c we can assume that $|A_1| \geq c_1 n^{1/2}$, where c_1 is still a sufficiently large constant. We are going to show that S_{A_1} contains an arithmetic progression of length n, which is a contradiction, as A_1 is a subset of A.

The rest of this section is organized as follows. In the first subsection we find a set B with high multiplicity. By the above argument, we can conclude that lB is a subset of S_A, for some large number l. This number l has the form $l = \frac{n^{1/2}}{4t\log t}$ where t is a parameter to be defined. It is important for the rest of the proof that we can assume t is sufficiently large. In the second subsection, we are going to show why this assumption is legitimate. A further consideration in this subsection shows that beside being large, t has some other useful properties.

One can show that lB contains a large generalized arithmetic progression of rank 2. However, this generalized arithmetic progression is still not large enough to allow Lemma 2.4 to be invoked. We shall use the so-called tiling argument (mentioned in the Introduction) to tile several translates of this generalized arithmetic progression to obtain a much larger generalized arithmetic progression (for which Lemma 2.4 works). The tiling argument is technical and we break it into two subsections. In the first one, we consider a simplified scenario so the reader can quickly grasp the idea. The treatment of the general case follows next. The fifth, and final, subsection is devoted to a construction showing the sharpness of Theorem 1.1.

5.1. *Defining B.* Denote by M_k the set of numbers whose multiplicities with respect to A_1 lie between $\frac{n^{1/2}}{2^k k}$ and $\frac{n^{1/2}}{2^{k+1}(k+1)}$, for all $k = 1, 2, 3, \dots, \lfloor \log n^{1/2} \rfloor$ (we may assume that $n^{1/2}$ is an irrational number to avoid possible overlaps). It is clear that M_k is subset of $A + A \subset 2P$ so

$$|M_k| \le \mathrm{Vol}(2P) \le 4n^{1/2} \log^\beta n$$

for all k. Moreover,

$$\sum_{k=1}^{\lfloor \log n^{1/2} \rfloor} \frac{n^{1/2}}{2^k k} |M_k| \ge \binom{|A_1|}{2} \ge \binom{c_1 n^{1/2}}{2}.$$

The total contribution from those k's where $2^k k \ge \log^{2+\beta} n$ is at most

$$\frac{n^{1/2}}{\log^{2+\beta} n} (4n^{1/2} \log^\beta n) \log n = o(n).$$

So

$$\sum_{k=1}^{\lfloor \log \log^{2+\beta} n \rfloor} \frac{n^{1/2}}{2^k k} |M_k| \ge \binom{|A_1|}{2} \ge (1 - o(1)) c_1^2 n/2, \tag{18}$$

which implies that there is an index k between 1 and $\lfloor \log \log^{2+\beta} n \rfloor$ such that $|M_k| \ge \frac{c_2 n^{1/2} 2^k}{k}$, where $c_2 = \frac{c_1^2}{3} (\sum_{k=1}^\infty \frac{1}{k^2})^{-1}$ (if there are many choose the largest k). Rename this particular set M_k to B and set $t = 2^k$. This is the set B we look for. The elements of B have multiplicity at least

$$\frac{n^{1/2}}{2^{k+1}(k+1)} \ge \frac{n^{1/2}}{4t \log_2 t} = l$$

with respect to A_1, so lB is a subset of S_{A_1}. Moreover $|lB| = O(n^{3/2})$ since lB is a subset of S_{A_1} and A_1 is a set of $O(n^{1/2})$ numbers not exceeding n. Without loss of generality, we can assume that l is a power of 2.

In the rest of the proof we shall need the assumption that t is bounded below by a large constant. In the next subsection, we are going to show this assumption is legitimate.

5.2. *A consideration of t.* If $t > \log n$, then we are done since n is arbitrarily large; so, we assume that $t \leq \log n$. Let $B_0 = B$ and $B_{i+1} = 2B_i$. Let $\gamma_i = |B_i|/|B_{i-1}|$ and s be the first index where $\gamma_s \leq 7$. A simple calculation shows that $(2.1)^s < l$ since otherwise $|lB| \gg n^{3/2}$, a contradiction. By Lemma 2.2, B_s is a subset of a generalized arithmetic progression Q of rank 2 and $|B_s| \geq \alpha \mathrm{Vol}(Q)$ for some positive constant α. Lemma 2.5 implies that there is a constant g such that $2^g B_s$ contains a generalized arithmetic progression Q' of rank 2 and cardinality at least $\alpha'|B_s|$, where α' is another positive constant. Moreover, as $(2.1)^s < l$ and $t \leq \log n$, $l/2^s = \omega(1)$ so $l/2^s > 2^g$. Thus $\frac{l}{2^{s+g}} B_{s+g}$ is a subset of S_{A_1} and so is $\frac{l}{2^{s+g}} Q'$. We next want to apply Lemma 2.4. In order to verify the conditions of this lemma, let us consider the product $\frac{l}{2^{s+g}}|Q'|$. We have

$$\frac{l}{2^{s+g}}|Q'| \geq \alpha' \frac{l}{2^{s+g}}|B_s| \geq \frac{\alpha'}{2^g}\left(\frac{7}{2}\right)^s l|B_0|, \tag{19}$$

where in the last inequality we used the fact that $|B_s| \geq 7^s|B_0|$ which is a consequence of the definition of s. As $|B_0| = |B| = |M_k| \geq \frac{c_2 n^{1/2} t}{\log t}$ and $l = \frac{n^{1/2}}{4t \log t}$, we see that

$$l|B_0| \geq \frac{c_2 n}{4 \log^2 t}$$

and

$$\frac{l}{2^{s+g}}|Q'| \geq \frac{\alpha'}{2^g}\left(\frac{7}{2}\right)^s \frac{c_2 n}{4 \log^2 t}. \tag{20}$$

Notice that Q' is a subset of the interval $[2^{s+g} n]$. So if $\frac{\alpha'}{g}(\frac{7}{2})^s \frac{c_2 n}{4 \log^2 t} \geq 20(2^{s+g} n)$ then by Lemma 2.4 $\frac{l}{2^{s+g}}|Q'|$ contains an arithmetic progression of length $2^{s+g} n$ $> n$, a contradiction. Thus

$$\frac{\alpha'}{2^g}\left(\frac{7}{2}\right)^s \frac{c_2 n}{4 \log^2 t} \leq 20 \times 2^{s+g} n,$$

which implies that

$$\frac{\alpha'}{80g}\frac{c_2}{4^g} \leq \log^2 t.$$

By increasing c_2 (the constants α' and g do not depend on c_2) we can assume that t is sufficiently large, whenever needed. In particular, we may assume that $t \geq \log^{300} t \gg 1$.

The rest of the proof of Theorem 1.1 splits into two cases. The first and easy case is when $\gamma_1 \dots \gamma_s$ is relatively large.

Case 1. $\log^3 t \le \gamma_1 \dots \gamma_s 5^{-s}$. In this case

$$|B_s| \ge \gamma_1 \dots \gamma_s |B_0| \ge 5^s (\log^3 t)|B_0| \ge 5^s \log^3 t \frac{n^{1/2} t}{\log t} = 5^2 n^{1/2} t \log^2 t. \tag{21}$$

The analysis of this case is similar to the argument we just presented. Consider the set Q' as above. We have

$$\frac{l}{2^{s+g}} |Q'| \ge \frac{\alpha'}{2^g} \frac{l}{2^s} |B_s|. \tag{22}$$

By (21) and the fact that $l = \frac{n^{1/2}}{4t \log t}$ the right-hand side of (22) is at least

$$\frac{\alpha'}{2^g} \frac{n^{1/2}}{4t \log t} (\frac{5}{2})^s n^{1/2} t \log^3 t \ge \frac{\alpha' \log t}{4^g} (\frac{5}{2})^s 2^g n. \tag{23}$$

Provided that t is sufficiently large, we have $\frac{\alpha' \log t}{4^g} \ge 20$. Thus the right-hand side of (23) is at least $20(2^{s+g} n)$, which implies that $\frac{l}{2^{s+g}} |Q'| \ge 20(2^{s+g} n)$. Similar to the previous proof, we can conclude that $\frac{l}{2^{s+g}} Q'$ contains an arithmetic progression of length $20(2^{s+g} n) > n$, a contradiction. This completes the analysis of the first case.

Case 2. $\log^3 t \ge \gamma_1 \dots \gamma_s 5^{-s}$. Recall that by Lemma 2.2, B_s is a subset of constant density of a generalized arithmetic progression P of rank 2. The condition $\log^3 t \ge \gamma_1 \dots \gamma_s 5^{-s}$ and the fact that all $\gamma_i > 7$ together imply that $\gamma_1 \dots \gamma_s \le \log^6 t$. Thus B is a subset of density

$$\Omega(\frac{1}{\gamma_1 \dots \gamma_s}) = \Omega(\frac{1}{\log^6 t})$$

of P. This information will be critical in the rest of the proof.

The remaining arguments of the proof are somewhat easier to verify with a geometrical visualization. For that purpose, we introduce the following map. Assume that $P = \{x_1 a_1 + x_2 a_2 | 0 \le x_i \le l_i\}$, Φ is a map which maps P onto $\mathbb{Z}^2$ as follows

$$\Phi : (x_1 a_1 + x_2 a_2) \to (x_1, x_2).$$

We would like to emphasize here that Φ does take into account the structure of P. If we view P as a set of integers, Φ is not an one-to-one map. If the same number x has two different representations $x = x_1 a_1 + x_2 a_2 = x_1' a_1 + x_2' a_2$, then $\Phi(x)$ contains both (x_1, x_2) and (x_1', x_2'). Φ^{-1} maps $\mathbb{Z}^2$ to $\mathbb{Z}$ as follows

$$\Phi^{-1}(x, y) \to (x a_1 + y a_2).$$

We shall work with $\Phi(B)$ and $\Phi(P)$ which are easier to view as they are two dimensional geometrical objects. If $x = (u, v)$ and $x' = (u', v')$ are two points in $\mathbb{Z}^2$, then $x + x' = (u + u', v + v')$. Under Φ^{-1}, an (integral) parallelogram in $\mathbb{Z}^2$ corresponds to a generalized arithmetic progression of rank 2, whose differences are integral linear combinations of the differences of P.

Recall that the general form of a generalized arithmetic progression of rank 2 is $\{a + x_1 a_1 + x_2 a_2 | 0 \leq x_i \leq l_i\}$. We can make the assumption that $a = 0$ for the following reason. In what follows, we consider only numbers which can be represented as a sum of the same number of elements in P. Given this, all arguments are invariant under shifting, justifying the assumption.

5.3. *The tiling argument*: *Simplified case.* It is not very hard to show that lB contains a relatively large generalized arithmetic progression of rank 2. However, this generalized arithmetic progression is still not large enough that one can apply Lemma 2.4. The tiling argument, presented below, provides a method by which we can tile several translates of a generalized arithmetic progression of rank 2 to obtain a much larger generalized arithmetic progression (for which Lemma 2.4 works).

The argument is somewhat technical and we first present a simplified version so the reader could capture the main ideas with not too much trouble. The complete treatment follows in the next subsection.

Partition each edge of $\Phi(P)$ into $\log^{50} t$ intervals of equal length (we could assume, without loss of generality, that $\log t$ is an integer and the lengths of the edges of $\Phi(P)$ are divisible by $\log^{50} t$). The products of these intervals partition $\Phi(P)$ into $\log^{100} t$ identical rectangles. A small rectangle Q is *dense* if

$$\frac{|B \cap \Phi^{-1}(Q)|}{|Q|} \geq \frac{1}{\log^7 t}.$$

Since $|B|/\mathrm{Vol}(P) = \Omega(1/\log^6 t)$, it follows, via a routine counting argument, that there is a subset B' of B, $|B'| \geq \frac{9}{10}|B|$ such that for any $x \in B'$, at least one element of $\Phi(x)$ is contained in a dense rectangle (call such an element *good*). Let C be the collection of good elements. We focus on C and the dense rectangles, ignoring all other elements.

Consider a dense rectangle Q. For each element $x \in \Phi(B) \cap Q$, $\Phi^{-1}(x)$ has high multiplicity with respect to A_1. So to each x we may associate a collection N_x of pairs of elements of A_1, where the sum of each pair equals $\Phi^{-1}(x)$.

FACT 5.2. *For each dense Q, the union of N_x's for all $x \in Q$ contains at least $\frac{n^{1/2}}{\log^{109} t}$ mutually disjoint pairs.*

Before going into the proof, let us point out why this fact is useful. The critical information here is that $\bar{l} = \frac{n^{1/2}}{\log^{109} t}$ is much larger than $l = \frac{n^{1/2}}{4t \log t}$ (here we do need the assumption that t is large). On the other hand, if one considers a sum $s = x_1 + \cdots + x_{\bar{l}/2}$, where x_i is an element of some dense rectangle

Q_i, then by an argument similar to the one following Definition 5.1, one can find $s' = x'_1 + \cdots + x'_{\bar{l}/2}$ so that $x'_i \in Q_i$ and the integer corresponding to s' ($\Phi^{-1}(s')$) can be written as the sum of $\bar{l}$ different elements of A_1. Furthermore, the difference between s and s' is relatively small since x_i and x'_i belong to the same rectangle for all i. Thus, we are able to approximate s fairly well by a sum of $\bar{l}$ different elements of A. We shall make this argument precise and quantitative at the end of this subsection (see the paragraphs following (24)).

Proof of Fact 5.2. The number of elements of $B \cap \Phi^{-1}(Q)$ is at least

$$\frac{|Q|}{\log^7 t} = \frac{|P|}{\log^{107} t} \geq \frac{|B|}{\log^{107} t}.$$

Each element in B gives rise to $l = \frac{n^{1/2}}{4t \log t}$ pairs. So the elements of $B \cap \Phi^{-1}(Q)$ give us at least

$$\frac{|B|}{\log^{107} t} \times \frac{n^{1/2}}{4t \log t} \geq \frac{c_2 n^{1/2} t}{\log^{108} t} \times \frac{n^{1/2}}{4t \log t} = \frac{c_2 n}{4 \log^{109} t}$$

pairs (notice that in the first inequality we use the lower bound $|B| \geq \frac{c_2 n^{1/2} t}{\log t}$). It is important to keep in mind that if two pairs correspond to the same number, then they are disjoint (as their sums are equal). Moreover, if two pairs correspond to two different numbers, then they have at most one element in common.

Now we create a collection of disjoint pairs by the greedy algorithm. Choose the first pair arbitrarily. Discard all pairs having nontrivial intersection with this pair. Choose the second pair arbitrarily from the set of remaining pairs and so on. Since each number in A_1 could appear in at most $|A_1| - 1 \leq c_1 n^{1/2}$ pairs, we discard at most $2c_1 n^{1/2}$ pairs in each step. Thus the collection of disjoint pairs has cardinality at least

$$\frac{c_2 n}{4 \log^{109} t} \times \frac{1}{2c_1 n^{1/2}} = \frac{c_2}{8c_1} \times \frac{n}{\log^{109} t}.$$

Recall that $c_2 = \frac{c_1^2}{3}(\sum_{k=1}^{\infty} \frac{1}{k^2})^{-1}$. Since c_1 is sufficiently large, $\frac{c_2}{8c_1} \geq 1$. It follows that our collection has at least $\frac{n}{\log^{109} t}$ disjoint pairs, completing the proof of Fact 5.2. $\square$

For each dense rectangle Q, let N_Q be the largest collection of disjoint pairs. For a pair (a, b) in N_Q, there is a corresponding point in $\mathbb{Z}^2$: $x = \Phi(a+b)$. In the following, we denote by D_Q the collection of these points; D_Q is a multi-set in $\mathbb{Z}^2$ (different pairs may lead to the same point). We have that $|D_Q| \geq \frac{n^{1/2}}{\log^{109} t}$ for any dense rectangle Q. Let D be the union of the D_Q's.

Claim 5.3. *There is a number $h = O(\log^8 t)$ such that hC contains a parallelogram R_C with cardinality at least $\alpha_1 |C|$, where α_1 is a positive constant.*

Proof of Claim 5.3. Observe that $C' = \Phi^{-1}(C)$ is a subset of P and $|C'|/\mathrm{Vol}(P)$ is $\Omega(1/\log^7 t)$. Similar to the argument preceding Case 1, consider a sequence $C_0 = C', C_{i+1} = 2C_i$. If $|C_i| \geq 7|C_{i-1}|$ for all $i \leq s$, then

$$|C_s| \geq 7^s |C_0| = \Omega(7^s \log^7 t)\mathrm{Vol}(P) \geq \frac{7^s}{\log^8 t}\mathrm{Vol}(P).$$

On the other hand, C_s is a subset of $2^s P$ which has cardinality at most $4^s \mathrm{Vol}(P)$. Thus

$$\frac{7^s}{\log^8 t} \leq 4^s,$$

which implies that $2^s \leq \log^8 t$. So there is a number s' so that $2^{s'} \leq \log^8 t$ and $|2^{s'+1}C'| \leq 7|2^{s'}C'|$. Lemma 2.5 implies that $g2^{s'}C'$ contain a generalized arithmetic progression C'' of rank 2 and cardinality $\Omega(|2^{s'}C'|) = \Omega(|C'|) = \Omega(|C|)$. Moreover, the differences of this generalized arithmetic progression are multiples of the differences of P, so $\Phi(C'')$ is a parallelogram in $\mathbb{Z}^2$. To conclude, notice that $h = g2^{s'} = O(\log^8 t)$. $\square$

It follows from the claim above that lC contains the parallelogram $P_1 = \frac{l}{h} R_C$, whose sides are L_1 and L_2. However, this parallelogram is not sufficiently large so that one can apply Lemma 2.4 to the generalized arithmetic progression $\Phi^{-1}(P_1)$. In fact, we want to obtain the larger parallelogram $P_2 = \frac{K}{h} R_C$ where $K = l \log^{30} t$. Notice that

$$\frac{K}{h}|R_C| \geq (\log^{20} t) l \alpha_1 |C| \geq (\log^{20} t) l \alpha_1 |B| \geq 20hn,$$

since $h = O(\log^8 t)$ and both c_2 and t are sufficiently large. Since $\Phi^{-1}(R_C)$ is a subset of $[hn]$, Lemma 2.4 implies that $\frac{K}{h} R_C$ contains an arithmetic progression of length $hn \geq n$, a contradiction.

Up to this point, our arguments are general. Let us now make a simplifying assumption that the basis vectors of the parallelogram R_C are the same as those of P, namely $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$. We can assume that

$$P_1 = \{u\mathbf{i} + v\mathbf{j} | 0 \leq u \leq L_1, 1 \leq v \leq L_2\}.$$

We shall construct $P_2 = \frac{K}{h} R_C$ by a tiling operation as follows. We first use KD to obtain a dense subset X of P_2. Next, we use the translates of P_1, centered at the elements of X, to cover P_2.

Consider P_2. Each of its element is an element of KC and can be written as

$$z = x_1 + \cdots + x_K,$$

where $x_i \in C$. By the definition of C, each x_i is in some dense rectangle Q. For $x_i \in Q$, we shall replace it by some $y_i \in D_Q$.

Next, let us compare $|D_Q|$ and K. We have

$$|D_Q| \geq \frac{n^{1/2}}{\log^{109} t} \text{ and } K = l \log^{30} t = \frac{n^{1/2} \log^{29} t}{4t}. \tag{24}$$

Provided that t is sufficiently large ($t \geq \log^{300} t$), $|D_Q| > 3K$ for all dense Q. Now comes the essential point of the whole argument: since $|D_Q| \geq 3K$ for all Q, we can replace $x_1, \dots, x_K$ by elements $y_1, \dots, y_K$ with the following property. There are mutually disjoint pairs $(a_1, a_1'), \dots, (a_K, a_K')$, $a_i, a_i' \in A_1$, such that $a_i + a_i' = \Phi^{-1}(y_i)$. This guarantees that $\Phi^{-1}(\sum_{i=1}^K y_K)$ can be represented as the sum of exactly $2K$ different elements from A_1.

Now let us consider the difference $\sum_{i=1}^K (y_i - x_i)$. Notice that $x_i - y_i$ is small for each i (as they are in the same dense rectangle). So the sum is small and we want to show that it is a vector of P_1. Indeed, the horizontal component of $x_i - y_i$ is at most $l_1 / \log^{50} t$, so the horizontal component of x is at most

$$K l_1 / \log^{50} t \leq \frac{l_1 n^{1/2}}{t \log^{20} t} < L_1.$$

The same estimate holds for the vertical component.

To summarize, we have proved that KD contains a subset X such that $X + lC$ contains a large rectangle P_2, where $\Phi^{-1}(P_2)$ contains an arithmetic progression of length n. Moreover, the inverse of any element from X is in S_{A_1}.

5.4. *The tiling argument*: *General case.* In the previous proof, we made the assumption that the basis vectors of R_C are the same as those of P, namely $(1, 0)$ and $(0, 1)$. This assumption might not always hold and we need to modify the proof a little bit. To start, assume that the basis vectors of R_C are $\mathbf{v}_1 = (a_1, b_1)$ and $\mathbf{v}_2 = (a_2, b_2)$, where a_i, b_i's are integers. Since R_C has high density in hP, the a_i's and b_i's cannot be too large in absolute value. Indeed,

$$\frac{|R_C|}{|hP|} \geq \frac{1}{\log^{30} t},$$

so the absolute values of a_1, a_2, b_1, b_2 are at most $\log^{30} t$. Now consider the parallelogram P_1

$$P_1 = \{\mathbf{v} + y_1 \mathbf{v}_1 + y_2 \mathbf{v}_2 | 0 \leq y_i \leq L_i\}.$$

Without loss of generality, we can assume that $\mathbf{v} = 0$. Next, consider a point $z = x_1 + \dots + x_K$ in P_2, where $x_i \in C$ (recall that $P_2 = \frac{K}{h} R_C$ is a subset of KC). As already mentioned, each x_i is in some dense rectangle Q, so we can use the dense rectangles to partition the x_i's and rewrite z as follows

$$z = \sum_{j=1}^m \sum_{x \in Q_j} x,$$

where $Q_1, \dots, Q_m$ are the dense rectangles. Since we partition P into $\log^{100} t$ rectangles, $m \le \log^{100} t$.

The important issue here is that we need to make sure that the approximation y of x is a vector in the lattice $\mathcal{L}$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. A vector in this lattice has the form $(ga_1 + g'a_2, gb_1 + g'b_2)$, where g and g' are integers. For the sake of simplicity, let us assume that $a_1a_2b_1b_2 \neq 0$. We are going to produce vectors where both coordinates are divisible by the product $a_1a_2b_1b_2$. (If $a_1 = 0$ and $a_2b_1b_2 \neq 0$ then we consider the product $a_2b_1b_2$; the rest of the proof is the same.) It is trivial that these vectors belong to the lattice $\mathcal{L}$.

We first approximate the sum $\sum_{x \in Q_j} x$ for each $1 \le j \le m$. As the subindex j plays no role, we omit it for a better presentation. To satisfy the modularity condition, we shall use only a special subset of D_Q. We say that two elements in D_Q are equivalent if both coordinates of their difference are divisible by $a_1a_2b_1b_2$. There is a equivalence class D'_Q with at least

$$|D_Q|/(a_1a_2b_1b_2)^2 \ge |D_Q|/\log^{160} t$$

elements. It is easy to see both coordinates of the sum of any $|a_1a_2b_1b_2|$ elements in D'_Q are divisible by $a_1a_2b_1b_2$. So such a sum is in $\mathcal{L}$.

Partition the set $\{x, x \in Q\}$ in the same way. We have $(a_1a_2b_1b_2)^2$ equivalence classes. In each class, partition the elements into groups of size $|a_1a_2b_1b_2|$ (one group may have fewer elements and we call this the *exceptional* group). The sum of the vectors in a nonexceptional group is a vector in $\mathcal{L}$. Replace each nonexceptional group with a group of $|a_1a_2b_1b_2|$ elements from D'_Q. Using the fact that $t \ge \log^{300} t$, we can verify that $|D'_Q|$ is still much larger than K. Thus, similar to the previous case, we can guarantee that the participating elements from D'_Q are all different. The approximating vector is the sum of the (new) elements in the nonexceptional groups and the (old) elements in the exceptional groups. It is obvious that the difference between this vector and the original vector $\sum_{x \in Q} x$ is a vector in $\mathcal{L}$ as in each replacement we replace a vector in $\mathcal{L}$ with in another vector from the same lattice.

It remains to estimate the magnitude of the difference between $x_1 + \cdots + x_K$ and its approximation. This part is essentially the same as in the simplified case, since we still replace x_i with some y_i from the same dense rectangle.

Each element of P_2 can be written as $y + z$, where y is the vector we obtain by replacements and z is vector in lC. Furthermore, y can be written as

$$y = y_1 + \cdots + y_{K'} + u_1 + \cdots + u_{K-K'},$$

where the y_i's are the replacements and $u_1, \dots, u_{K-K'}$ are elements of C which did not get replaced. In each dense rectangle, at most $a_1a_2b_1b_2 - 1$ elements did not get replaced, so

$$K - K' \le a_1a_2b_1b_2 \log^{100} t \le \log^{180} t \le l$$

and thus $\Phi^{-1}(u)$ can be represented as sum of at most $2\log^{180} t$ elements from A_1. Provided that t is sufficiently large, $|D'_Q| \geq 4K > l$, we can find $y_1, \ldots, y_{K'}$ so that their corresponding pairs are disjoint and also disjoint from the elements used in the representation of $\Phi^{-1}(u)$. Thus, $\Phi^{-1}(y)$ is an element of S_{A_1}.

Consider $\Phi^{-1}(P_2)$. This set contains an arithmetic progression N of length n. Since $\Phi^{-1}(y)$ is an element of S_{A_1}, each element of N is a sum of $\Phi^{-1}(y)$ and $\Phi^{-1}(z)$ where z and y are as above. Furthermore, both $\Phi^{-1}(y)$ and $\Phi^{-1}(z)$ are in S_{A_1}. However, we are not completely done. The (only) remaining obstacle is that an element from A_1 might appear in the representations of $\Phi^{-1}(y)$ and $\Phi^{-1}(z)$ simultaneously. We can, however, overcome this obstacle by the following simple, but useful argument.

The cloning argument. At the very beginning, we split the set A into two sets A' and A'' in such the way that $|A'| \approx |A''|$ and any number x which has high multiplicity with respect to A' should have almost the same multiplicity with respect to A''. Next, we continue with A' and keep A'' for reserve. Repeat the whole proof with A' (so A_1 will be a subset of A' etc) until the previous paragraph. To overcome the obstacle, it suffices to show that $S_{A''}$ contains an exact copy of $\Phi^{-1}(lC)$. In other words, we clone an exact copy of $\Phi^{-1}(lC)$ in $S_{A''}$.

We are going to show that a random splitting provide the sets A' and A'' as required with probability close to one. A random splitting is constructed as follows: For each element of A throw a fair coin. If head, we put the element into A', otherwise we put it into A''. If a number x has multiplicity $N_x \gg \log n$ with respect to A, then it is easy to see (via standard large deviation inequalities) that with probability at least $1 - n^{-2}$, x has multiplicities

$$\frac{N_x}{4} \pm 10\sqrt{N_x \log n} = (1 + o(1))\frac{N_x}{4}$$

with respect to both A' and A''. Since there are only $O(n)$ possible x, with probability close to 1, every x with multiplicity $\gg \log n$ has approximately the same multiplicities in A' and A''.

When we obtain the set M_k (which we rename B), the elements in M_k have multiplicity at least $\frac{n^{1/2}}{2^{k+1}(k+1)} \gg \log n$ with respect to A'. Furthermore, as we define $l = \frac{n^{1/2}}{6.2^k k}$, we have $l \leq \frac{1}{2}\frac{n^{1/2}}{2^{k+1}(k+1)}$. So the elements of M_k should have multiplicities at least l with respect to A''. Therefore $\Phi^{-1}(lC)$ is a subset of $S_{A''}$, completing the proof. □

5.5. *The sharpness of Theorem* 1.1. Here we construct a set $A \subset [n]$ with cardinality roughly $(\frac{1}{2})^{1/2} n^{1/2}$ such that S_A does not contain an arithmetic progression of length $(\frac{1}{2})^{7/4} n^{3/4}$. Assume that n is sufficiently large. Choose two different primes $p_1 \approx p_2 \approx (\frac{1}{2})^{3/4} n^{3/4}$. Consider the set

$$A = \Big\{ x_1 p_1 + x_2 p_2 | 1 \le x_i \le (1-\varepsilon) \left(\frac{1}{2}\right)^{1/4} n^{1/4} \Big\},$$

where ε is a small positive constant. One can show that

$$x_1 p_1 + x_2 p_2 = x_1' p_1 + x_2' p_2$$

if and only if $(x_1, x_2) = (x_1', x_2')$. Thus A is proper and its cardinality is $(1-\varepsilon)^2 \left(\frac{1}{2}\right)^{1/2} n^{1/2}$. On the other hand, A is a subset of $[n]$ and S_A is a subset of the generalized arithmetic progression

$$B = \Big\{ x_1 p_1 + x_2 p_2 | 1 \le x_i \le \frac{1-\varepsilon}{2} \left(\frac{1}{2}\right)^{3/4} n^{3/4} \Big\}.$$

Since

$$2 \frac{1-\varepsilon}{2} \left(\frac{1}{2}\right)^{3/4} n^{3/4} \le p_i,$$

it follows that $2B$ is still proper and this implies that if

$$(x_1 p_1 + x_2 p_2) + (x_1' p_1 + x_2' p_2) = 2(x_1'' p_1 + x_2'') p_2,$$

holds for three elements $(x_1 p_1 + x_2 p_2), (x_1' p_1 + x_2' p_2), (x_1'' p_1 + x_2'' p_2)$ of B then $x_1 + x_1' = 2x_1''$ and $x_2 + x_2' = 2x_2''$. So the length of the longest arithmetic progression in B is at most the length of an edge of B, which is less than $\left(\frac{1}{2}\right)^{7/4} n^{3/4}$.

6. Erdős-Folkman's conjectures

We prove Corollary 1.4, using Theorem 1.1. Corollary 1.5 follows from Corollary 1.4 via Folkman's partition argument. The proof presented here combines arguments from Hegyvári's paper [11] and new ideas. Let us start with a corollary of Lemma 2.3.

COROLLARY 6.1. *Let P be a generalized arithmetic progression of rank* 2, $P = \{x_1 a_1 + x_2 a_2 | 0 \le x_i \le l_i\}$, *where $l_i \ge 5a_{3-i}$ for $i = 1, 2$. Then P contains an arithmetic progression of length $l_1 + l_2$ whose difference is* $\gcd(a_1, a_2)$.

Proof of Corollary 6.1. The corollary is easy to check if either a_1 or a_2 is divisible by the other. We omit the proof of this case. If both $a_1 / \gcd(a_1, a_2)$ and $a_2 / \gcd(a_1, a_2)$ is at least 2, then by Lemma 2.3, P contains an arithmetic progression of length at least

$$\frac{3}{5 \gcd(a_1, a_2)} (l_1 a_1 + l_2 a_2) \ge \frac{6}{5} (l_1 + l_2) > (l_1 + l_2),$$

concluding the proof. □

The next lemma is a consequence of the Chinese remainder theorem and we omit the simple proof.

LEMMA 6.2. *Let* $1 \leq x_1 < x_2 < \cdots < x_h < d$ *be positive integers. If* $\gcd(x_1, \ldots, x_h) = 1 (\bmod\ d)$, *then there are integers* $0 \leq a_1, \ldots, a_h < d$ *such that* $\sum_{j=1}^h a_j x_j \equiv 1 (\bmod\ d)$.

Another useful observation is the following, due to Graham [10].

LEMMA 6.3. *Let* $Y = y_1 < y_2 < \ldots$ *be an infinite sequence of positive integers and* $S_Y = \{s_1 < s_2 < \ldots\}$. *If* $y_{m+1} \leq \sum_{i=1}^m y_i$ *for all sufficiently large* m, *then there is some* L *such that* $s_{i+1} - s_i \leq L$ *for all* i.

The proof of this lemma is short and we include it here for the sake of completeness. This proof is different from the proof in [10].

Proof of Lemma 6.3. There is some m_0 such that $y_{m+1} \leq \sum_{i=1}^m y_i$ for all $m \geq m_0$. Let $L = \sum_{i=1}^{m_0} y_i$. We are going to prove that $s_{i+1} - s_i \leq L$ for all i. Our strategy is as follows: if $s_{i+1} - s_i > L$ for some i, we construct a finite set B such that

$$s_i < \sum_{y_j \in B} y_j < s_{i+1}, \tag{25}$$

which would contradict the assumption that s_i and s_{i+1} are two consecutive elements of $S(Y)$. We denote by B_1 the set of elements of Y appearing in the representation of s_i (if s_i has many representations, choose an arbitrary one).

If there is some y_j, $j \leq m_0$, not in B_1, then $B = B_1 \cup y_j$ satisfies (25) since $y_j \leq y_{m_0} \leq L$. Let m_1 be the largest index such that $\{y_1, \ldots, y_{m_1}\} \subset B_1$, from now on we can assume that $m_1 \geq m_0$.

By the definition of m_1, y_{m_1+1} is not an element of B_1. Moreover, $m_1 \geq m_0$, so $y_{m_1+1} \leq \sum_{i=1}^{m_1} y_i$. Among all subsets C of $\{y_1, \ldots, y_{m_1}\}$ satisfying $y_{m_1+1} + \sum_{y_j \in C} y_j \leq \sum_{i=1}^{m_1} y_i$, let B_2 be the one which maximizes $\sum_{y_j \in B_2} 2^j$ (if B_2 is the empty set we set $\sum_{y_j \in B_2} 2^j = 1$). Let us consider two cases:

Case 1. There is some y_k, $k \leq m_0$, not in B_2. In this case $B = (B_1 \backslash \{y_1, \ldots, y_{m_1}\}) \cup \{y_{m_1+1} \cup B_2 \cup y_k\}$ satisfies (25) since

$$y_{m_1+1} + \sum_{y_j \in B_2} y_j \leq \sum_{i=1}^{m_1} y_i \leq y_{m_1+1} + \Big(\sum_{y_j \in B_2} y_j \Big) + y_k \leq y_{m_1+1} + \Big(\sum_{y_j \in B_2} y_j \Big) + L.$$

Case 2. $\{y_1, \ldots, y_{m_0}\} \subset B_2$. In this case, there is an index $m_2 \geq m_0$ such that $\{y_1, \ldots, y_{m_2}\} \subset B_2$ but $y_{m_2+1} \notin B_2$. Since $y_{m_1+1} + \sum_{y_j \in B_2} y_j \leq \sum_{i=1}^{m_1} y_i$, $m_2 < m_1$. Furthermore, since $m_2 \geq m_0$, $y_{m_2+1} \leq \sum_{i=1}^{m_2} y_i$, so the set $B_2' =$

$(B_2 \backslash \{y_1, \ldots, y_{m_2}\}) \cup y_{m_2+1}$ satisfies

$$y_{m_1+1} + \sum_{y_j \in B_2'} y_j \leq \sum_{i=1}^{m_1} y_i.$$

On the other hand, $\sum_{y_j \in B_2'} 2^j > \sum_{y_j \in B_2} 2^j$, which contradicts the definition of B_2. This completes the proof of Lemma 6.3. □

Now we are going to use Theorem 1.1 to prove a critical lemma.

Lemma 6.4. *For any sufficiently large constant c the following holds. For any sequence A of density at least $cn^{1/2}$ there is a positive integer d such that for every l the set S_A contains an arithmetic progression of length l with distance d.*

Proof of Lemma 6.4. We can assume that $A = \{a_1 < a_2 < \ldots\}$, where $a_m \leq m^2/c^2$ for all sufficiently large m. Let $A[m]$ be the set consisting of the first m elements of A. Fix a sufficiently large m and define $A_0 = A[m]$ and $A_i = A[2^i m] \backslash A[2^{i-1} m]$. The set A_i has $2^{i-1} m$ elements and is a subset of the interval $[4^i m^2/c^2]$.

By Theorem 1.1 (provided that c is sufficiently large), S_{A_i} contains an arithmetic progression P_i of length $l_i = 4^i m^2/c^2$ for all i. Set $Q_0 = P_0$ (and assume that d_0 is the difference of Q_0) and consider the generalized arithmetic progression $Q_0 + P_1$. This is a generalized arithmetic progression of rank 2 with volume $l_1 l_2$. Moreover, this two dimensional generalized arithmetic progression is a subset of a relatively short interval $[2l_1^{3/2}]$, so one can easily check that its differences are relatively small and satisfy the assumption of Corollary 6.1. This corollary implies that $Q_0 + P_1 = P_0 + P_1$ should contain an arithmetic progression Q_1 of length $l_0 + l_1 - 2$ with difference d_1 which is a divisor of d_0. (The -2 term comes from the fact that in Corollary 6.1, the edges of P have length $l_1 + 1$ and $l_2 + 1$, respectively.) Similarly, by considering $Q_1 + P_2$ we obtain an arithmetic progression Q_2 of length $l_0 + l_1 + l_2 - 3$ with difference d_2 which is a divisor of d_1 and so on. The difference sequence $d_0, d_1, d_2, \ldots$ is nonincreasing, so there must be an index j so that $d_i = d_j = d$ for all $i \geq j$. The arithmetic progressions $Q_j, Q_{j+1}, Q_{j+2}, \ldots$ have increasing lengths and the same difference d. Moreover, each Q_i is a subset of S_A and this completes the proof. □

We are, finally, in a position to complete the proof. The following definition will play an important role.

Definition 6.5. An infinite sequence $B = \{b_1 < b_2 < b_3 < \ldots\}$ is a (d, L)-net if $|b_{i+1} - b_i| < L$ and is divisible by d for all $i = 1, 2 \ldots$.

It is clear that if B is a (d, L)-net and Q is an arithmetic progression with difference d and length larger than L/d, then $B + Q$ contains an infinite arithmetic progression with difference d. This observation will be the leading idea in what follows.

Consider a sequence $A = \{a_1 < a_2 < a_3 \dots\}$ with density at least $cn^{1/2}$. Partition A into two parts A_1 and A_2, where A_1 (A_2) contains the elements with odd (even) indices, respectively. Since A has density $cn^{1/2}$, both A_1 and A_2 have density $cn^{1/2}/2$.

Use A_1 to create the arithmetic progressions $Q_0, Q_1, Q_2, \dots$ with the same difference d and strictly increasing lengths, as shown in Lemma 6.4.

Next, we focus on A_2. Let X be the set of divisors d' of d with the following property. All but at most finitely many elements of A_2 are divisible by d'. Since $1 \in X$, X is not empty and thus has a maximum element d_1. By throwing away finitely many elements, we can assume that all elements are divisible by d_1. Next, discard every element y (in the remaining sequence) with the property that there is only a finite number elements of A_2 equal y modulo d. Again, we discard only a finite number of elements so the remaining sequence still has the same density as A_2. Thus, we can assume that $A_2 = \{b_1 d_1 < b_2 d_1 < \dots\}$ where the b_i's have the following property: Let b'_i be the remainder when dividing b_i by d. For each i, there are infinitely many j's such that $b'_i = b'_j$. Moreover, the greatest common divisor of the b'_i's equals one modulo d by the definition of d_1.

By Lemma 6.2 and the property of A_2, we can find $(d-1)$ mutually disjoint finite subsets $X_1, \dots, X_{d-1}$ of A_2 so that the sum of the elements in each subset equals d_1 modulo d. Denote these sums by $x_1 d + d_1, \dots, x_{d-1} d + d_1$, where the x_i's are nonnegative integers. For any arithmetic progression Q_j with length $l \geq 3(x_1 + \cdots + x_{d-1})$, the set $Q_j + S_{\{x_1 d + d_1, \dots, x_{d-1} d + d_1\}}$ contains an arithmetic progression with difference d_1 and length at least $l/2$. Thus we can conclude that $S_{A_1} + S_{\{x_1 d + d_1, \dots, x_{d-1} d + d_1\}}$ contains an arbitrarily long arithmetic progression with difference d_1.

Set $A'_2 = A_2 \setminus \cup_{i=1}^{d-1} X_i$; to complete the proof, we show that $S_{A'_2}$ contains a (d_1, L)-net for some constant L. Let $S_{A'_2} = \{s_1 < s_2 < \dots\}$. Clearly all the s_i's are divisible by d_1 so it suffices to show that there is some L such that $s_{i+1} - s_i \leq L$ for all i. We do this by applying Lemma 6.3.

Given this lemma, all we need is to verify the assumption $y_{m+1} \leq \sum_{i=1}^m y_i$, where y_j denotes the j^{th} element of A'_2. Recall that A'_2 has the same density as A_2, which is $cn^{1/2}/2$. So for a sufficiently large m, $y_{m+1} \leq 4(m+1)^2/c^2 \leq m^2/3$, provided that we set c large enough. On the other hand,

$$\sum_{i=1}^m y_i \geq \sum_{i=1}^m i = \binom{m+1}{2} > m^2/3.$$

The proof is complete. □

7. Concluding remarks

- We do not need the full strength of Theorem 1.1 in the proof of Corollary 1.4. The only place where we used Theorem 1.1 is the proof of Lemma 6.4. The reader can check that in this application, it is already sufficient to have P_i containing an arithmetic progression of length $kl_i^{3/4}$, for some sufficient large constant k. Thus, what we actually required, instead of Theorem 1.1, is the following statement: For any constant k there is a constant c such that the following holds. If $A \subset [n]$ and $|A| \geq cn^{1/2}$, then S_A contains an arithmetic progression of length $kn^{3/4}$.

- For the proof of Theorem 1.1, it suffices to have a generalized arithmetic progression of constant rank in Lemma 2.6. However, we prefer to state this lemma the current form as it might be interesting in its own right. Furthermore, the proof for constant rank is not significantly simpler than the proof for the optimal rank 2.

- Sárközy [17] and Freiman [6] proved that if A is a subset of $[n]$ and $l|A| \geq cn$, where c is sufficiently large constant, then lA contains an arithmetic progression of length $\Omega(l|A|)$. Some of the facts used in our proof are corollaries of this result (for instance, Lemma 2.4). However, we avoid using this result for two reasons. The first reason is that we want our proof to be self-contained. The second, and more important, reason is that the techniques developed in our proof already provide a new and relatively simple proof of the Freiman-Sárközi result. The reader who is interested in the details of this proof is referred to [18] (§1.1 of [18]).

- By slightly modifying the proof of Theorem 1.1, we could obtain a little bit stronger result that if $|A| \geq cn^{1/2}$, then l^*A contains an arithmetic progression of length n, for some $l \leq |A|$, where l^*A denotes the set of numbers which can be represent as a sum of exactly l distinct elements of A. To see this, note that the only place in the whole proof where we do not consider sums of the same number of elements is the statement of Lemma 4.2. But, as we pointed out in Remark 4.3 following this lemma, one can modify the proof to obtain a similar statement where $S_{A'}$ is a replaced by l_0^*A', for some $l_0 = O(\log n)$.

- Together with Conjecture 1.3, Folkman [9] (see also §6 of [5]) also made the following conjecture about nondecreasing sequences

  Conjecture 7.1. *There is a constant c such that the following holds. Any nondecreasing sequence $A = \{a_1 \leq a_2 \leq a_3 \leq \dots\}$ satisfying $A(n) \geq cn$ is subcomplete.*

We confirm this conjecture in [19]. Given the proof in the previous section, it suffices to have the following variant of Theorem 1.1 for multi-sets.

THEOREM 7.2 ([19]). *There is a positive constant c such that the following holds. For any positive integer n, if A is a multi-set consisting of positive integers between* 1 *and n with and $|A| \geq cn$, then S_A contains an arithmetic progression of length n.*

Acknowledgement. We would like to thank the referee for his suggestions and D. Galvin for his careful proofreading.

COMPUTER SCIENCE DEPARTMENT, RUTGERS UNIVERSITY, PISCATAWAY, NJ
E-mail address: szemered@cs.rutgers.edu

DEPARTMENT OF MATHEMATICS, UCSD, LA JOLLA, CA
E-mail address: vanvu@ucsd.edu
Web page: http://www.math.ucsd.edu/∼vanvu/

REFERENCES

[1] Y. BILU, Structure of sets with small sumset, in *Structure Theory of Set Addition*, *Astérisque* **258** (1999), 77–108.

[2] J. BOURGAIN, On arithmetic progressions in sums of sets of integers, in *A Tribute to Paul Erdős*, 105–109, Cambridge Univ. Press, Cambridge, 1990.

[3] J. W. S. CASSELS, On the representation of integers as the sums of distinct summands taken from a fixed set, *Acta Sci. Math. Szeged* **21** (1960) 111–124.

[4] P. ERDŐS, On the representation of large interges as sums of distinct summands taken from a fixed set, *Acta. Arith.* **7** (1962), 345–354.

[5] P. ERDŐS and R. GRAHAM, *Old and New Problems and Results in Combinatorial Number Theory*. *Monographies de L'Enseignement Mathématique* [Monographs of L'Enseignement Mathématique] **28**, Université de Genève, L'Enseignement Mathématique, Geneva, 1980.

[6] G. FREIMAN, New analytical results in subset-sum problem, *Combinatorics and Algorithms* (Jerusalem, 1988), *Discrete Math.* **114** (1993), 205–217.

[7] ———, Foundations of a structural theory of set addition (Translated from the Russian), *Transl. of Mathematical Monographs* **37**, Amer. Math. Soc., Providence, R. I., 1973.

[8] G. FREIMAN, H. HALBERSTAM, and I. RUZSA, Integer sum sets containing long arithmetic progressions, *J. London Math. Soc.* **46** (1992), 193–201.

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

[10] R. GRAHAM, Complete sequences of polynomial values, *Duke Math. J.* **31** (1964), 275–285.

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

[12] V. Lev and P. Smeliansky, On addition of two distinct sets of integers, *Acta Arith.* **70** (1995),85–91.

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

[14] M. Nathanson, *Additive Number Theory. Inverse Problems and the Geometry of Sumsets*, *Graduate Texts in Math.* **165**, Springer-Verlag, New York, 1996.

[15] C. Pomerance and A. Sárközy, Combinatorial number theory, in *Handbook of Combinatorics*, Vol. 1, 2, 967–1018, Elsevier, Amsterdam, 1995.

[16] I. Ruzsa, Generalized arithmetical progressions and sumsets, *Acta Math. Hungar.* **65** (1994), 379–388.

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

[18] E. Szemerédi and V. H. Vu, Long arithmetic progressions in sum-sets and the number of x-sum-free sets, *Proc. London Math. Soc.* **90** (2005), 273–296.

[19] ———, Long arithmetic progressions in sumsets: Thresholds and bounds, *J. Amer. Math. Soc.* **19** (2006), 119–169.

(Received March 19, 2003)
(Revised February 10, 2004)