MULTIFOLD SUMS AND PRODUCTS OVER $\mathbb{R}$, AND COMBINATORIAL PROBLEMS ON SUMSETS

A Thesis
Presented to
The Academic Faculty

by

Albert Bush

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
August 2015

MULTIFOLD SUMS AND PRODUCTS OVER $\mathbb{R}$, AND COMBINATORIAL PROBLEMS ON SUMSETS

Approved by:

Professor Ernie Croot, Advisor
School of Mathematics
Georgia Institute of Technology

Professor William Trotter
School of Mathematics
Georgia Institute of Technology

Professor Michael Lacey
School of Mathematics
Georgia Institute of Technology

Professor Neil Lyall
Department of Mathematics
University of Georgia

Professor Prasad Tetali
School of Mathematics
Georgia Institute of Technology

Date Approved: 17 July 2015

To my family

ACKNOWLEDGEMENTS

First and foremost, I want to express gratitude to my advisor, Ernie Croot. His endless ideas, enthusiasm, and support were instrumental to my success. Thank you to my committee members – Michael Lacey, Neil Lyall, Prasad Tetali, and Tom Trotter – for their participation in my defense as well as their assistance throughout graduate school as teachers and mentors. I would also like to thank Chris Pryby and Gagik Amirkhanyan for our fruitful discussions and collaborations. Thank you to Robert Krone, Chun-Hung Liu, Ning Tan, Ruidong Wang, Peter Whalen, and other (former) graduate students in the School of Mathematics. My parents have always been supportive, and I would be remiss to ignore the effect this has had on my career. Lastly, thank you to my wife for her encouragement.

TABLE OF CONTENTS

SUMMARY

We prove a new bound on a version of the sum-product problem studied by Chang. By introducing several combinatorial tools, this expands upon a method of Croot and Hart which used the Tarry-Escott problem to build distinct sums from polynomials with specific vanishing properties. We also study other aspects of the sum-product problem such as a method to prove a dual to a result of Elekes and Ruzsa and a conjecture of J. Solymosi on combinatorial geometry. Lastly, we study two combinatorial problems on sumsets over the reals. The first involves finding Freiman isomorphisms of real-valued sets that also preserve the order of the original set. The second applies results from the former in proving a new Balog-Szemerédi type theorem for real-valued sets.

CHAPTER I

INTRODUCTION

For a set A in some group G, the sumset, the difference set, and the h-fold sumset are defined as

$$A + A := \{a + a' : a, a' \in A\},$$

$$A - A := \{a - a' : a, a' \in A\},$$

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

Additive combinatorialists are often interested in analyzing the size and structure of the sumset when A is finite. How large and how small can $|A + A|$ be? How does $|A + A|$ depend on $|A|$? Does $A + A$ contain an arithmetic progression, and if so, of what length?

On the other hand, often one knows something about the sumset – generally its size in relation to the original set – and one would then like to infer that the original set contains some structure based on this information. The most well-known result in this direction is Freiman's theorem [18] which states that if $|A + A| \leq K|A|$, then A is contained in a generalized arithmetic progression of size $O_K(|A|)$ and dimension $O_K(1)$.

Modern proofs of Freiman's theorem rely on a so-called 'good-modeling' or 'small-modeling' lemma. Such a lemma shows that one can, for most combinatorial purposes, map a set in one additive group or set to a more convenient one – usually $\mathbb{Z}_N$ or $[1, N]$ – while preserving additive properties of the original set. When doing so with real-valued sets, one encounters two inconveniences. First, the order of the set is not usually preserved under the mapping. Second, one would like to control the value of N when condensing the original set into $[1, N]$. Previous results have overcome the

latter difficulty. In Chapter 3, we present a mapping that overcomes both difficulties for sets with a sufficiently small sumset.

In the same chapter, we also study several applications. When analyzing the sumset, one naturally comes across the equation

$$a + b = c + d;\ a, b, c, d \in A.$$

The number of solutions to this equation is called the additive energy of a set, $E(A, A) := \{(a, b, c, d) \in A^4 : a + b = c + d\}$. When the additive energy is $o(|A|^3)$, it is easy to see that the sumset has substantial size $|A + A| = \Omega(|A|)$. Conversely, one can have sets where the additive energy is $\Theta(|A|^3)$, and the sumset still is large – consider the union of an arithmetic progression and a geometric progression of the same length. A theorem of Balog and Szemerédi [2] states that any set with large additive energy must contain some large subset whose sumset hsa size $O(|A|)$. That is, if $E(A, A) = \Theta(|A|^3)$, then there exists a subset $A' \subseteq A$ such that $|A'| = \Theta(|A|)$ and $|A' + A'| = O(|A'|)$. In Chapter 3, we study an analogue of this theorem which incorporates the indices of the original set.

If A is contained in a ring R instead of simply a group, we can define the product set, the quotient set, and the h-fold product set as

$$A.A := \{a \cdot a' : a, a' \in A\},$$
$$A/A := \{a/a' : a, a' \in A\},$$
$$A^{(h)} := \{a_1 \cdot \ldots \cdot a_h : a_i \in A\}.$$

Usually, one is working in a field (or at least an integral domain), and results about the sumset carry over in a straightforward manner to the product set. More interestingly, it turns out that there is some interplay between the behavior of the sumset and the product set. Consider two examples:

$$A = \{1, \ldots, n\} \subseteq \mathbb{Z},$$

$$B = \{2^i : i = 0, \ldots, n-1\}.$$

We see that $A + A = \{2, \ldots, 2n\}$, and so $|A + A| = 2|A| - 1$. Similarly, $B.B = \{2^i : i = 0, \ldots, 2n-2\}$ and so $|B + B| = 2|B| - 1$. However, by the prime number theorem, there are $\Theta(n/\log n)$ primes in A, and hence $|A.A| = \Omega(n^2/(\log n)^2)$. By uniqueness of binary representation, $|B + B| = \Theta(|B|^2)$. Hence, we see that in both examples, either the sumset or the product set is large. By considering $A \cup B$ we can of course have both the sumset and the product set be large, but it seems impossible to have both sets be small. That is, either the sumset or the product set has size quadratic in n. Erdős and Szemerédi conjectured this is always the case. We study this problem, its generalizations and variants, and its connections to combinatorial geometry in Chapter 2.

CHAPTER II

THE SUM-PRODUCT PROBLEM

2.1 Pairs of Sums and Products

In 1983, Erdős and Szemerédi stated the following two conjectures [14]:

Conjecture 1. *(Sum-Product Problem) For any $\epsilon > 0$, there exists an $n_0 = n_0(\epsilon)$ such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$, then*

$$|A.A| + |A + A| \geq |A|^{2-\epsilon}.$$

Conjecture 2. *(h-fold Sum-Product Problem) For any $\epsilon > 0$ and for any $h \in \mathbb{N}$, there exists an n_0 such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$, then*

$$|hA| + |A^{(h)}| \geq |A|^{h-\epsilon}.$$

Although resolution of either conjecture is currently out of reach, there has been considerable progress on Conjecture 1.

Theorem 3. *([14],[24],[11],[15],[28],[29],[22]) There exists an $0 < \epsilon < 1$ and an absolute constant $c > 0$ such that for any $A \subseteq \mathbb{R}$*

$$|A + A| + |A.A| \geq c|A|^{1+\epsilon}.$$

Initially, results were only proven when $A \subseteq \mathbb{Z}$. In that case Theorem 3 was first proved by Erdős and Szemerédi with an unspecified, but fixed value $\epsilon > 0$ [14]. Their method was refined by Nathanson and then Chen who showed one could take $\epsilon = 1/31$ [24] and $\epsilon = 1/5$ [8] respectively. In the case when one assumes $A \subseteq \mathbb{R}$, Ford used a similar method to prove one could take $\epsilon = 1/15$ [15]. Elekes showed one could take $\epsilon = 1/4$ in $\mathbb{R}$ by exhibiting a beautiful correspondence between incidence geometry and the sumset and product set [11].

Theorem 4. *There exists a $c > 0$ such that for any finite $A \subseteq \mathbb{R}$,*

$$|A + A||A.A|, |A - A||A.A|, |A + A||A/A|, |A - A||A/A| \geq c|A|^{5/2}.$$

Elekes' proof of Theorem 4 relies on the Szemerédi-Trotter theorem, and hence, the constant c depends on the so-called Szemerédi-Trotter constant. An incidence between a point and a line is a pair (p, ℓ) such that p is a point on ℓ.

Theorem 5 (The Szemerédi-Trotter Theorem)**.** *[30] For any set of points $P \subseteq \mathbb{R}^2$ and any set of lines L in the plane, the number of incidences between P and L, $I(P, L)$, satisfies*

$$I(P, L) \leq c\left(|P|^{2/3}|L|^{2/3} + |P| + |L|\right)$$

for some absolute constant $c > 0$.

Originally proved by a complicated combinatorial argument relying on cell decompositions in the plane, Szekèly simplified the proof by showing that Theorem 5 follows quickly from the crossing number inequality in graph theory. We now prove Theorem 4 using the Szemerédi-Trotter theorem.

Proof. Let $A = \{a_1, \ldots, a_n\}$. Consider the following sets of lines and points:

$$\mathcal{L}_1 := \{y = a_i(x - a_j) : a_i, a_j \in A\} \; \mathcal{P}_1 := A + A \times A.A$$

$$\mathcal{L}_2 := \{y = a_i(x + a_j) : a_i, a_j \in A\} \; \mathcal{P}_2 := A - A \times A.A$$

$$\mathcal{L}_3 := \{y = \frac{1}{a_i}(x - a_j) : a_i, a_j \in A\} \; \mathcal{P}_3 := A + A \times A/A$$

$$\mathcal{L}_4 := \{y = \frac{1}{a_i}(x + a_j) : a_i, a_j \in A\} \; \mathcal{P}_4 := A - A \times A/A$$

We finish the argument for $\mathcal{L}_1$ and $\mathcal{P}_1$, and the others follow an identical argument. Since $y = a_i(x - a_j)$ goes through the points $(a_k + a_j)$ for all $k = 1, \ldots, n$, we have that the number of incidences is at least $|A||\mathcal{L}_1|$. Since a_i determines the slope, and a_j determines the y-intercept, $|\mathcal{L}_1| = |A|^2$, and so

$$I(\mathcal{P}_1, \mathcal{L}_1)) \geq |A|^3.$$

On the other hand, by Theorem 5,

$$I(\mathcal{P}_1, \mathcal{L}_1) \leq c(|A+A|^{2/3}|A.A|^{2/3}|A|^{4/3} + |A+A||A.A| + |A|^2).$$

Hence,

$$\frac{|A|^3}{2} \leq |A|^3 - c|A|^2 \leq c(|A+A|^{2/3}|A.A|^{2/3}|A|^{4/3} + |A+A||A.A|).$$

If $|A+A||A.A|$ is the dominant term, we get a stronger result. Hence,

$$\frac{|A|^3}{4} \leq c|A+A|^{2/3}|A.A|^{2/3}|A|^{4/3}$$

which implies that

$$c'|A|^{5/2} \leq |A+A||A.A|.$$

□

This breakthrough on Conjecture 1 was significant not only because it provides an improvement on the value of ϵ in Theorem 3, but it also interpolates in an intermediate ground where the sumset or the product set is small.

Corollary 6. *There exists a $c > 0$ such that for any $A \subseteq \mathbb{R}$, if $|A+A|$ (or $|A.A|$) is at most $|A|^{1+\delta}$, then $|A.A|$ (or $|A+A|$ respectively) is at least $|A|^{3/2-\delta}$.*

Elekes and Ruzsa [13] created another geometric argument to prove a best possible result, up to the constant c, when the sumset is small.

Theorem 7. *There exists a $c > 0$ such that for any $\delta > 0$, if $A \subseteq \mathbb{R}$ and $|A+A| \leq |A|^{1+\delta}$, then $|A.A| \geq |A|^{2-c\delta}$ for A sufficiently large.*

Solymosi [28] expanded upon this geometric connection between sums and products and point-line incidences by showing one can take $\epsilon = 3/11 - \delta$ in Theorem 3 for any $\delta > 0$ given that A is sufficiently large. He improved upon this further a few years later and showed that one can take $\epsilon = 1/3 - \delta$ [29]. The latter argument surprisingly avoided the use of the Szemerédi-Trotter theorem despite also relying on combinatorial geometry.

Theorem 8. *For any set $A \subseteq \mathbb{R}$,*

$$|A.A||A+A|^2 \geq \frac{|A|^4}{8\lceil \log |A| \rceil}.$$

Up to a constant factor and the power of the logarithm, one cannot prove a stronger lower bound on $|A.A||A+A|^2$ as seen by taking A to be an arithmetic progression.

Theorem 9. *[16] For $A = \{1, \ldots, n\}$, $|A+A| = 2n-1$ and*

$$|A.A| = \Theta\left(\frac{n^2}{(\log n)^c (\log\log n)^{3/2}}\right)$$

where $c = \frac{-\log\log 2}{\log 2} - \frac{1}{\log 2} + 1$.

Theorem 8 also provides a significant improvement to Corollary 6 (and to the constant c in Theorem 7) when one assumes that the sumset is small.

Corollary 10. *For any $A \subseteq \mathbb{R}$, if $|A+A| \leq |A|^{1+\delta}$, then*

$$|A.A| \geq \frac{|A|^{2-2\delta}}{8\lceil \log |A| \rceil}.$$

Notice however that if one assumes the product set is small and apply Theorem 8, we do not get any real improvement over Corollary 6. This direction is still open:

Conjecture 11. *There exists a $c > 0$ such that if $A \subseteq \mathbb{R}$ is sufficiently large and if $|A.A| \leq |A|^{1+\delta}$, then $|A+A| \geq |A|^{2-c\delta}$.*

There have been a sequence of results towards this conjecture. First, Chang proved it for integral sets [7], but her technique relied heavily on prime factorization properties of the integers. When one assumes that $|A.A| = O(|A|)$, Chang [7] showed that one can use Freiman's theorem to deduce that $|A+A| = \Omega(|A|^2)$. We remind the reader that Corollary 6 showed that $|A.A| \leq |A|^{1+\delta}$ implies that $|A+A| \geq |A|^{3/2-\delta}$. Shkredov and Konyagin improved this recently [22] to the following.

Theorem 12. *If $A \subseteq \mathbb{R}$ and $|A.A| \leq |A|^{1+\delta}$, then*

$$|A+A| \geq |A|^{\frac{3}{2}+\frac{1}{12}-\frac{5}{6}\delta}$$

and

$$|A+A| \geq |A|^{\frac{3}{2}+\frac{1}{32}-\frac{19}{32}\delta}.$$

In the same paper, they were even able to improve Theorem 8 slightly past 4/3 in the exponent.

Theorem 13. *If $A \subseteq \mathbb{R}$, then*

$$|A+A| + |A.A| \geq c|A|^{\frac{4}{3}+\frac{1}{20599}}$$

We now prove Conjecture 11 under the assumption of a certain other conjecture. This is significant because Iosevich, Roche-Newton, and Rudnev recently presented a beautiful but flawed argument of the following conjecture [20].

Conjecture 14. *Let $A \subseteq \mathbb{R}$. Then,*

$$|\{(x_1, \ldots, x_8) \in A^8 : x_1x_2 - x_3x_4 = x_5x_6 - x_7x_8\}| = O(|A|^6 \log |A|).$$

Proposition 15. *If Conjecture 14 is true, then there exists an absolute constant c such that for every $0 < \epsilon < \frac{1}{2}$, there exists an $n \in \mathbb{N}$ such that if $A \subseteq \mathbb{R}$ is sufficiently large with $|A.A| \leq |A|^{1+\epsilon}$, then $|A-A| \geq c\frac{|A|^{2-3\epsilon}}{(\log |A|)^8}$. Equivalently,*

$$|A-A||A.A|^3 \geq c\frac{|A|^5}{(\log |A|)^8}.$$

Proposition 15 is almost a dual to Theorem 8. It would be interesting to see if one could in fact prove such a dual.

Conjecture 16. *If Conjecture 14 is true, then there exists $c, d > 0$ such that for every $0 < \epsilon < \frac{1}{2}$, if $A \subseteq \mathbb{R}$ is sufficiently large, then*

$$|A-A||A.A|^2 \geq c\frac{|A|^4}{(\log |A|)^d}.$$

Proof of Proposition 15. Let $A \subseteq \mathbb{R}$ be a finite set such that $|A.A| \leq |A|^{1+\epsilon}$. Letting $r(x) := |\{(a,b) \in A^2 : ab = x\}|$, we have that

$$\sum_{x \in A.A} r(x) = |A|^2$$

and

$$\sum_{x \in A.A} r(x)^2 = E_\times(A) = |\{(a,b,c,d) : ab = cd\}|.$$

By a dyadic pigeonhole argument, there exists a $t \in \{0, \dots 2\log|A|\}$ such that

$$\sum_{x \in A.A : r(x) \in [2^t, 2^{t+1}]} r(x) \geq \frac{|A|^2}{2\log|A|}. \tag{1}$$

Let $C := \{(x,y) \in A \times A : r(xy) \in [2^t, 2^{t+1})\}$, and let $D := \{xy : (x,y) \in C\} \subseteq A.A$. Note that (1) is counting the size of C, so $|C| > \frac{|A|^2}{2\log|A|}$. Observe that

$$\frac{|A|^2}{2\log|A|} \leq |C| = \sum_{x \in D} r(x) \leq 2^{t+1}|D|. \tag{2}$$

Since $D \subseteq A.A$, and $|A.A| \leq |A|^{1+\epsilon}$, we get that

$$2^t \geq \frac{|A|^{1-\epsilon}}{4\log|A|} \tag{3}$$

Now, let $I_w := \{(a,b) \in C : ab \in wA\}$. We have that

$$\sum_{w \in A} |I_w| = \sum_{w \in A} \sum_{d \in D} |\{(x,y) \in C : xy = d \in wA\}|.$$

Switching the order of summation, we get that

$$= \sum_{d \in D} \sum_{w \in A} |\{(x,y) \in C : xy = d \in wA\}|.$$

Observe that this is equal to

$$= \sum_{d \in D} |\{(x,y) \in C : xy = d\}|^2 \geq 2^{2t+2}|D|.$$

So,

$$\mathbb{E}_w(|I_w|) \geq \frac{4 \cdot 2^{2t}|D|}{|A|}. \tag{4}$$

Let

$$J_w := \{(x_1, x_2, \ldots, x_8) \in C^4 : x_1x_2 - x_3x_4 = x_5x_6 - x_7x_8; x_1x_2 \in wA\}.$$

We will use Conjecture 14 applied to A to give an upper bound on $\mathbb{E}(|J_w|)$. First, note that

$$\sum_{w\in A} |J_w| = \sum_{w\in A} \sum_{(x_1,x_2)\in C} |\{(x_1, \ldots, x_8) \in C^4 : x_1x_2 \in wA; x_1x_2 - x_3x_4 = x_5x_6 - x_7x_8\}|.$$

Again, changing the order of summation,

$$= \sum_{(x_1,x_2)\in C} \sum_{w\in A} |\{(x_1, \ldots, x_8) \in C^4 : x_1x_2 \in wA; x_1x_2 - x_3x_4 = x_5x_6 - x_7x_8\}|$$

$$= \sum_{(x_1,x_2)\in C} |\{(x_1, \ldots, x_8) \in C^4 : x_1x_2 - x_3x_4 = x_5x_6 - x_7x_8\}| \cdot r(x_1x_2)$$

$$\leq |\{(x_1, \ldots, x_8) \in C^4 : x_1x_2 - x_3x_4 = x_5x_6 - x_7x_8\}|2^{t+1} \ll |A|^6 \log|A|2^t$$

by Conjecture 14. Thus,

$$\mathbb{E}(|J_w|) \ll |A|^5 2^t \log|A|. \tag{5}$$

We will show that for some $w \in A$,

$$|I_w|^4 \gg \frac{|A|^{2-3\epsilon}}{(\log|A|)^8}|J_w|. \tag{6}$$

The theorem will quickly follow once this is established.

For a contradiction, suppose that for all $w \in A$

$$|I_w|^4 \ll \frac{|A|^{2-3\epsilon}}{(\log|A|)^8}|J_w|.$$

If this is true, then

$$\sum_w |I_w|^4 \ll \frac{|A|^{2-3\epsilon}}{(\log|A|)^8} \sum_w |J_w|.$$

Using Holder's inequality on the left sum and (4), we get that

$$\sum_w |I_w|^4 \geq \frac{(\sum_w |I_w|)^4}{|A|^3} \gg \frac{2^{8t}|D|^4}{|A|^3}.$$

On the other hand, using (5) we have that

$$\sum_w |I_w|^4 < \frac{c|A|^{2-3\epsilon}}{(\log|A|)^8} \sum_w |J_w| \ll \frac{|A|^{8-3\epsilon} 2^t \log|A|}{(\log|A|)^8}$$

Thus, we have that

$$\frac{2^{8t}|D|^4}{|A|^3} \ll \frac{|A|^{8-3\epsilon}2^t}{(\log|A|)^7}.$$

However, using (2) and (3) on the left side of this inequality gives us a contradiction for an appropriately chosen constant c. Thus, we may assume there exists a w such that (6) holds. For such a w, let

$$G := I_w \times I_w = \{(x_1, x_2, x_3, x_4) \in C^2 : x_1x_2 \in wA, x_3x_4 \in wA\}$$

and

$$n(s) := |\{(x_1, x_2, x_3, x_4) \in G : x_1x_2 - x_3x_4 = s\}|$$

and let

$$S := \{s : \text{ There exists } (x_1, x_2, x_3, x_4) \in G : x_1x_2 - x_3x_4 = s \ \}.$$

Then, we have that by Cauchy-Schwarz

$$\sum_{s\in S} n(s)^2 \geq \frac{|G|^2}{|S|}.$$

On the other hand,

$$\sum_{s\in S} n(s)^2 \leq |J_w| \ll |I_w|^4 \cdot \frac{(\log|A|)^8}{|A|^{2-3\epsilon}}.$$

Note that $|G| = |I_w|^2$, so combining the above two inequalities, we get that

$$|S| \gg \frac{|A|^{2-3\epsilon}}{(\log|A|)^8}.$$

Lastly, observe that $|S| \leq |A-A|$ since for each $s \in S$, $s = wa_1 - wa_2 = w(a_1 - a_2) \in w(A-A)$. □

2.2 Combinatorial Geometry

As shown in the proof of Theorem 4, results in combinatorial geometry can have a significant impact on sum-product theorems. In fact, Elekes' treatise [12] contained many connections between combinatorial problems on points and lines and additive combinatorics. Amirkhanyan et al [1] demonstrated – not for the first time – that the relationship is reciprocated by using additive combinatorial tools to prove the following geometric theorem.

Theorem 17. *[1] For every $\epsilon > 0$, there exists $0 < \delta < \epsilon$ such that for sufficiently large $n = n(\epsilon, \delta)$ the following holds: if $A \subseteq \mathbb{R}$ has size n, then every set of at least n^{ϵ} lines in $\mathbb{R}^2$, each of which intersects $A \times A$ in at least $n^{1-\delta}$ points, contains either two parallel lines or three lines with a common intersection point.*

A variant of this theorem was conjectured in Elekes' treatise. The conjecture was attributed to Solymosi. We say a set of lines is in *general position* if no two are parallel and no three intersect at a point.

Conjecture 18. *For any $c > 0$, if there exists a set of lines, all of whom contain at least cN points in an $N \times N$ Cartesian product, then at most $d = d(c)$ can be in general position.*

Although clearly Theorem 17 and Conjecture 18 are related, Conjecture 18 is still technically open. It seems possible to perhaps use the techniques in the proof of Theorem 17 to prove Conjecture 18 directly. One of the main lemmas used in the proof of Theorem 17 is the following reduction which we prove here. We say a line is k-rich in a set of points P if it contains at least k points in P.

Theorem 19. *For every $\varepsilon > 0$, there exists $0 < \delta < \varepsilon$ such that for sufficiently large $n = n(\varepsilon, \delta)$, the following holds:*

If $A \subseteq \mathbb{R}$, $|A| = n$, then every set of at least $n^{1-\varepsilon}$ lines in $\mathbb{R}^2$, each of which intersects $A \times A$ in at least $n^{1-\delta}$ points, contains either two parallel lines or $C = C(\varepsilon) > 0$ lines with a common intersection point.

First, we will prove Theorem 19; then, we will prove that Theorem 19 implies Theorem 17. A key idea in our proofs is that we can combine rich lines in a grid in a natural way in order to to get more. However, the number of rich lines in a grid is bounded, and after several iterations, we obtain structural properties on the sumset and product set of the grid. Function composition is the natural group action on lines; however, our action will be slightly modified so that $(f, g) \to f^{-1} \circ g$. We define the operation $*$ by $f * g = f^{-1} \circ g$. Given two sets L, L' of $n^{1-\delta}$-rich lines, we would like to consider the set of lines $f * g$ which retain a large amount of richness in $A \times A$.

$$L * L := \left\{ f * g : f, g \in L \text{ and } |f * g \cap (A \times A)| \geq n^{1-2\delta}/2 \right\}. \tag{7}$$

A priori, one may have that $L * L$ does not contain many lines compared to L – why should $f * g$ contain many points in $A \times A$? However, the following lemma shows that this is not the case since the $*$ operation preserves richness.

Proposition 20. *Let $A \subseteq \mathbb{R}$, and let L be a set of lines $y = ax + b$ such that each line in L is $n^{1-\delta}$-rich in $A \times A$. Then for at least $\frac{|L|^2}{2n^{2\delta}}$ pairs of lines $f, g \in L$, $f * g$ is $\frac{n^{1-2\delta}}{2}$ rich in $A \times A$.*

We need the following combinatorial lemma. Let $A \subseteq \mathbb{R}$ be a set of size n.

Lemma 21. *Given sets $A_1, \ldots, A_k \subseteq A$ such that $|A_i| \geq n^{1-\delta}$ for all $i = 1, \ldots, k$, we must have at least $\frac{k^2}{2n^{2\delta}}$ pairs (A_i, A_j) with $|A_i \cap A_j| \geq \frac{n^{1-2\delta}}{2}$.*

Proof. Let $B = \{(i, j) : |A_i \cap A_j| \geq \frac{n^{1-2\delta}}{2}\}$. For a contradiction, suppose that

$$|B| < \frac{k^2}{2n^{2\delta}}.$$

Then,

$$\begin{aligned}\sum_{i,j} |A_i \cap A_j| &= \sum_{(i,j)\in B} |A_i \cap A_j| + \sum_{(i,j)\in B^c} |A_i \cap A_j| \\ &< n \cdot \frac{k^2}{2n^{2\delta}} + k^2 \cdot \frac{n^{1-2\delta}}{2} \\ &< k^2 n^{1-2\delta}.\end{aligned}$$

However, letting $d(x) := |\{i \in [k] : x \in A_i\}|$, we have by Cauchy-Schwarz

$$\sum_{i,j} |A_i \cap A_j| = \sum_{x\in A} d(x)^2 \geq \left(n^{-1/2} \sum_{x\in A} d(x)\right)^2 = n^{-1} \left(\sum_{i\in[k]} |A_i|\right)^2 \geq k^2 n^{1-2\delta}$$

which is a contadiction. □

Now, we prove Proposition 20.

Proof. For each $f \in L$, let $X_f := \{x \in A : f(x) \in A\}$, and similarly $Y_f := \{y \in A : f^{-1}(y) \in A\}$. Observe that $X_f = Y_{f^{-1}}$. Thus, for $f, g \in L$, if $|Y_f \cap Y_g| \geq \frac{n^{1-2\delta}}{2}$, then $f^{-1} \circ g = f * g$ is $\frac{n^{1-2\delta}}{2}$-rich in $A \times A$. Apply Lemma 21 to all the sets X_f, and the result follows. □

We need the following deep and technical theorem from [1] as well as its corollary. A set of lines forms a *star family* if there is a single point p that is contained in every line in the family. A set of lines is in *near-general position* with star families bounded by C if no two lines are parallel and any star family has size at most C. Roughly speaking, the following theorem says that not only does the $*$ operation preserve richness, but it also preserves the property of general position to an extent.

Theorem 22. *[1] For all $0 < \varepsilon < 1$, there exists $0 < \alpha_0 < \varepsilon$ such that for all $0 < \alpha < \alpha_0$, there exists $0 < \delta_0 < \alpha$ such that for all $0 < \delta < \delta_0$ and for finite sets A with $|A| = n$ sufficiently large, the following holds:*

If L is a set of at least n^ε lines in near-general position (with star families bounded by some constant $C = C(\varepsilon, \alpha) > 0$) which are $n^{1-\delta}$-rich in $A \times A$, then:

(i) *If $L * L$ contains a family P of parallel lines, then $|P| \leq 2|L * L| n^{2\delta}/|L|$.*

(ii) *If $L * L$ contains a star family S, then $|S| \leq 2C|L * L| n^{2\delta}/|L|$.*

(iii) *If P_λ denotes the set of lines in $L * L$ with common slope λ, then $|P_\lambda| \geq n^\alpha$ for at most n^α numbers λ.*

(iv) *If S_p denotes the set of lines in $L * L$ with common meeting point p, then $|S_p| \geq n^\alpha$ for at most n^α points p.*

Corollary 23. *For all $0 < \varepsilon < 1$ there exists $0 < \alpha_0 < \varepsilon$ such that, for all $0 < \alpha < \alpha_0$, there exists $0 < \delta_0 < \alpha$ such that for all $0 < \delta < \delta_0$ and for sufficiently large n, the following holds:*

Let $A \subseteq \mathbb{R}$ be a finite set with $|A| = n$, and let L be a set of at least n^ε lines which are all $n^{1-\delta}$-rich in $A \times A$. If L contains no parallel lines and all star families in L are bounded above in size by $C = C(\varepsilon, \alpha)$, then there exists a subset $R \subseteq L * L$ such that

- *$|R| \geq |L| n^{-c\alpha}$ for some absolute constant c,*
- *R contains no two lines which are parallel, and*
- *at most $k = \lceil \varepsilon/\alpha \rceil$ lines of R pass through any given point of $\mathbb{R}^2$.*

2.2.1 Proving Theorem 19, the Weakened Theorem

Using Corollary 23, we are now ready to prove Theorem 19. A major tool used will be the commutator graph, which we draw from [12].

Let $A \subseteq \mathbb{R}$, let $\delta > 0$, and let L be a set of $n^{1-\delta}$-rich lines in $\mathbb{R}^2$. The *commutator graph* on L is the graph $G = (V, E)$, where

$$V(G) = L * L \cup L^{-1} * L^{-1}$$

(with the minor change that we require minimum richness only $n^{1-5\delta}$ for each line in $L * L$ and $L^{-1} * L^{-1}$) and

$$E(G) = \left\{ \{f * g, g^{-1} * f^{-1}\} : f, g \in L, f * g \in L * L, g^{-1} * f^{-1} \in L^{-1} * L^{-1} \right\}.$$

We draw attention to the fact that the lines $f * g$ and $g^{-1} * f^{-1}$ have the same slope. Hence, any edge of the commutator graph is between two parallel lines.

Proof of Theorem 19. Let $\varepsilon > 0$, let $0 < \delta \ll \varepsilon$, and let $A \subset \mathbb{R}$ with $n = |A| > 0$. Suppose for a contradiction that L is a set of at least $n^{1-\varepsilon}$ lines, all $n^{1-\delta}$-rich in $A \times A$, and that L is in near-general position with star families bounded in size by a constant $C > 0$ independent of n. Consider the commutator graph on L.

If $|V(G)| \geq n^{1+4\delta}$, then we contradict Theorem 5, so let us assume that $|V(G)| < n^{1+4\delta}$. We claim that $|E(G)| \geq n^{2-6\delta}$. If this is true, then there is a vertex with degree at least $|E(G)| / |V(G)|$, so there is a connected component (corresponding to a set of parallel lines) of size $n^{1-10\delta}$, in contradiction with Theorem 22(i).

Let $S(f) = X(f) \times Y(f)$ for each $f \in L$. By applying Proposition 20 to the collection of sets $S(f)$, where each set $S(f)$ has size at least $n^{2-2\delta}$, we must have at least $n^{2-4\delta}/2 \geq n^{2-5\delta}$ pairs $S(f), S(g)$ with $|S(f) \cap S(g)| \geq n^{2-4\delta}/2 \geq n^{2-5\delta}$. Note that for any sets A_1, A_2, A_3, A_4, $(A_1 \times A_3) \cap (A_2 \times A_4) = (A_1 \cap A_2) \times (A_3 \cap A_4)$. Thus, since $|S(f) \cap S(g)| \geq n^{2-5\delta}$, we have $|X(f) \cap X(g)| \geq n^{1-5\delta}$ and $|Y(f) \cap Y(g)| \geq n^{1-5\delta}$. Thus, we have at least $n^{2-5\delta}$ pairs $f, g \in L$ such that $f * g$ and $g^{-1} * f^{-1}$ are each $n^{1-5\delta}$-rich.

Let f_i, g_i denote the lines such that $P_i := \{f_i * g_i, g_i^{-1} * f_i^{-1}\}$ is a pair of $n^{1-5\delta}$-rich lines. Given an index i, f_i and g_i intersect at a unique point (x, y); it then follows that y is the unique fixed point of $f_i * g_i$ and x is the unique fixed point of $g_i^{-1} * f_i^{-1}$. Suppose there were $2C + 2$ indices $i_1, ..., i_{2C+2}$ such that $P_{i_j} = P_{i_k}$ for all $1 \leq j, k \leq 2C + 2$. Then there would exist $C + 1$ indices $i_{j_1}, ..., i_{j_{C+1}}$ such that

$$f_{i_{j_1}} * g_{i_{j_1}} = \cdots = f_{i_{C+1}} * g_{i_{C+1}} \quad \text{and} \quad g_{i_{j_1}}^{-1} * f_{i_{j_1}}^{-1} = \cdots = g_{i_{C+1}}^{-1} * f_{i_{C+1}}^{-1}.$$

Since for each $1 \le k \le C+1$ there is a unique (x,y) such that $f_{i_{j_k}} * g_{i_{j_k}}(y) = y$ and $g_{i_{j_k}}^{-1} * f_{i_{j_k}}^{-1}(x) = x$, it follows that $f_{i_{j_k}}$ and $g_{i_{j_k}}$ all intersect the point (x,y). Since the $f_{i_{j_k}} * g_{i_{j_k}}$ must all have the same slope and L has no parallel lines, we cannot have that $f_{i_{j_k}} = f_{i_{j'_k}}$ for $k \neq k'$ or else $g_{i_{j_k}} = g_{i_{j'_k}}$ as well, contradicting distinctness of the pairs. Similarly we must have $g_{i_{j_k}} \neq g_{i_{j'_k}}$ for $k \neq k'$. The collection

$$\{f_{i_{j_k}} : 1 \le k \le C+1\} \cup \{g_{i_{j_k}} : 1 \le k \le C+1\}$$

must therefore contain at least $C+1$ distinct lines (a single line may appear as an f_{i_j} at most once and as a g_{i_j} at most once). But then we have a set of more than C concurrent lines at (x,y), contradicting the hypothesis that L is in almost-general position.

Thus, for each edge e, there are at most $2C+2$ pairs $\{f_i \circ g_i^{-1}, g_i^{-1} \circ f_i\}$ equal to e, so the total number of edges in G is at least $n^{2-5\delta}/(2C+2) \gg n^{2-6\delta}$, yielding a contradiction with Theorem 22(i). □

We remark that taking $\delta < \varepsilon/12$ is sufficient for the proof to go through.

2.2.2 The Weakened Theorem Implies the Strong Version

For $\ell \in L * L$, let $\mathcal{P}(\ell)$ be the set of all pairs $(f,g) \in L \times L$ such that $f * g = \ell$.

Lemma 24. *For all $0 < \varepsilon < 1$, there exists $0 < \alpha_0 < \varepsilon$ such that, for all $0 < \alpha < \alpha_0$, there exists $0 < \delta_0 < \alpha$ such that for all $0 < \delta < \delta_0$ and for sufficiently large n, the following holds:*

Let $A \subseteq \mathbb{R}$ have size n, and let L be a set of at least n^{ε} near-general position lines, all of which are $n^{1-\delta}$-rich in $A \times A$. Then there exists a set $L' \subseteq L * L$ such that L' is a set of lines in near-general position, $|L'| > |L|\, n^{-5\alpha-4\delta}$, and for all $\ell \in L'$,

$$|\mathcal{P}(\ell)| \ge \frac{|L|^2}{2\,|L * L|\, n^{3\delta}}.$$

Proof. Let

$$S := \{(f,g) \in L \times L : f * g \text{ is } n^{1-5\delta}\text{-rich}\}.$$

By Proposition 20, $|S| \geq |L|^2 n^{-3\delta}$. Let

$$T := \left\{ (f,g) \in S : |\mathcal{P}(f*g)| \leq \frac{|L|^2}{2\,|L*L|\,n^{3\delta}} \right\}.$$

If $|T| > |S|/2$, then we obtain an absurdity:

$$|L*L| = \sum_{(f,g)\in S} \frac{1}{|\mathcal{P}(f*g)|} = \sum_{(f,g)\in S\setminus T} \frac{1}{|\mathcal{P}(f*g)|} + \sum_{(f,g)\in T} \frac{1}{|\mathcal{P}(f*g)|} \geq$$
$$\frac{1}{|L|}\,|S\setminus T| + \frac{2\,|L*L|\,n^{3\delta}}{|L|^2}\,|T| > |L*L|\,.$$

Thus, $|S\setminus T| \geq |L|^2 n^{-3\delta}/2 > |L|^2 n^{-4\delta}$. Letting $L' = \{f*g : (f,g) \in S\setminus T\}$, we then have $|L'| \geq |L|\,n^{-4\delta}$. Apply Corollary 23 to deduce that L' contains a subset of $|L|\,n^{-5\alpha-4\delta}$ lines in near-general position. □

Proposition 25. *Theorem 19 implies Theorem 17.*

Proof. Let L be a set of n^ε lines in general position, all of which are $n^{1-\delta}$-rich for some $\delta > 0$ to be chosen later. Fix $\alpha < \varepsilon$, and suppose $\left|L^{*(k+1)}\right| \geq \left|L^{*k}\right| n^{5\alpha}$ for all k up to $m = \lfloor 2/\alpha \rfloor$. (By Corollary 23, we may further assume that L^{*j} is in near-general position for all $j \leq k$ at the cost of a factor of $n^{4\alpha}$ each iteration.) Redefining δ if necessary, we can take $1 - 4\cdot 5^m\delta > 0$. For sufficiently large n, we then have

$$\left|L^{*(m+1)}\right| \geq n^{\varepsilon+m\alpha} \geq n^2.$$

But this violates Theorem 5, so such an m cannot exist. Therefore there exists $k < 2/\alpha$ such that

$$\left|L^{*(k+1)}\right| < \left|L^{*k}\right| n^{5\alpha}.$$

In this case, let $L' = L^{*k}$ for the smallest such k (such that the above inequality would now read $|L'*L'| < |L'|\,n^{5\alpha}$), let $\alpha' < 5\alpha$ such that $\alpha' \ll \varepsilon$, let $N = |L'|$, and choose $\delta' \leq 5^k\delta$ such that $\delta' \ll \alpha'$.

By applying Lemma 24, we can restrict our attention to a subset $L'' \subseteq L' * L'$ of size at least $Nn^{-5\alpha'-4\delta'}$ such that all lines in L'' are in near-general position and, for all $\ell \in L''$,

$$|\mathcal{P}(\ell)| \geq \frac{N^2}{2\,|L' * L'|\, n^{3\delta'}} \geq \frac{N}{2n^{5\alpha+3\delta'}} > \frac{N}{2n^{\alpha'+3\delta}}.$$

If ℓ is a line in L'', then $\ell = f * g$ for some $f, g \in L'$. We will then have at least

$$\frac{1}{C}\,|L''|\,(Nn^{-\alpha'-4\delta'})^2 \geq \frac{1}{C} N^3 n^{-5\alpha'-8\delta'} \gg N^3 n^{-6\alpha'}$$

solutions $(f, g, f', g') \in L \times L \times L \times L$ to the equation

$$f' * f(0) = g' * g(0) \tag{8}$$

(The factor of $1/C$ comes from the fact that L'' is a set of lines in near-general position, so at most C lines will share a y-intercept.)

Now, fixing f', g' in (8) and letting f, g vary, we can interpret (8) as the line $f' * g'$, where the x and y variables are the y-intercepts of f and g. Letting B be the set of y-intercepts among lines in L'', we may interpret the above count of solutions to (8) as stating that many of the lines $f' * g'$ are $Nn^{-7\alpha'}$-rich in the new grid $B \times B$. Indeed, let

$$S := \{(f', g') \in L' \times L' : f' * g' \text{ is } Nn^{-7\alpha'}\text{-rich in } B \times B\},$$

and let $p(f', g')$ denote the number of points that $f' * g'$ intersects in $B \times B$. Then, for a contradiction, assume $|S| < N^2 n^{-8\alpha'}$. This implies the absurdity:

$$N^3 n^{-6\alpha'} = \sum_{(f',g') \in S} p(f', g') + \sum_{(f',g') \in S^c} p(f', g') < |S|\,|B| + |S^c|\,(Nn^{-7\alpha'}) <$$

$$|S|\,Nn^{\alpha'} + (N^2 - |S|)Nn^{-7\alpha'} < |S|\,Nn^{\alpha'} + N^3 n^{-7\alpha'} < 2N^3 n^{-7\alpha'} \ll N^3 n^{-6\alpha'}.$$

(Note that this requires $N \gg n^{4\alpha'}$, which is satisfied provided $\alpha' \ll \varepsilon$ because $N \gg n^{\varepsilon}$.) Thus, $|S| \geq N^2 n^{-8\alpha'}$, and that implies that we have at least $Nn^{-8\alpha'}$ lines that are all $Nn^{-7\alpha'}$-rich in $B \times B$. Moreover, since L' is in near-general position, we may

extract a set of lines from $S * S$ that are in near-general position, and this set has size at least $Nn^{-11\alpha'}$. However, this is in contradiction with Theorem 19, since $S*S$ is a set of $N^{1-\gamma}$-rich lines in near-general position for some $\gamma > 0$, and $|S * S| \geq N^{1-12\gamma}$. $\square$

2.3 Multifold Sums and Products

Progress on the h-fold version of Erdős and Szemerédi's conjecture has been much slower. Essentially the only real progress towards Conjecture 2 is the following theorem of Bourgain and Chang that only holds for sets of integers.

Theorem 26. *[4] For every $b > 0$, there exists and $h \in \mathbb{N}$ such that for any $A \subseteq \mathbb{Z}$*

$$|hA| + |A^{(h)}| \geq |A|^b$$

One can take b to be on the order of $(\log h)^{1/4}$ – far from the conjectured bound of $\Theta(h)$. This result relies heavily on analytic estimates that only hold over $\mathbb{Z}$, and it is still open whether Theorem 26 holds over $\mathbb{R}$. Generalizations of the geometric approach used in the 2-fold sum-product conjecture have typically resulted in bounds where b is bounded, not an unbounded function of h. Theorem 26 was a generalization of techniques introduced by Chang who first proved the following weaker theorem.

Theorem 27. *[6] For any $h \in \mathbb{N}$, there exists a $K = K(h) > 0$ such that if $A \subseteq \mathbb{Z}$ and $|A.A| \leq K|A|$, then*

$$|hA| \geq c(K, h)|A|^h.$$

A significant shortcoming in Chang's approach is the dependence of $c(K, h)$ on K. The conclusion of Theorem 27 becomes trivial when K is taken any larger than $\log |A|$ due to Chang's use of Freiman's Structure Theorem. When $|A.A|$ is very close to $|A|$, Freiman's Structure Theorem provides a precise characterization of A – that it is a large subset of a geometric progression – which allows one to analyze the nature of the h-fold sumset.

A second shortcoming in Theorem 27 is that it only holds for integral sets. Several years afterward, Chang was able to remedy both shortcomings individually. By an ingenious use of the Subspace Theorem, Chang proved that Theorem 27 also holds for real-valued sets.

Theorem 28. *[7] For any $h \in \mathbb{N}$, there exists a $K = K(h) > 0$ such that if $A \subseteq \mathbb{R}$ and $|A.A| \leq K|A|$, then*

$$|hA| \geq c(K,h)|A|^h.$$

Additionally, Chang proved a version of Theorem 27 that allowed one to take K to be up to a small power of $|A|$.

Theorem 29. *[7] For any $h \in \mathbb{N}$, there exists an $\epsilon > 0$ such that if $A \subseteq \mathbb{Z}$ and $|A.A| \leq |A|^{1+\epsilon}$, then*

$$|hA| \geq |A|^{h-\delta_h(\epsilon)}$$

where $\delta_h \to 0$ as $\epsilon \to 0$.

Proving a real-valued version of the Bourgain-Chang theorem has been a major research goal in this area. Hence, one might try to prove a real-valued version of Theorem 29 as a first step.

Conjecture 30. *For any $h \in \mathbb{N}$, there exists an $\epsilon > 0$ such that if $A \subseteq \mathbb{R}$ and $|A.A| \leq |A|^{1+\epsilon}$, then*

$$|hA| \geq |A|^{h-\delta_h(\epsilon)}$$

where $\delta_h \to 0$ as $\epsilon \to 0$.

Croot and Hart used the Szemerédi Cube Lemma to prove a weaker form of Conjecture 30.

Theorem 31. *[9] For every $h \in \mathbb{N}$ there exists an $\epsilon' := \epsilon'(h)$ such that the following holds: for any $0 < \epsilon < \epsilon'$, there exists an $n_0 := n_0(\epsilon, h)$ such that if $A \subseteq \mathbb{R}$ is of size*

$n \geq n_0$ and $|A.A| \leq |A|^{1+\epsilon}$, then

$$|hA| \geq |A|^{c \log h/2 - f_h(\epsilon)}$$

where c is an absolute constant, and $f_h(\epsilon) \to 0$ as $\epsilon \to 0$.

A similar theorem was also proved by Li [23] who generalized Solymosi's geometric approach to Theorem 8 to multifold sums by a clever induction argument. Konyagin [21] proved this theorem in yet another way, but neither Li nor Konyagin could improve the exponent to a super-logarithmic function. However, in the same paper, Croot and Hart were able to get a polynomial function in the exponent when one considered h-fold sums of $A.A$.

Theorem 32. *[9] For every $h \in \mathbb{N}$ there exists an $\epsilon = \epsilon(h) > 0$ such that the following holds: there exists an $n_0 := n_0(\epsilon, h)$ such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$ then*

$$|h(A.A)| \geq |A|^{\Omega((h/\log h)^{1/3})}.$$

A key ingredient to Theorem 32 is a corollary to the Tarry-Escott problem as well as combinatorially finding geometric progressions in sets with product sets of size $|A|^{1+\delta}$. Expanding upon this idea with several technical combinatorial lemmas involving graph theory, dependent random choice, Ruzsa calculus and other additive combinatorial tools, we prove the following strengthening of Theorem 31.

Theorem 33. *For any $h \in \mathbb{N}$, there exists an $\epsilon = \epsilon(h) > 0$ such that the following holds: there exists an $n_0 = n_0(\epsilon, h)$ such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$ and $|A.A| \leq |A|^{1+\epsilon}$, then*

$$|hA| \geq |A|^{c \exp \sqrt{\frac{1}{100} \log h}}$$

for some absolute constant c.

2.3.1 Layout and Notation.

In Section 2, we list some well-known additive combinatorial results that we will need. We also include several lemmas that are directly from [9]. For completeness, we include the proofs of these lemmas. In Section 3 and 4, we prove new, key lemmas that we will need to prove Theorem 33. Section 5 contains the proof of Theorem 33. In addition to the notation introduced in the beginning, we define the difference and quotient set as follows:

$$A - B := \{a - b : a \in A, b \in B\}$$

$$A/B := \{a/b : a \in A, b \in B \setminus \{0\}\}$$

All sets are assumed to be finite subsets of $\mathbb{R}$ unless indicated otherwise. The additive energy $E(A, B)$ is defined as

$$|\{(a, b, a', b') \in A \times B \times A \times B : a + b = a' + b'\}|.$$

We say that $f \gg g$ if $g = O(f)$ and $f \gg_k g$ if $f(n) \geq c(k)g(n)$ for n sufficiently large. We say that a polynomial $p(x)$ vanishes at $x = a$ to order j if $x = a$ is a root of order j but not $j + 1$. All graphs are finite and undirected. For a graph (G, E), $\Delta(G)$ denotes the maximum degree of G. We will abuse notation and denote $|G|$ as $|V(G)|$.

2.3.2 Lemmas and Known Results

The Plünnecke-Ruzsa inequality is ubiquitious in additive combinatorics and will be needed in our proof.

Lemma 34 (Plünnecke-Ruzsa Inequality)**.** *[25][31] Let A be a subset of a finite abelian group such that $|A + A| \leq c|A|$. Then, $|kA - \ell A| \leq c^{k+\ell}|A|$.*

We will also need the following lemma which exists in many different forms ([31], Chap. 2).

Lemma 35. *Let $X, Y \subseteq \mathbb{R}$. Then,*

$$|X+Y| \geq \frac{|X||Y|}{|(X-X)\cap(Y-Y)|}.$$

In particular, if $(X-X)\cap(Y-Y) = \{0\}$, then $|X+Y| = |X||Y|$.

Proof. The additive energy of X and Y can be bounded from above by

$$\begin{aligned} E(X,Y) &:= |\{(x,x',y,y') \in X\times X\times Y\times Y : x+y = x'+y'\}| \\ &= |\{(x,x',y,y') : x-x' = y-y'\}| \\ &= \sum_{t\in X-X\cap Y-Y} |\{(x,x',y,y') : x-x' = t = y-y'\}| \\ &\leq |(X-X)\cap(Y-Y)||X||Y| \end{aligned}$$

On the other hand, one can use Cauchy-Schwarz to bound the additive energy from below:

$$E(X,Y) = \sum_{s\in X+Y} |\{(x,y)\in X\times Y : x+y = s\}|^2 \geq \frac{|X|^2|Y|^2}{|X+Y|}.$$

Combining the two inequalities proves the lemma. □

We will use several lemmas from [9] whose proofs we include for completeness. First, we state a result of Wooley on the Tarry-Escott problem [32].

Theorem 36. *For every $k \geq 3$, there exists two distinct sets*

$$\{a_1,\dots,a_s\}, \{b_1,\dots,b_s\} \subseteq \mathbb{Z}$$

such that for all $j = 1,\dots,k$

$$\sum_{i=1}^{s} a_i^j = \sum_{i=1}^{s} b_i^j$$

but

$$\sum_{i=1}^{s} a_i^{k+1} \neq \sum_{i=1}^{s} b_i^{k+1}.$$

Moreover, $s < (5/8)(k+1)^2$.

We will need a useful corollary of this result.

Corollary 37. *For all $k \geq 2$, there exists a monic polynomial $f(x)$ having coefficients only $0, 1, -1$ having at most $2k^2$ nonzero terms such that $f(x)$ vanishes at $x = 1$ to order exactly k.*

Proof. For $k = 2, 3$, verify the corollary by hand by considering $(x-1)(x^2-1)$ and $(x-1)(x^2-1)(x^4-1)$. For $k \geq 4$, we use Theorem 36 as follows. Note that for $k \geq 4$, we have that $k^2 \geq (5/8)(k+1)^2$. Apply Theorem 36 to get two distinct sets $\{a_1, \ldots, a_s\}$ and $\{b_1, \ldots, b_s\}$ with the properties stated in the lemma. If these sets are not in $\mathbb{Z}_{\geq 0}$, then let $a := \min\{a_1, \ldots, a_s, b_1, \ldots, b_s\}$ otherwise, $a := 0$. Let

$$f(x) := x^{-a} \sum_{i=1}^{s} x^{a_i} - x^{b_i}.$$

Since the sets are distinct, it is clear that the polynomial is monic, has at most $2k^2$ nonzero terms, and only has coefficients 1, and -1. To see that f has the correct order of vanishing at $x = 1$, we use the fact that f vanishes at $x = 1$ to order exactly k if and only if f and its first $k-1$ derivatives vanish at $x = 1$, but the kth derivative does not. Let $1 \leq \ell \leq k-1$. Consider the ℓth derivative of f evaluated at $x = 1$:

$$\begin{aligned} f^{(\ell)}(1) &= \sum_{i=1}^{s} (a_i - a) \ldots (a_i - (\ell - 1) - a) - (b_i - a) \ldots (b_i - (\ell - 1) - a) \\ &= \sum_{i=1}^{s} a_i^{\ell} - b_i^{\ell} + g_{k-1} \cdot (a_i^{\ell-1} - b_i^{\ell-1}) + \ldots + g_1 \cdot (a_i - b_i) + g_0 \end{aligned}$$

where g_i is some constant depending only on i and a. Since the a_i, b_i satisfy the conditions of Theorem 36, the ℓth derivative is equal to zero if $1 \leq \ell \leq k-1$. Moreover, the kth derivative of f at $x = 1$ then simplifies to

$$f^{(k)}(1) = \sum_{i=1}^{s} a_i^k - b_i^k \neq 0.$$

So f has a zero at $x = 1$ of order precisely k. □

Lemma 38. *For every $k \in \mathbb{N}$, there exists an $n_0 = n_0(k)$ such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$ and no dyadic interval $[x, 2x)$ contains more than s elements of A, then,*

$$|kA| \gg_k \frac{|A|^k}{s^k}$$

Proof. Without loss of generality, we may assume half the elements of A are nonnegative, else, replace A with $-A$ and repeat the proof since $|kA| = |k(-A)|$. Denote the nonnegative elements as $A' := \{a_1 < \ldots < a_N\}$, and let

$$B := \{a_{2s}, a_{4s}, a_{6s}, \ldots, a_{(2\lfloor \frac{N}{2s} \rfloor)s}\}.$$

Now, consider kB. Suppose

$$b_1 + \ldots + b_k = b'_1 + \ldots + b'_k. \tag{9}$$

for some $b_1 < \ldots < b_k, b'_1 < \ldots < b'_k \in B$. We claim that this implies $b_i = b'_i$ for all $i = 1, \ldots, k$. Let $t \in \{1, \ldots, k\}$ be the largest integer such that $b_t \neq b'_t$. Without loss of generality, if $b_t > b'_t$, then in fact $b_t > 2b'_t$ since they belong to nonconsecutive dyadic intervals. Moreover,

$$b'_1 + \ldots + b'_{t-1} + b'_t \leq b'_t + b'_t < b_t < b_1 + \ldots + b_t.$$

Hence, all the sums $b_1 + \ldots + b_k$ are unique, and so

$$|kA| \geq |kB| = \binom{|B|}{k} \gg_k |B|^k \gg_k \frac{|A|^k}{s^k}.$$

□

Let $C \subseteq \mathbb{R}$. We call $C_0, \ldots, C_{k-1}$ a **decreasing partition** of C if

$$C = \bigcup_{i=0}^{k-1} C_i$$

and for any distinct $i, j \in \{0, \ldots, k-1\}$, if $i < j$, then $|c| > |d|$ for all $c \in C_i, d \in C_j$.

Lemma 39. *Suppose that $C \subseteq \mathbb{R} - \{0\}$, and let*

$$1 = \delta_0 > \delta_1 > \ldots > \delta_{k-1} > 0.$$

Moreover, suppose that C has the property that for any $c > d \in C$,

$$\frac{c}{d} - 1 > 2k\frac{\delta_i}{\delta_{i-1}}. \tag{10}$$

for all $i = 1, \ldots, k-1$. Then for any decreasing partition $C_0, \ldots, C_{k-1}$ of C, we must have that the sums

$$c_0\delta_0 + c_1\delta_1 + \ldots + c_{k-1}\delta_{k-1}$$

are distinct for all $(c_0, c_1, \ldots, c_{k-1}) \in C_0 \times C_1 \times \ldots \times C_{k-1}$.

Proof. Suppose

$$\sum_{i=0}^{k-1} c_i\delta_i = \sum_{i=0}^{k-1} c_i'\delta_i \tag{11}$$

where $c_i, c_i' \in C_i$. Let j be the smallest integer in $\{0, \ldots, k-1\}$ such that $c_j \neq c_j'$. Hence, we need only consider

$$\sum_{i=j}^{k-1} c_i\delta_i = \sum_{i=j}^{k-1} c_i'\delta_i \tag{12}$$

We will now derive a contradiction proving no such j exists and so (11) only holds when $c_i = c_i'$ for all i. For a contradiction, suppose $c_j > c_j'$. Dividing by $c_j'\delta_j$ on both sides and rearranging, the sum becomes

$$\frac{c_j}{c_j'} - 1 = \sum_{i=j+1}^{k-1} \frac{c_i' - c_i}{c_j'} \cdot \frac{\delta_i}{\delta_j}.$$

By (10), this implies that

$$\sum_{i=j+1}^{k-1} \frac{c_i' - c_i}{c_j'} \cdot \frac{\delta_i}{\delta_j} > 2k\frac{\delta_{j+1}}{\delta_j}.$$

On the other hand, since the C_i form a decreasing partition,

$$\left|\frac{c_i' - c_i}{c_j'}\right| < 2$$

for all $i \geq j+1$. Also, since $\delta_{j+1} > \delta_\ell > 0$ for all $\ell > j+1$

$$\frac{\delta_i}{\delta_j} < \frac{\delta_{j+1}}{\delta_j}.$$

So we get a contradiction since this would imply

$$\left|\sum_{i=j+1}^{k-1} \frac{c_i' - c_i}{c_j'} \cdot \frac{\delta_i}{\delta_j}\right| < 2k\frac{\delta_{j+1}}{\delta_j}.$$

□

2.3.3 Finding a Long Geometric Progression in A/A

The following two lemmas are variants of Lemma 2 in the work of Croot and Hart [9]. The first one is a repackaged version of the main idea in [10] which allows one to combinatorially find long progressions in difference (or quotient) sets. The second lemma builds upon the first by taking $(N+1)$-tuples and showing that one can project them in a way that satsfies properties we will need later on.

Lemma 40. *For any $\epsilon > 0$ and any integer $N \geq 2$, there exists a $\delta = \delta(\epsilon, N) > 0$ such that if $B \subseteq A \subseteq \mathbb{R}$ with $|B| \geq |A|^{\delta}$, $|A.A| < |A|^{1+\epsilon}$, and A sufficiently large, then the following holds. There exists $\alpha \in \mathbb{R}$ and $\theta \in B/B \setminus \{1\}$ such that there are at least $|A|^{N+2-7\epsilon N^2}$ tuples $(a, y_0, \ldots, y_N) \in A^{N+2}$ such that*

$$ay_i\theta^i \in \alpha A$$

for all $i = 0, \ldots, N$.

Proof. Let $\epsilon > 0$, $N \in \mathbb{N}$, and let $\delta = 7\epsilon N^2$. Let $B \subseteq A \subseteq \mathbb{R}$ be such that $|A.A| < |A|^{1+\epsilon}$ and $|B| \geq |A|^{\delta}$. Consider the following set E:

$$\{(b_1, b_2, a_1, a_2, u, v, y_0, \ldots, y_N, z_0, \ldots, z_N) \in B^2 \times A^{2N+6} : va_1b_1^iz_i = ua_2b_2^iy_i\}.$$

For a vector $\mathbf{t} = (t_0, \ldots, t_N) \in A^{(3)} \times A^{(4)} \ldots \times A^{(N+3)}$, let

$$r(\mathbf{t}) := |\{(b, v, a, z_0, \ldots, z_N) \in B \times A^{N+3} : vab^iz_i = t_i \text{ for } i = 0, \ldots, N\}|.$$

Note that here is where we use the fact that $B \subseteq A$ in order to assume that $vab^iz_i \in A^{(i+3)}$. Now, one can use the Cauchy-Schwarz inequality to bound the size of E:

$$|E| = \sum_{\mathbf{t}} r(\mathbf{t})^2 \geq \frac{|B|^2|A|^{2N+6}}{|A^{(3)}||A^{(4)}| \ldots |A^{(N+3)}|}$$

By the Plünnecke-Ruzsa inequality, since $|A.A| < |A|^{1+\epsilon}$, we then have that for all i, $|A^{(i)}| < |A|^{1+i\epsilon}$. Thus, for $N \geq 2$,

$$|E| \geq |B|^2|A|^{N+5-\epsilon(3+4+\ldots+N+3)} \geq |B|^2|A|^{N+5-6\epsilon N^2}.$$

We call a tuple in E trivial if $b_1 = b_2$ since this would lead to $\theta = 1$. The number of trivial solutions in E is at most $|B||A|^{N+5}$. Hence, for our choice of $\delta = 7\epsilon N^2$ and A sufficiently large, the number of nontrivial solutions is at least

$$|B|^2|A|^{N+5-6\epsilon N^2} - |B||A|^{N+5}$$

By the pigeonhole principle, there exists a $(b_1, b_2, u, v, a_1) \in B^2 \times A^3$ such that there are at least

$$\frac{1}{|B|^2|A|^3}\left(|B|^2|A|^{N+5-6\epsilon N^2} - |B||A|^{N+5}\right) \geq |A|^{N+2-7\epsilon N^2}$$

tuples $(a_2, y_0, \ldots, y_N, z_0, \ldots, z_N)$ such that for $i = 0, \ldots, N$

$$va_1b_1^i z_i = ua_2b_2^i y_i.$$

Rearranging the above, we get that

$$z_i = a_2\frac{u}{va_1}\left(\frac{b_2}{b_1}\right)^i y_i.$$

Letting $\alpha = \frac{va_1}{u}$ and $\theta = \frac{b_2}{b_1}$ proves the lemma. □

Lemma 41. *Let $N, \ell \in \mathbb{N}$, $\epsilon > 0$ and let $c = 2\ell^{\lceil \log_2 N \rceil}$. There exists an $n_0 = n_0(N, \ell, \epsilon)$, $\delta = \delta(\epsilon, N)$ such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$ then the following holds. If $|A.A| \leq |A|^{1+\epsilon}$, then for any $B \subseteq A$ with $|B| \geq |A|^\delta$ there exist $Y_0, \ldots, Y_N \subseteq A$ such that*

1. *$|Y_i| \geq |A|^{1-O(\epsilon c N^4)}$.*
2. *For any collection of subsets $Y_i' \subseteq Y_i$ satisfying $|Y_i'| \leq c$ there exists an $\alpha \in \mathbb{R}$, $\theta \in \frac{B}{B} \setminus \{1\}$, and an $A' \subseteq A$ of size at least $\sqrt{|A|}$ such that*

$$ay_i\theta^i \in \alpha A$$

for all $a \in A'$, $y_i \in Y_i'$, $i \in \{0, \ldots, N\}$.

We first need a graph theoretic lemma. It is a slight variant of a lemma found in the excellent survey by Fox and Sudakov about the technique of dependent random choice [17]. For a graph G and $T \subseteq G$, let $\Gamma(T)$ denote the set of common neighbors of T; that is, the set of all vertices adjacent to every vertex in T.

Lemma 42. *Let $\nu, m, r \in \mathbb{N}$. Let $G = [X, Y]$ be a bipartite graph with $|E(G)|$ edges. If there exists a $t \in \mathbb{N}$ such that*

$$\frac{|E(G)|^t}{|X|^t |Y|^{t-1}} - \binom{|Y|}{r} \left(\frac{m}{|X|}\right)^t \geq \nu$$

then there exists a set of vertices in Y of size ν such that every r of them have at least m common neighbors.

Proof. Let $T \subseteq X$ be a set of t vertices chosen uniformly at random with repetition. Let $\Gamma(T)$ denote the set of common neighbors of T, and let $Z = |\Gamma(T)|$. Then, by linearity of expectation and Hölder's inequality

$$\mathbb{E}(Z) = \sum_{y \in Y} \mathbb{P}(T \subseteq N(y)) = \sum_{y \in Y} \left(\frac{|N(y)|}{|X|}\right)^t \geq \frac{|E(G)|^t}{|X|^t |Y|^{t-1}}.$$

Now, let W be the random variable associated to the number of sets of r vertices in $\Gamma(T)$ with less than m common neighbors. We want W to be small so that we may modify all these deficient sets and prove the lemma. First, if a set $S \subseteq Y$ is also a subset of $\Gamma(T)$, then S is adjacent to every vertex in T by the definition of common neighborhood. Hence, the common neighborhood of S, $\Gamma(S)$, must contain T. The probability that we chose only elements from $\Gamma(S)$ when we chose T is

$$\left(\frac{|\Gamma(S)|}{|X|}\right)^t$$

Moreover, there are at most $\binom{|Y|}{r}$ such sets S if S has size r. Hence, the expected number of sets of r vertices in $\Gamma(T)$ with less than m common neighbors can be bounded as follows:

$$\mathbb{E}(W) \leq \left(\frac{|\Gamma(S)|}{|X|}\right)^t \binom{|Y|}{r} < \frac{m^t}{|X|^t} \binom{|Y|}{r}.$$

Therefore, there exists a choice of T such that

$$Z - W > \frac{|E(G)|^t}{|X|^t|Y|^{t-1}} - \binom{|Y|}{r}\left(\frac{m}{|X|}\right)^t \geq \nu.$$

Let T be chosen such that the above holds. For each set $S \subseteq \Gamma(T)$ of size r with less than m common neighbors, remove a vertex arbitrarily from S. After this process, $\Gamma(T)$ still has at least ν vertices left, and every set of size r has at least m common neighbors. □

Proof of Lemma 41. Apply Lemma 40 to get an $\alpha \in \mathbb{R}$ and a $b_1 > b_2 \in B$ such that there are $|A|^{N+2-7\epsilon N^2}$ tuples

$$T := (a, y_0, \ldots, y_N) \in A^{N+2}$$

such that

$$\alpha a y_i \theta^i \in A \text{ for } i = 0, \ldots, N \tag{13}$$

where $\theta = b_1/b_2$. Let $G[X, Y]$ be the bipartite graph defined by $X = A$, $Y = A^{N+1}$, and edges defined by the set T. Observe that for any constant r depending only on ℓ and N there exists a t and an ϵ such that if A is sufficiently large, then

$$\begin{aligned}\frac{|A|^{t(N+2-7\epsilon N^2)}}{|A|^t|A|^{(t-1)(N+1)}} - \binom{|A|^{N+1}}{r}\left(\frac{|A|^{t/2}}{|A|^t}\right) &\geq |A|^{N+1-7\epsilon t N^2} - |A|^{r(N+1)-t/2} \\ &\geq \frac{1}{2}|A|^{N+1-7\epsilon t N^2}.\end{aligned}$$

In particular, one may choose $t = 2r(N+1)$. Hence, we may apply Lemma 42 with $\nu = \frac{1}{2}|A|^{N+1-28\epsilon r N^3}$, $m = |A|^{1/2}$, and $r = c(N+1)$. Let $Y' \subseteq Y$ denote the set found by Lemma 42 with the specified property.

Each vertex $v \in Y'$ is associated to a corresponding $(N+1)$-tuple; for $i = 0, \ldots, N$, let Y_i be the projection of Y' onto the ith coordinate axis. One can see that $|Y_i| \geq |A|^{1-O(\epsilon c N^4)}$. Consider an arbitrary collection of subsets $Y_i' \subseteq Y_i$ satisfying $|Y_i'| \leq c$. Let $y_{i,j} \in Y_i'$. Our goal is to show there is a fixed set $A' \subseteq A$ of $|A|^{1/2}$ elements such that (13) holds for all $y_{i,j}$, $a \in A'$, $i = \{0, \ldots, N\}$.

Since $y_{i,j} \in Y_i' \subseteq Y_i$, there exists a corresponding $(N+1)$-tuple

$$(u_0, u_1, \ldots, u_{i-1}, y_{i,j}, u_{i+1}, \ldots, u_N) \in Y'.$$

For each $y_{i,j}$, arbitrarily choose such a tuple in Y', and denote the tuple as $v_{i,j}$. Let V be the collection of all such $v_{i,j}$. So, letting $|V| \leq c(N+1)$ be the constant r in the application of Lemma 42, we can conclude that there is a set of $|A|^{1/2}$ vertices in X adjacent to every vertex in V. Let A' be this set of $|A|^{1/2}$ vertices. Hence, there is a set of $|A|^{1/2}$ elements such that for any $y_{i,j} \in Y_i'$ (13) holds for all $a \in A'$, $i \in \{0, \ldots, N\}$. □

2.3.4 Intersections of Multifold Sumsets

We now prove the following lemma that gives us information when lots of multifold sumsets intersect trivially. This lemma is what introduces a significant amount of loss in the strength of our overall bound in Theorem 33 – that is, it is the main obstruction in improving the exponent $\exp(c\sqrt{\log h})$ to some fixed power of h.

Lemma 43. *Let $A \subseteq \mathbb{R}$ and $\ell, t \in \mathbb{N}$. Let $A_i \subseteq A$ for $i = 1, \ldots, 2^t$ be such that*

$$\bigcap_{i=1}^{2^t} f(t,i)\ell^{g(t,i)}A_i - f(t,i)\ell^{g(t,i)}A_i = \{0\}.$$

Then, if $|A_i| \geq n$, then there exists an $i \in \{2, \ldots, t+1\}$ and an $j \in \{1, \ldots, 2^t\}$ such that

$$|(\ell^{i-1} + \ell^i)A| \geq n^{\frac{1}{3^{t+1}}}|\ell^i A_j|.$$

The functions f and g in the above lemma are defined as follows. For $a \in \mathbb{N}$, $b = 1, \ldots 2^a$, define $f(a,b)$ recursively as follows:

$$f(1,1) := 1, f(1,2) := 2,$$

$$f(a, 2b-1) := f(a-1, b); b = 1, \ldots, 2^{a-1} \tag{14}$$

$$f(a, 2b) := 2f(a, 2b-1) = 2f(a-1, b); b = 1, \ldots, 2^{a-1} \tag{15}$$

For the benefit of the reader, we list the first few values of $f(a,b)$:

$$f(1,1)=1; f(1,2)=2$$

$$f(2,1)=1; f(2,2)=2; f(2,3)=2; f(2,4)=4$$

$$f(3,1)=1; f(3,2)=2; f(3,3)=2; f(3,4)=4;$$

$$f(3,5)=2; f(3,6)=4; f(3,7)=4; f(3,8)=8$$

Observe that

$$f(a,b)=2^k \text{ for some } k \leq a. \tag{16}$$

Denote $g(a,b) := \log_2 f(a,b)+1$. Observe that by (15),

$$g(a,2b)=g(a,2b-1)+1 \tag{17}$$

and by (16),

$$g(a,b) \leq a+1. \tag{18}$$

The following covering lemma, which is potentially of independent interest, is the main tool in proving Lemma 43.

Lemma 44. *For any X, Y in an abelian group G and any integer $1 \leq K \leq \sqrt{|X|}$, there exists an $X' \subseteq X$ such that either*

1. *$|X'| \geq K$ and $X' - X' \cap Y - Y = \{0\}$, or*
2. *$|X'| \geq \frac{|X|}{K}$ and $X' - X' \subseteq 2Y - 2Y$.*

This follows quickly from the following graph theoretic lemma.

Lemma 45. *For any graph G and any $1 \leq K \leq \sqrt{|G|}$, G contains an independent set of size at least K or a vertex of degree at least $|G|/K$.*

Proof. If G has a vertex of degree at least $|G|/K$, we are done. Hence, the maximum degree of G, $\Delta(G)$ is less than $|G|/K$. By the greedy algorithm, we can find an independent set of size

$$\left\lfloor \frac{|G|-1}{\Delta+1} \right\rfloor + 1 = \left\lfloor \frac{|G|+\Delta}{\Delta+1} \right\rfloor. \tag{19}$$

If $\Delta \geq K$, by the right hand side of (19), we can find an independent set of size

$$\left\lfloor \frac{|G| + \Delta}{\Delta + 1} \right\rfloor > \left\lfloor \frac{|G| + \Delta}{\frac{|G|}{K} + 1} \right\rfloor = \left\lfloor \frac{K(|G| + \Delta)}{|G| + K} \right\rfloor \geq K.$$

If $\Delta < K$, then by the left hand side of (19), we can find an independent set of size

$$\left\lfloor \frac{|G| - 1}{\Delta + 1} \right\rfloor + 1 > \left\lfloor \frac{|G| - 1}{K} \right\rfloor + 1.$$

Let $|G| = q \cdot K + r$ where $0 \leq r \leq k - 1$, and for any $x \in \mathbb{R}$, let $[x] = x - \lfloor x \rfloor$ denote the fractional part of x. Then,

$$\begin{aligned}
\left\lfloor \frac{|G| - 1}{K} \right\rfloor + 1 = \frac{|G| - 1}{K} + 1 - \left[\frac{|G| - 1}{K}\right] &= \frac{|G|}{K} + \frac{K - 1}{K} - \left[q + \frac{r - 1}{K}\right] \\
&= \frac{|G|}{K} + \frac{K - 1}{K} - \left[\frac{r - 1}{K}\right] \\
&= \frac{|G|}{K} + \frac{K - 1}{K} - \frac{r - 1}{K} > \frac{|G|}{K} \geq K.
\end{aligned}$$

Note that we used the fact that $K \leq \sqrt{|G|}$ in the last inequality. Hence, we can find an independent set of size K if $\Delta < |G|/K$. □

Proof of Lemma 44. Let $G = (V, E)$ be the graph defined by $V(G) := X$ and $\{u, v\} \in E(G)$ if $u - v \in Y - Y$. Observe that since $Y - Y$ is symmetric, these edges are undirected. If G contains an independent set X' of size at least K, for any distinct $u, v \in X'$, $u - v \notin Y - Y$. Hence, $X' - X' \cap Y - Y = \{0\}$. Otherwise, G contains a vertex, a, of degree at least $\frac{|X|}{K}$. Letting the neighborhood of this vertex be X', for any $u, v \in X'$, $\{u, a\}$ and $\{a, v\}$ are edges. Since $u - v = u - a + a - v$, we have that $u - v \in 2Y - 2Y$. □

Proof of Lemma 43. We perform the following algorithm to find such an i, j as in the conclusion of the lemma. We outline steps $j = 0, \ldots, t - 2$.

Step 0: Let $A_{0,i} := \ell^{g(t,i)} A_i$. For $i = 1, \ldots, 2^{t-1}$, apply Lemma 44 with

$$X := A_{0,2i-1},\ Y := A_{0,2i},\ \text{and}\ K := K_0 = n^{\frac{1}{3^t}},$$

and observe which case holds. If for any i, Case 1 holds, we halt since this implies that there exists an $X' \subseteq X$ with $|X'| \geq n^{\frac{1}{3^t}}$ and

$$|(\ell^{g(t,2i-1)} + \ell^{g(t,2i)})A| \geq |A_{0,2i-1} + A_{0,2i}| \geq |X' + Y| = |X'||Y| \geq n^{\frac{1}{3^t}}|\ell^{g(t,2i)}A_{2i}|.$$

This satisfies the conclusion of the lemma with $k = g(t, 2i)$ and $j = 2i$. Hence, we may assume Case 2 holds for all i. Therefore, there exists an $X' \subseteq X$ such that $X' - X' \subseteq 2Y - 2Y$. Adding $X' - X'$ to itself multiple times also implies for any positive integer s, $sX' - sX' \subseteq 2sY - 2sY$. In particular for $s = f(t, 2i-1)$,

$$\begin{aligned} f(t,2i-1)X' - f(t,2i-1)X' &\subseteq 2f(t,2i-1)Y - 2f(t,2i-1)Y \\ &= 2f(t,2i-1)A_{0,2i} - 2f(t,2i-1)A_{0,2i} \\ &= f(t,2i)A_{0,2i} - f(t,2i)A_{0,2i}. \end{aligned} \tag{20}$$

where we used (15) in the last equality. Also,

$$\begin{aligned} f(t,2i-1)X' - f(t,2i-1)X' &\subseteq f(t,2i-1)X - f(t,2i-1)X \\ &= f(t,2i-1)A_{0,2i-1} - f(t,2i-1)A_{0,2i-1} \end{aligned} \tag{21}$$

Letting $A_{1,i} := X'$, we then have that by (14), (20), and (21)

$$\begin{aligned} \bigcap_{i=1}^{2^{t-1}} f(t-1,i)A_{1,i} - f(t-1,i)A_{1,i} &\subseteq \bigcap_{i=1}^{2^t} f(t,i)A_{0,i} - f(t,i)A_{0,i} \\ &= \bigcap_{i=1}^{2^t} f(t,i)\ell^{g(t,i)}A_i - f(t,i)\ell^{g(t,i)}A_i = \{0\}. \end{aligned}$$

And we also have that

$$|A_{1,i}| \geq \frac{|A_{0,2i-1}|}{K_0}.$$

The next steps, Steps $j = 1, \ldots, t-2$, are iterations of this argument with a very slight change in the choice of X and Y in the application of Lemma 44.

Step j: For $1 \leq j \leq t-2$, let $A_{j,i} \subseteq A_{j-1,2i-1}$ be as specified in Step (j-1) of the algorithm. In particular, $A_{j,i}$ satisfies

$$|A_{j,i}| \geq \frac{|A_{j-1,2i-1}|}{K_{j-1}}.$$

An easy inductive argument shows that there exists an r such that

$$A_{j,i} \subseteq A_{j-1,2i-1} \subseteq \ldots \subseteq A_{0,r} \subseteq \ell^{g(t-j,i)} A_r. \tag{22}$$

where we draw the reader's attention to the fact that the subscript $A_{j,i}$ determines the exponent at the end, $g(t-j,i)$. For $i = 1, \ldots, 2^{t-j-1}$, apply Lemma 44 with $X = A_{j,2i-1}$, $Y = A_{j,2i}$, $K := K_j = n^{\frac{1}{3^{t-j}}}$. We first check that $K \leq \sqrt{X}$. We have that

$$|X| \geq \frac{n}{K_0 K_1 \cdot \ldots \cdot K_{j-1}} = n^{1-\sum_{i=t-j+1}^{t} 3^{-i}} \geq \sqrt{n}.$$

On the other hand, since $j \leq t-2$,

$$K = n^{\frac{2}{3^{t-j}}} \leq n^{\frac{2}{9}} \leq n^{\frac{1}{4}} \leq \sqrt{|X|}.$$

Now, observe which case holds in our application of Lemma 44. If for any i, Case 1 holds, we halt since by Lemma 35 this implies that

$$\begin{aligned} |X' + Y| = |X'||Y| &\geq K_j |A_{j,2i}| \\ &\geq \frac{K_j}{K_{j-1}} |A_{j-1,4i-1}| \\ &\vdots \\ &\geq \frac{K_j}{K_{j-1} K_{j-2} \ldots K_0} |A_{0,r}| \\ &\geq n^{\frac{1}{3^{t+1}}} |\ell^{g(t-j,2i)} A_r|. \end{aligned} \tag{23}$$

On the other hand, using (22) and (17), we have

$$\begin{aligned} |X' + Y| \leq |X + Y| &= |A_{j,2i-1} + A_{j,2i}| \\ &\leq |A_{j-1,4i-3} + A_{j-1,4i-1}| \\ &\vdots \\ &\leq |A_{0,r'} + A_{0,r}| \leq |(\ell^u + \ell^{u+1})A| \end{aligned} \tag{24}$$

for $u = g(t-j, 2i-1)$ and some integer r'. Combining (23) and (24) shows that we have satisfied the conclusion of the Lemma.

Hence, we may assume Case 2 holds for all i. Therefore, there exists an $X' \subseteq X$ with $|X'| \geq |X|/K_j$ such that $X' - X' \subseteq 2Y - 2Y$. Moreover, for any positive integer s, $sX' - sX' \subseteq 2sY - 2sY$. For $s = f(t-j-1,i)$

$$\begin{aligned} f(t-j-1,i)X' - f(t-j-1,i)X' &\subseteq 2f(t-j-1,i)Y - 2f(t-j-1,i)Y \\ &= 2f(t-j-1,i)A_{j,2i} - 2f(t-j-1,i)A_{j,2i} \\ &= f(t-j,2i)A_{j,2i} - f(t-j,2i)A_{j,2i}. \end{aligned} \tag{25}$$

where we used (14) in the last equality. Also,

$$\begin{aligned} f(t-j-1,i)X' - f(t-j-1,i)X' &\subseteq f(t-j-1,i)X - f(t-j-1,i)X \\ &= f(t-j,2i-1)A_{j,2i-1} - f(t-j,2i-1)A_{j,2i-1} \end{aligned} \tag{26}$$

Letting $A_{j+1,i} := X'$, we then have that by (15), (25), and (26)

$$\bigcap_{i=1}^{2^{t-j-1}} f(t-j-1,i)A_{j+1,i} - f(t-j-1,i)A_{j+1,i} \subseteq \bigcap_{i=1}^{2^{t-j}} f(t-j,i)A_i - f(t-j,i)A_i = \{0\}.$$

We now proceed to Step j+1 with $A_{j+1,i}$, $i = 1, \ldots, t-j-1$.

Step t − 1: If we have not halted, then at this point, we only have 2 sets, $A_{t-1,1}, A_{t-1,2}$, such that

$$f(1,1)A_{t-1,1} - f(1,1)A_{t-1,1} \cap f(1,2)A_{t-1,2} - f(1,2)A_{t-1,2} = \{0\}.$$

Since $f(1,1) = 1$, $f(1,2) = 2$, and

$$A_{t-1,1} - A_{t-1,1} \cap A_{t-1,2} - A_{t-1,2} \subseteq A_{t-1,1} - A_{t-1,1} \cap 2A_{t-1,2} - 2A_{t-1,2} = \{0\}$$

we then have by Lemma 35

$$|A_{t-1,1} + A_{t-1,2}| = |A_{t-1,1}||A_{t-1,2}|.$$

Tracing back our steps in the algorithms as we did in (23) and (24), we get that

$$\begin{aligned} |A_{t-1,1}||A_{t-1,2}| \geq \frac{|A_{t-2,1}||A_{t-2,3}|}{K_{t-1}^2} &\geq \frac{|A_{t-3,1}||A_{t-3,5}|}{K_{t-1}^2 K_{t-2}^2} \\ &\geq \frac{|A_{t-3,1}||A_{t-3,9}|}{K_{t-1}^2 K_{t-2}^2 K_{t-3}^2} \\ &\vdots \\ &\geq \frac{|A_{0,1}||A_{0,2^{t-1}+1}|}{K_{t-1}^2 K_{t-2}^2 K_{t-3}^2 \dots K_0^2} \geq n^{\frac{1}{3^{t+1}}} |\ell^2 A_{2^{t-1}+1}| \end{aligned} \tag{27}$$

Note that we used the fact that $|A_{0,1}| \geq n$ in the last inequality. On the other hand,

$$\begin{aligned} |A_{t-1,1} + A_{t-1,2}| \leq |A_{t-2,1} + A_{t-2,3}| &\leq |A_{t-3,1} + A_{t-3,5}| \\ &\vdots \\ &\leq |A_{0,1} + A_{0,2^{t-1}+1}| \\ &\leq |\ell A_1 + \ell^2 A_{2^{t-1}+1}| \leq |(\ell + \ell^2)A| \end{aligned} \tag{28}$$

Combining (27) and (28) completes the proof of the lemma. □

2.3.5 Proof of Main Theorem

The proof of our main theorem is iterative. The argument splits into two cases: in one case, we prove our bound directly similar to [9]; the other case we have to iteratively use Lemma 43 to get a small amount of growth each iteration while passing to subsets of our original set. After enough iterations, we prove our bound.

Proposition 46. *Let $h \in \mathbb{N}$. Let*

$$k := \exp \sqrt{\frac{1}{100} \log \frac{h}{2}} \text{ and } \ell := k^8.$$

There exists an $\epsilon' := \epsilon'(h)$ such that for any $0 < \epsilon < \epsilon'$ there exists an $n_0 := n_0(\epsilon, h)$ such that if $A \subseteq \mathbb{R}$ is of size $n \geq n_0$ and $|A.A| \leq |A|^{1+\epsilon}$, then either

$$|hA| \geq |A|^{\Omega(k)}$$

or there exists an $A' \subseteq A$ and a $c := c(h)$ such that $|A'| \geq |A|^{1-c\epsilon}$, and

$$|(\ell^i + \ell^{i-1})A| \gg_h |A|^{\frac{1}{22k^6}} |\ell^i A'|$$

for some $i \in \{2, \ldots, \log(8k^5) + 1\}$.

Proof of Proposition 46. Let $A \subseteq \mathbb{R}$ be such that $|A.A| \leq |A|^{1+\epsilon}$. Let k, ℓ be constants depending on h as specified in the statement of the proposition. Apply Corollary 37 to get a set of polynomials $f_j(x)$ for $j = 2, \ldots, k-1$ such that each polynomial has coefficients in $\{-1, 0, 1\}$, $f_j(x)$ has a root at $x = 1$ of order exactly j, and $f_j(x)$ has at most $2j^2 \leq 2k^2$ nonzero terms. Let $f_0(x) := 1$, and $f_1(x) := x - 1$.

$$N := \max_j \{deg(f_j) : j = 0, \ldots, k-1\}$$

and let $S \subseteq \{0, \ldots, N\}$ be such that $i \in S$ if and only if there is an $f_j(x)$ such that the coefficient of x^i is nonzero. Let $M := |S|$ and observe that $M \leq 2k^3$.

Denote $A := \{a_1 < \ldots < a_n\}$, let $0 < \delta < 1/4$ be a parameter chosen later, and let $s := \lfloor n^\delta \rfloor$. Let

$$B' := \{a_i, a_{i+1}, \ldots, a_{i+s-1}\}$$

be chosen such that a_{i+s-1}/a_i is minimal. By Lemma 38, if no dyadic interval contains more than s elements of A, we are done. Hence, $B' \subseteq [x, 2x)$ for some $x \in \mathbb{R}$. Let $0 < \gamma < 1$ be a small constant depending on h to be chosen later. There exists a subinterval

$$[y, y + \gamma x) \subseteq [x, 2x)$$

with at least γs elements of A in it. Let B be the intersection of A with this subinterval. So $B \subseteq A$ has the properties that $|B| \geq \gamma s$ and for any $b, b' \in B$,

$$\left| \frac{b}{b'} - 1 \right| < \gamma.$$

The latter property will be important when we later consider polynomials with roots at 1 evaluated at $\frac{b}{b'}$.

Apply Lemma 41 with N, ℓ, ϵ, B to find a set of $Y_i \subseteq A$, $\alpha \in \mathbb{R}$, $\theta \in B/B$, satisfying the conclusion of the lemma. We will discard some of the sets from $Y_0, \ldots, Y_N$ in the

following way. If $i \notin S$, then we throw out Y_i. Abusing our notation, relabel the remaining sets as $Y_1, \ldots, Y_M$. Let $t = \lceil \log_2 M \rceil \leq \lceil \log_2 2k^3 \rceil$. If

$$\bigcap_{i=1}^{M} \ell^t Y_i - \ell^t Y_i = \{0\}$$

then we may apply Lemma 43 to conclude that there exists an $i \in \{2, \ldots, t+1\}$ and a $j \in \{1, \ldots, 2^t\}$ such that for ϵ sufficiently small,

$$|(\ell^{i-1} + \ell^i)A| \geq (n^{1-O(\epsilon c N^4)})^{\frac{1}{3^{t+1}}} |\ell^i Y_j| \geq |A|^{\frac{1}{22k^6}} |\ell^i Y_j|.$$

This satisfies the second conclusion of the proposition, so we may assume that there exists a nonzero β in the above intersection. That is, a nonzero β such that for $i = 1, \ldots, M$,

$$\beta = \sum_{j=1}^{\ell^t} y_{i,j} - \sum_{j=\ell^t+1}^{2\ell^t} y_{i,j}$$

where $y_{i,j} \in Y_i$. Letting $Y_i' := \{y_{i,j} : j = 1, \ldots, 2\ell^t\}$, by the conclusion of Lemma 41, there exists an $A' \subseteq A$ of size at least $|A|^{1/2}$ such that

$$a y_{i,j} \theta^i \in \alpha A \text{ for } i = 1, \ldots, M, \text{ and any } a \in A'. \tag{29}$$

Denote $A' := \{a_1 < a_2 < \ldots < a_{|A'|}\}$, and let $C := \{a_{i_1}, a_{i_2}, \ldots, a_{i_r}\}$ where

$$i_j = j\lfloor n^{1/4} \rfloor \text{ and } r = \left\lfloor \frac{|A'|}{n^{1/4}} \right\rfloor.$$

This ensures that we have

$$\frac{c}{c'} > \theta \text{ for any } c, c' \in C \tag{30}$$

by our choice of B' along with the fact that $s < \lfloor n^{1/4} \rfloor$. Decompose C into disjoint sets $C_0, \ldots, C_{k-1}$ where all elements of C_i are greater than all elements of C_j for $i < j$, and for all $i = 0, \ldots, k-2$, $|C_i| = \lfloor |C|/k \rfloor$. For $i = 0, \ldots, k-1$, let $\delta_i := f_i(\theta)$. Now consider sums of the form

$$\Sigma = \{\beta(c_0\delta_0 + c_1\delta_1 + \ldots + c_{k-1}\delta_{k-1}) : c_i \in C_i\}. \tag{31}$$

We verify that C and δ_i satisfy the requirements of Lemma 39 as follows. Since

$$\frac{\delta_i}{\delta_{i-1}} = \frac{f_i(x)}{f_{i-1}(x)} = (x-1)\frac{g_i(x)}{g_{i-1}(x)}$$

where the coefficients of g_i and g_{i-1} depend only on k, we may choose γ small enough such that

$$\theta - 1 < \frac{g_{i-1}(\theta)}{g_i(\theta)}.$$

So we have that $\delta_{i-1} > \delta_i$ for all $i = 1, \ldots, k-1$. Let $c, d \in C$. From (30), we have that $\frac{c}{d} > \theta$. However, by choosing δ small enough, we can assume that in fact $\frac{c}{d} > \theta^r$ for some $r = r(k)$ to be specified later. Hence,

$$\frac{c}{d} - 1 > \theta^r - 1 = (\theta - 1)(1 + \theta + \ldots + \theta^{r-1}) \geq (\theta - 1)r.$$

By choosing $r > 2k \cdot \frac{g_i(\theta)}{g_{i-1}(\theta)}$, we have

$$\frac{c}{d} - 1 \geq (\theta - 1)2k \cdot \frac{g_i(\theta)}{g_{i-1}(\theta)} = 2k\frac{\delta_i}{\delta_{i-1}}.$$

So by Lemma 39, all the sums of the form (31) are distinct, and so

$$|\Sigma| \geq \prod_{i=0}^{k-1} |C_i|.$$

We can rewrite (31) by grouping like powers of θ as

$$\beta\left[\left(\sum_{i=0}^{k-1} \epsilon_{0,i}c_i\right)\theta^0 + \left(\sum_{i=0}^{k-1} \epsilon_{1,i}c_i\right)\theta^1 + \ldots + \left(\sum_{i=0}^{k-1} \epsilon_{M,i}c_i\right)\theta^N\right]$$

where $\epsilon_{i,j} \in \{-1, 0, 1\}$. Recall that S is the set of powers of θ that have at least one nonzero coefficient in some polynomial f_j. Denoting S as $i_1 < i_2 < \ldots < i_M$, we can rewrite the above as

$$\beta\left[\left(\sum_{i=0}^{k-1} \epsilon_{i_1,i}c_i\right)\theta^{i_1} + \left(\sum_{i=0}^{k-1} \epsilon_{i_2,i}c_i\right)\theta^{i_2} + \ldots + \left(\sum_{i=0}^{k-1} \epsilon_{i_M,i}c_i\right)\theta^{i_M}\right].$$

Distribute β to each summand, and expand it uniquely for each power of θ to get

$$= \sum_{j=0}^{k-1}\sum_{i=1}^{\ell^t} \epsilon_{i_1,j}c_j(y_{1,i} - y_{1,\ell^t+i})\theta^{i_1} + \ldots + \sum_{j=0}^{k-1}\sum_{i=1}^{\ell^t} \epsilon_{i_M,j}c_j(y_{M,i} - y_{M,\ell^t+i})\theta^{i_M} \qquad (32)$$

Since our choices of θ and $y_{i,j}$ satisfy (29), we have that each element in this sum is in $\pm\alpha * A$. Hence, we have that for ℓ_1, ℓ_2 large enough,

$$|\ell_1(\alpha * A) - \ell_2(\alpha * A)| = |\ell_1 A - \ell_2 A| \geq \prod_{i=0}^{k-1} |C_i| \geq \left\lfloor \frac{|C|}{k} \right\rfloor^{k-1} \gg_k |A|^{\frac{k-1}{4}}.$$

Recall that $\ell = k^8$, $M \leq 2k^3$, and $t = \lceil \log_2 M \rceil \leq \lceil \log_2 2k^3 \rceil \leq \log_e 8k^5$. Also, recall that M is the number of powers of θ that occur in σ. Each θ has at most k coefficients, and $\beta \in \ell^t A - \ell^t A$. So, we have at most $kM \cdot 2\ell^t$ nonzero terms in $\sigma \in \Sigma$. We bound this as

$$kM \cdot 2\ell^t \leq 2k^4 k^{8 \log_e 8k^5} \leq 2k^{100 \log k}$$

So, choosing $k := \exp \sqrt{\frac{1}{100} \log \frac{h}{2}}$ proves our proposition:

$$|hA| \geq \sqrt{|hA - hA|} \geq |A|^{\Omega(\exp \sqrt{\frac{1}{100} \log h})}$$

□

2.3.6 The Iterative Case

We are now able to prove Theorem 33.

Proof of Theorem 33. We iteratively apply Proposition 46 in the following algorithm.

Step 0: Let k and ℓ be functions of h as specified in the statement of Proposition 46, and let $0 < \epsilon < \epsilon'$ where ϵ' is some unspecified function of h taken to be sufficiently small. Let $\ell_0 := \ell$, $A_0 := A$, and $\epsilon_0 := \epsilon$. Since $|A_0.A_0| \leq |A_0|^{1+\epsilon_0}$, we may apply Proposition 46 to A_0. If $|hA_0| \geq |A_0|^{\Omega(k)}$, then we are done. Else, there exists an $i \in \{2, \ldots, \log_e(8k^5) + 1\}$ and an $A_0' \subseteq A_0$ such that

$$|(\ell^i + \ell^{i-1})A_0| \gg_h |A_0|^{\frac{1}{22k^6}} |\ell^i A_0'| \text{ and } |A_0'| \geq |A_0|^{1-c\epsilon}$$

where c is a constant depending on h. Let $A_1 := A_0'$ and continue to Step 1.

For $j = 1, \ldots, \frac{1}{2}\ell$, we do the following.

Step j: Let A_j be as specified in the previous step. Since

$$|A_j.A_j| \leq |A_{j-1}.A_{j-1}| \leq |A_{j-1}|^{1+\epsilon_{j-1}} \leq |A_j|^{\frac{1+\epsilon_{j-1}}{1-c\epsilon_{j-1}}} \leq |A_j|^{1+2c\epsilon_{j-1}}$$

where we assumed ϵ_{j-1} is sufficiently small in the last inequality. Let $\epsilon_j := 2c\epsilon_{j-1}$. Let $\ell_j := \ell - j$. This determines h_j and k_j as

$$h_j = e^{\frac{25}{36}(\log(\ell-j))^2} \; ; \; k_j = (k^8 - j)^{1/8}.$$

Applying Proposition 46 to A_j with and h_j, we get that either

$$|hA| \geq |h_j A_j| \geq |A_j|^{\Omega(k_j)} \geq |A_j|^{\Omega((k^8-j)^{1/8})} = |A_j|^{\Omega(k)} \geq |A|^{(1-(2c)^j\epsilon)\cdot\Omega(k)} = |A|^{\Omega(k)}$$

which proves the theorem for ϵ sufficiently small – so we exit the algorithm. Or, there exists an $A'_j \subseteq A_j$ of size $|A'_j| \geq |A_j|^{1-c\epsilon_j}$ and a $t_j \in \{2, \ldots, \log(8k^5) + 1\}$ such that

$$|(\ell_j^{t_j} + \ell_j^{t_j-1})A_j| \geq |A_j|^{\frac{1}{22k^6}} |\ell_j^{t_j} A'_j| \geq |A_j|^{\frac{1}{22k^6}} |\ell_j^{t_j} A'_j| \geq n^{\frac{1}{23k^6}} |\ell_j^{t_j} A'_j|$$

where we used the fact that ϵ is sufficiently small and n is sufficiently large depending on h in the last inequality. Letting $A_{j+1} := A'_j$ we continue to Step $j+1$.

Analysis of Algorithm: Since $\ell_j = \ell - j$, and we perform at most $\ell/2$ steps, $\ell_j \geq \ell/2$. Assume the algorithm runs and finishes Step $\ell/2$. Each step in the algorithm produces a $t_j \in \{2, \ldots, \log(8k^5) + 1\}$. By averaging, there is some integer $s \in \{2, \ldots, \log(8k^5)+1\}$ that appears in the algorithm at least $\frac{\ell}{2(\log(8k^5)+1)}$ times. Denote $j_1, \ldots, j_q$ as the steps in which s is chosen. It is easy to verify that by the definition of ℓ_j,

$$\ell_j^2 + \ell_j \leq \ell_{j-1}^2,$$

and so we must also have that

$$\ell_j^s + \ell_j^{s-1} \leq \ell_{j-1}^2 \cdot \ell_j^{s-2} \leq \ell_{j-1}^s.$$

So,

$$\begin{aligned}
|(\ell_{j_1}^s + \ell_{j_1}^{s-1})A_{j_1}| &\geq n^{\frac{1}{23k^6}} |\ell_{j_1}^s A'_{j_1}| \geq \\
&\geq n^{\frac{1}{23k^6}} |(\ell_{j_2}^s + \ell_{j_2}^{s-1})A_{j_2}| \geq n^{\frac{2}{23k^6}} |\ell_{j_2}^s A'_{j_2}| \geq \\
&\vdots \\
&\geq n^{\frac{q}{23k^6}} |(\ell_{j_q}^s + \ell_{j_q}^{s-1})A_{j_q}| \geq n^{\frac{\ell}{2(\log(8k^5)+1)} \cdot \frac{1}{23k^6}} = n^{\Omega(k)}
\end{aligned}$$

where we used the fact that $q \geq \frac{\ell}{2(\log(8k^5)+1)}$ and $\ell = k^8$ in the last inequality. Since

$$\ell_{j_1}^s \leq \ell^{\log 8k^5} \leq k^{8(\log 8k^5)} \leq k^{100 \log k} = h$$

we have that

$$|hA|^2 \geq |(\ell_{j_1}^s + \ell_{j_1}^{s-1})A_{j_1}| \geq n^{\Omega(k)}$$

proving our theorem. □

CHAPTER III

ORDER-PRESERVING FREIMAN ISOMORPHISMS

Let G_1 and G_2 be additive groups, and let $A \subseteq G_1$ and $B \subseteq G_2$. A Freiman k-homomorphism is a map $\phi : A \to B$ such that

$$\phi(x_1) + \ldots + \phi(x_k) = \phi(y_1) + \ldots + \phi(y_k)$$

whenever

$$x_1 + \ldots + x_k = y_1 + \ldots + y_k.$$

Such a map ϕ is called a Freiman k-isomorphism if the converse holds as well. If A and B have an ordering, then ϕ is order-preserving when

$$\phi(a) < \phi(b) \text{ if and only if } a < b.$$

A Freiman 2-isomorphism will frequently be refered to as just a Freiman isomorphism. Freiman isomorphisms are used to transfer an additive set A in some arbitrary abelian group into a more amenable ambient group or set (such as $\mathbb{R}$, $\mathbb{Z}_N$, or $[1, n]$) while preserving the additive structure of A. We refer the interested reader to Chapter 5 of [31] for a detailed exposition on the various uses of Freiman isomorphisms.

The main tool we introduce in this paper allows one to find an order-preserving Freiman isomorphism from a set of n integers to the interval $[-cn, cn] \cap \mathbb{Z}$ where c is not too large provided that the original set is additively structured. We call this tool a 'Condensing Lemma' since, in a sense, it allows one to view sets with small doubling as dense subsets of an interval.

Theorem 47. *[Condensing Lemma] For any $K > 0$, there exists a c_1, c_2 such that if $A \subseteq \mathbb{Z}$ is such that $|A + A| \leq K|A|$ then the following holds: there exists $A' \subseteq A$ with*

$|A'| \geq c_1|A|$, *and there exists an order-preserving Freiman 2-isomorphism* $\phi : A' \to [-c_2|A'|, c_2|A'|] \cap \mathbb{Z}$.

Since the constants c_1 and c_2 depend exponentially on K, we do not bother specifying their exact value. In order to prove the Condensing Lemma, we need a version of the so-called Bogolyubov-Ruzsa lemma which guarantees us a large generalized arithmetic progression G in $2A - 2A$ when A has a small doubling. For more on this important result, we refer the reader to the recent work by Sanders [26] who gives the best-known bounds for the constants c_1, c_2, and c_3 stated below.

Theorem 48 (Bogolyubov-Ruzsa Lemma)**.** *Suppose* $A \subseteq \mathbb{Z}$ *satisfies* $|A+A| \leq K|A|$. *Then, there exists absolute constants* c_1, c_2, c_3 *dependent only on* K *such that* $2A-2A$ *contains a proper, symmetric, generalized arithmetic progression* G *of dimension at most* c_1 *and size at least* $c_2|A|$. *Moreover, for each* $x \in G$, *there are at least* $c_3|A|^3$ *quadruples* $(a, b, c, d) \in A^4$ *with* $x = a + b - c - d$.

Along with the Plünnecke-Ruzsa estimates, one can also deduce that $|G| \leq K^4|A|$.

Theorem 49 (Plünnecke-Ruzsa Inequality [31])**.** *If* $|A + A| \leq K|A|$, *then for any positive integers* ℓ, m, *we have that* $|\ell A - mA| \leq K^{\ell+m}|A|$.

For our purposes, we will define a generalized arithmetic progression G as $G = \{\sum_{i=1}^{k} x_i d_i : |x_i| \leq L_i\}$. The proof of the Condensing Lemma consists of first applying Sanders' theorem so that we may approximate A by a generalized arithmetic progression G. Then, after passing to certain subsets, we use some techniques from convex geometry to show that there is a generalized arithmetic progression $G' = \{\sum_{i=1}^{k} x_i d_i' : |x_i| \leq L_i/4\}$ that shares the additive properties of G and is contained in an interval of length $O(|G|)$.

After we prove the Condensing Lemma, we provide some applications. Let $A = \{a_1 < a_2 < \ldots < a_n\}$ be a finite subset of the integers, and denote the *indexed energy*

of A as

$$EI(A) := \{(i,j,k,l) : a_i + a_j = a_k + a_l \text{ and } i + j = k + l\}.$$

The reader may be more familiar with the additive energy of a set which can be used to control the size of the sumset:

$$E(A) = |\{(i,j,k,l) : a_i + a_j = a_k + a_l\}| \geq \frac{|A|^4}{|A+A|}.$$

We determine the precise relationship between $E(A)$ and $EI(A)$. Although the indexed energy of a set has not been directly studied, the additive properties of a set and how they interact with the related indices has appeared in various forms. Solymosi [27] studied the situation when $a_i + a_j \neq a_k + a_l$ for $i - j = k - l = c$ for a fixed constant c, and in particular when a set A has the property that $a_{i+1} + a_i \neq a_{j+1} + a_j$ for all pairs i, j. Brown et al [5] asked if one finitely colors the integers $\{1, \ldots, n\}$, must one be forced to find a monochromatic 'double' 3-term arithmetic progression $a_i + a_j = 2a_k$ where $i + j = 2k$?

Layout and Notation. In section 2, we state some basic notions from convex geometry, and then we prove the Condensing Lemma. In section 3, we study the indexed energy of a set, providing both an extremal construction of a set with large additive energy and small indexed energy as well as proving a Balog-Szemerédi-Gowers type theorem to find a subset with large indexed energy. Section 4 contains further applications and conjectures related to the Condensing Lemma as well as the indexed energy.

We write $[a,b]$ for $[a,b] \cap \mathbb{Z}$, and similarly for $[a,b), (a,b)$, and $(a,b]$. For two functions f, g, we write $f \gg g$ if $f(n) \geq cg(n)$ for some constant c and n sufficiently large. We write $f \gg_K g$ if c is allowed to depend on K. The doubling constant of a set A is $\frac{|A+A|}{|A|}$. A set has *small doubling* if its doubling constant is $O(1)$. A *generalized arithmetic progression* G is a set $\{a + x_1 d_1 + \ldots x_k d_k : |x_i| \leq L_i\}$; we call k the dimension of G; $|G|$ is the volume of G. Moreover, G is proper if the volume of

G is maximal – $(2L_i + 1)^k$.

3.1 Condensing Lemma

The following lemma in conjunction with Theorem 48 will allow us to prove Theorem 47.

Lemma 50. *Let G and G' be proper generalized arithmetic progressions of the form*

$$G := \{\sum_{i=1}^{k} a_i d_i : |a_i| \leq L_i\} \text{ and } G' := \{\sum_{i=1}^{k} a_i d_i : |a_i| \leq 4L_i\}$$

where $a_i, d_i \in \mathbb{Z}$. Then, there exists a constant $c = c(k)$, $d'_1, \ldots, d'_k \in \mathbb{Z}$, and a map ϕ with the following properties:

1. *$\phi(\sum_{i=1}^{k} a_i d_i) = \sum_{i=1}^{k} a_i d'_i$ for $|a_i| \leq L_i$.*
2. *ϕ is an order-preserving Freiman 2-isomorphism.*
3. *For any $x \in G$, $|\phi(x)| \leq c|G|$.*

In order to prove this lemma we need some definitions and results from convex geometry, from which we refer the reader to [3] as a reference.

3.1.1 Convex Geometry

A set $K \subset \mathbb{R}^n$ is said to be a convex cone if for all $\alpha, \beta \geq 0$ and $\mathbf{x}, \mathbf{y} \in K$ we have $\alpha\mathbf{x} + \beta\mathbf{y} \in K$.

Fact 51. *The set of solutions to the system of linear inequalities*

$$\sum_{i=1}^{k} a_{i,j} x_i > 0; a_{i,j} \in \mathbb{R} \text{ and } j = 1, \ldots, n \tag{33}$$

is a convex cone.

Proof. Let $\mathbf{x}$ and $\mathbf{y}$ be solutions to the system of linear inequalities defined above and let $\alpha, \beta \geq 0$. It is trivial to verify that $\alpha\mathbf{x}$ and $\mathbf{x} + \mathbf{y}$ are also solutions to (33). □

For points $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ and non-negative real numbers $\alpha_1, ..., \alpha_m$, the point

$$\mathbf{x} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i$$

is called a conic combination of the points $\mathbf{x}_1, ..., \mathbf{x}_m$. The set $co(D)$ is defined as all conic combinations of points in $D \subset \mathbb{R}^n$ and is called the conical hull of the set D. For a non-zero $\mathbf{x} \in \mathbb{R}^n$ the conical hull of $\mathbf{x}$ is called a ray spanned by $\mathbf{x}$. A ray R of the cone K is called an extreme ray if whenever $\alpha\mathbf{x} + \beta\mathbf{y} \in R$ for $\alpha > 0$, $\beta > 0$ and $\mathbf{x}, \mathbf{y} \in K$ then $\mathbf{x}, \mathbf{y} \in R$. An extreme ray is a 1-dimensional face of the cone. A set $B \subset K$ is called a base of K if $0 \notin B$ and for every point $\mathbf{x} \in K$, $\mathbf{x} \neq 0$, there is a unique representation $\mathbf{x} = \lambda\mathbf{y}$ with $\mathbf{y} \in B$ and $\lambda > 0$.

Fact 52. *Let*

$$A := \{(x_1, \ldots, x_k) \in \mathbb{R}^k : \sum_{i=1}^{k} a_{i,j} x_i > 0 \text{ for } j = 1, \ldots, \ell\}$$

be the solution set to a system of linear inequalities in $\mathbb{R}^k$ with a nonempty set of solution space in the positive quadrant of $\mathbb{R}^k$. Then, the closure of A has a compact base.

Proof. Observe that A is an open set, and since there is at least one solution, it is nonempty. By Fact 51, A is also a convex cone. Let $cl(A)$ be the closure of A, and let $H := \{(x_1, \ldots, x_k) \in \mathbb{R}^k : x_1 + \ldots + x_k = 1\}$. We claim that $B := cl(A) \cap H$ is a compact base of $cl(A)$. Clearly B is a subset of $cl(A) - \{0\}$. Let $\mathbf{y} \in cl(A)$ and consider the line $\lambda\mathbf{y}$. Since A is a convex cone, this line is contained in $cl(A)$ for all $\lambda \geq 0$. If this line intersects B, then B must be a compact base, but clearly it does at $\lambda = \frac{1}{y_1 + \ldots + y_k}$. □

Theorem 53 (Cor. 8.5 [3])**.** *If K is a convex cone with a compact base. Then every point $\boldsymbol{x} \in K$ can be written as a conic combination*

$$\boldsymbol{x} = \sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i, \quad \lambda_i \geq 0, \quad i = 1, ..., m,$$

where the $\boldsymbol{x}_i$ each span an extreme ray of K.

Lastly, we need the well-known linear algebraic result known as Cramer's rule.

Theorem 54 (Cramer's Rule)**.** *Let A be a $k \times k$ matrix over a field F with nonzero determinant. Then, $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution given by*

$$\boldsymbol{x}_i = \frac{det(A_i)}{det(A)} \; i = 1, \ldots, k$$

where A_i is obtained by replacing the ith column in A with $\boldsymbol{b}$.

The broad idea of the proof of Lemma 50 is as follows. We are given a generalized arithmetic progression $G := \{\sum_{i=1}^{k} a_i d_i : -L_i \leq y_i \leq L_i\}$. In a sense, this can be indentified with the point $(d_1, \ldots, d_k)$. What we would like to find is another generalized arithmetic progression, $H := \{\sum_{i=1}^{k} b_i d'_i : -L'_i \leq b_i \leq L'_i\}$ which maintains the same additive structure as G, but is much more compact. Viewed another way, we want to find a point $(d'_1, \ldots, d'_k)$ much closer to the origin than $(d_1, \ldots, d_k)$ that also satisfies certain inequalities (these are what maintain the additive structure). Hence, we reduce our problem to finding an integer solution, relatively close to the origin, to a set of linear inequalities.

3.1.2 Proof of the Condensing Lemma

The crux in the proof of the Condensing Lemma is to first prove it for generalized arithmetic progressions; that is, to first prove Lemma 50.

Proof of Lemma 50. Given G as in the statement of the Lemma, consider the following set of inequalities:

$$\{\sum_{i=1}^{k} a_i x_i > 0 : a_1 d_1 + \ldots + a_k d_k > 0; -4L_i \leq a_i \leq 4L_i\}. \tag{34}$$

We will first prove that if $(d'_1, \ldots, d'_k)$ is an integer solution to the above system of inequalities, then the map $\phi : G \to \mathbb{Z}$ defined by

$$\phi\left(\sum_{i=1}^{k} a_i d_i\right) = \sum_{i=1}^{k} a_i d'_i$$

is an order-preserving Freiman 2-isomorphism.

To see that ϕ is order-preserving, if

$$\sum_{i=1}^{k} a_i d_i < \sum_{i=1}^{k} b_i d_i$$

for two elements in G, then

$$\sum_{i=1}^{k} (b_i - a_i) x_i > 0$$

is one of the inequalities in (34) that $(d'_1, \ldots, d'_k)$ must satisfy; so

$$\phi\left(\sum_{i=1}^{k} a_i d_i\right) = \sum_{i=1}^{k} a_i d'_i < \sum_{i=1}^{k} b_i d'_i = \phi\left(\sum_{i=1}^{k} b_i d_i\right).$$

For the converse, if

$$\sum_{i=1}^{k} a_i d'_i < \sum_{i=1}^{k} b_i d'_i \tag{35}$$

and

$$\sum_{i=1}^{k} (b_i - a_i) d_i \leq 0,$$

then we get a contradiction as follows. First, if

$$\sum_{i=1}^{k} (b_i - a_i) d_i = 0,$$

then $b_i = a_i$ because G is a proper generalized arithmetic progression. Hence, (35) cannot hold in this case. If

$$\sum_{i=1}^{k} (b_i - a_i) d_i < 0, \text{ then } \sum_{i=1}^{k} (a_i - b_i) d_i > 0$$

which implies that

$$\sum_{i=1}^{k} (a_i - b_i) x_i > 0$$

is an inequality in (34) satisfied by $(d'_1, \ldots, d'_k)$, again contradicting (35).

If we have points in G such that

$$\sum_{i=1}^{k} a_i d_i + \sum_{i=1}^{k} b_i d_i = \sum_{i=1}^{k} s_i d_i + \sum_{i=1}^{k} t_i d_i$$

then

$$\sum_{i=1}^{k}(a_i + b_i)d_i = \sum_{i=1}^{k}(s_i + t_i)d_i. \tag{36}$$

Moreover, $|a_i + b_i|, |s_i + t_i| \leq 2L_i$. Hence, each side of (36) corresponds to an element in G', and by the fact that G' is proper, we must have that $a_i + b_i = s_i + t_i$ for $i = 1, \ldots, k$. This implies that indeed, ϕ is a Freiman 2-homomorphism:

$$\sum_{i=1}^{k} a_i d'_i + \sum_{i=1}^{k} b_i d'_i = \sum_{i=1}^{k} s_i d'_i + \sum_{i=1}^{k} t_i d'_i. \tag{37}$$

For the converse, if (37) holds and (36) does not, then without loss of generality, we may assume

$$\sum_{i=1}^{k}(a_i + b_i - s_i - t_i)d_i > 0.$$

However, $a_i + b_i - s_i - t_i \in [-4L_i, 4L_i]$, and so the inequality

$$\sum_{i=1}^{k}(a_i + b_i - s_i - t_i)x_i > 0$$

is satisfied by $(d'_1, \ldots, d'_k)$ which contradicts (37). This proves ϕ is a Freiman 2-isomorphism.

Now, we bound the image of ϕ. Consider the system of inequalities defined in (34); by Fact 51 the solution space forms a convex cone. Moreover, this interior is nonempty since there is a solution – $(d_1, \ldots, d_k)$. Also, $x_i > 0$ is one of our inequalities for all $i = 1, \ldots, k$ so the solution space is in the positive quadrant of $\mathbb{R}^k$. Let K be the closure of the cone defined by the inequalities in (34). By Fact 52, K has a compact base. So, we may apply Theorem 53 to conclude that each $\mathbf{x} \in K$ can be represented as conic combinations of the points on its extreme rays. Because all extreme rays have dimension 1, they are each intersections of $k - 1$ linearly independent hyperplanes corresponding to the system (34). For each extreme ray, we show how to find an integer point on it; then, taking a conic combination of these integer points will allow us to find an integer point in the interior of the cone.

Let the following hyperplanes define one of our extreme rays:

$$\{a_{i,1}x_1 + \ldots + a_{i,k}x_k = 0 : i = 1, \ldots, k-1\}. \tag{38}$$

This system of equations will have all the points along our extreme ray as a solution. Hence, we may treat one of the variables x_i as a free variable while the other variables depend on it. Without loss of generality, assume that x_k is the free variable, and let us solve the system for the case when $x_k = 1$. We will use Cramer's rule. Let

$$\Delta := \begin{vmatrix} a_{1,1} & \ldots & a_{1,k-1} \\ a_{2,1} & \ldots & a_{2,k-1} \\ \vdots & & \\ a_{k-1,1} & \ldots & a_{k-1,k-1} \end{vmatrix}$$

and let Δ_i be the determinant of the same matrix with the ith row and column replaced by $-a_{j,k}$ for $j = 1, \ldots, k-1$:

$$\Delta_i := \begin{vmatrix} a_{1,1} & \ldots & a_{1,i-1} & -a_{1,k} & a_{1,i+1} & \ldots & a_{1,k-1} \\ \vdots & \ddots & & & & & \\ a_{k-1,1} & \ldots & a_{k-1,i-1} & -a_{k-1,k} & a_{k-1,i+1} & \ldots & a_{k-1,k-1} \end{vmatrix}.$$

By Cramer's rule, the solution to the system is given by $x_i = \frac{\Delta_i}{\Delta}$ for $i = 1, \ldots, k-1$. By instead choosing $x_k = c$ instead of $x_k = 1$, we see that we can require that any multiple of this is also a solution to (38). Hence, $(|\Delta_1|, \ldots, |\Delta_{k-1}|, |\Delta|)$ is an integer solution to our system that lies along our edge. For convenience, let $\Delta_k := \Delta$.

Now, we may get such an integer solution for each of our extreme rays. Because cone K has interior points, then not all extreme rays belong to the same face, in particular, we may take a set of $k+1$ of such rays that do not all lie along the same face and get $k+1$ integer solutions as we did above. Call these solutions $\mathbf{p}_1, \ldots, \mathbf{p}_{k+1}$. We can bound the entries of $\mathbf{p}_i$ by using a trivial bound on the determinant of our matrices formed above. We have that for $i = 1, \ldots, k$, since each entry $|a_{i,j}| \leq 4L_j$,

the determinant is bounded as follows:

$$|\Delta_i| \leq 4^k k! \prod_{j \neq i} L_j$$

Moreover, the sum, $\mathbf{p}_1 + \ldots + \mathbf{p}_{k+1} =: (d'_1, \ldots, d'_k)$ does not belong to any of the faces of K; so, it belongs to the interior of the cone, and hence, satisfies (34). Lastly, this implies that the image of ϕ is bounded as follows:

$$\left|\phi\left(\sum_{i=1}^{k} y_i d_i\right)\right| = \left|\sum_{i=1}^{k} y_i d'_i\right| \leq \left|\sum_{i=1}^{k} L_i d'_i\right| \leq \left|\sum_{i=1}^{k} L_i (k+1)(4^k k! \prod_{j \neq i} L_j)\right|$$
$$\leq (k+1)! 4^k \prod_{j=1}^{k} L_j.$$

So if $g \in G'$, $\phi(g) \in [-4^k(k+1)!|G|, 4^k(k+1)!|G|]$. □

The proof of Theorem 47 follows easily from applying Theorem 48 to a set with small doubling. We need the following trivial fact.

Fact 55. *Let ϕ_1 be an order-preserving Freiman isomorphism, and let $\phi_2(x) = x + a$. Then $\phi_2, \phi_1 \circ \phi_2$ and $\phi_2 \circ \phi_1$ are order-preserving Freiman isomorphisms.*

Proof of Theorem 47. Let $A \subseteq \mathbb{Z}$ be such that $|A + A| \leq K|A|$. All constants c_i in the following depend only on K. We may apply Theorem 48 to A to get a generalized arithmetic progression $G \subseteq 2A - 2A$ with $|G| \geq c_1|A|$, dimension at most c_2, and for each $x \in G$, there are at least $c_3|A|^3$ quadruples $(a, b, c, d) \in A^4$ with $x = a+b-(c+d)$. Hence,

$$|\{(a, b, c, d) \in A^4 : a + b - (c + d) \in G\}| \geq c_3|A|^3|G|.$$

So, we can find a triple (b, c, d) such that

$$|\{a \in A : a + b - (c + d) \in G\}| \geq c_3|G|.$$

Let $A' := \{a \in A : a + b - c - d \in G\}$. Let $G' = G - b + c + d$. So, $A' \subseteq G'$, $|A'| \geq c_3|G'|$, and G' is a proper generalized arithmetic progression of the same size and dimension as G.

Denote G' as

$$G' = \{u + \sum_{i=1}^{k} x_i d_i : |x_i| \leq L_i\}.$$

By Fact 55, we may assume $u = 0$, else simply shift everything in A' and G' by $-u$, and work with those sets instead. Let

$$G'' := \left\{ \sum_{i=1}^{k} x_i d_i : |x_i| \leq \lfloor L_i/4 \rfloor \right\}.$$

Apply Lemma 50 to G'' to get an order-preserving Freiman isomorphism $\phi : G'' \to [-c_4|G''|, c_4|G''|]$. We have that $A' \subseteq G'$, but $A' \cap G''$ may not be large. However, by considering the 4^k different translates, $G'' + v$, where $v = j\lfloor L_i/4 \rfloor$ for $j = 0, 1, 2, 3$, $i = 1, \ldots, k$ there exists an integer v such that

$$|A' \cap (G'' + v)| = |(A' - v) \cap G''| \gg_k |A'|.$$

Let $A'' := A' \cap (G'' + v)$. So, ϕ is an order-preserving Freiman isomorphism from $A'' - v$ to $[-c_4|G''|, c_4|G''|]$, and by Fact 55, $\phi_0(x) := \phi(x) - v$ is an order-preserving Freiman isomorphism from A'' to $[-c_4|G''|, c_4|G''|]$. By Theorem 49, since $G \subseteq 2A - 2A$ and $|A + A| \leq K|A|$, we must have $|G| \ll_K |A|$, and so $[-c_4|G''|, c_4|G''|] = [-c_5|A''|, c_5|A''|]$, proving the lemma. □

3.2 Indexed Energy

One always has the following relationship between the additive energy and indexed energy:

$$|A|^2 \leq EI(A) \leq E(A) \leq |A|^3.$$

If A is an arithmetic progression the relationship is strengthened to $EI(A) = E(A)$. Moreover, for an arithmetic progression A, $E(A)$ is maximized. Thus, it is natural to wonder if one loosens the restriction to $E(A) \gg |A|^3$ then is $EI(A) \gg |A|^3$? We provide a counterexample to show that this is false.

Theorem 56. *There exists an integer N such that for every $n \geq N$, there exists $A \subset [n]$ such that, $E(A) \geq \frac{1}{18}|A|^3$ and $EI(A) \leq 2000|A|^2(\log|A|)^2$.*

Thus, one can indeed have the additive energy $\Omega(|A|^3)$ while the indexed energy is $O((|A|\log|A|)^2)$. However, when the additive energy is large, it turns out that one can still pass to a large subset $A' \subseteq A$, $|A'| = \Omega(|A|)$, which has indexed energy $\Omega(|A'|^3)$. We note that when passing to a subset, the subset does not inherit the same indices as the superset, but rather it is reindexed in the natural way. Hence, $EI(A')$ is not bounded from above or below by $EI(A)$.

Theorem 57. *For any $K > 0$, there exists c_1, c_2 dependent only on K such that if A is a finite set of integers with $|A+A| \leq K|A|$ then the following holds. There exists an $A' \subseteq A$ such that $EI(A') \geq c_1|A'|^3$ and $|A'| \geq c_2|A|$.*

We mention in passing that the condition that $|A+A| \ll_K |A|$ may easily be loosened to $E(A) \gg_K |A|^3$ by applying the following well-known result of Balog-Szemerédi [2] and Gowers [19] to pass to a subset with small doubling.

Theorem 58 (Balog-Szemerédi[2], Gowers[19])**.** *For any $K > 0$, there exists c_1, c_2 such that if $A \subseteq \mathbb{Z}$ is such that $E(A) \geq K|A|^3$ then there exists $A' \subseteq A$ with $|A'| \geq c_1|A|$ and $|A'+A'| \leq c_2|A'|$.*

3.2.1 Indexed energy in subsets of $[1,n]$

It turns out that if A is a dense subset of an interval, then there is a simple algorithm that can find a subset $A' \subseteq A$ with $|A'| \gg |A|$ and $EI(A') \gg |A'|^3$. Thus, the general case may then be quickly deduced by applying the Condensing Lemma. We first begin with a lemma that states, loosely speaking, that if A is a dense subset of $[1,n]$, then one can choose a large subset $A' \subseteq A$ that is equidistributed over the interval.

Lemma 59. *For every $\delta > 0$, there exists c_1, c_2, c_3, N such that if $A \subseteq [1,n]$ with $n > N$ and $|A| = \delta n$, then the following holds. There exists an $A' \subseteq A$, $|A'| \geq c_1|A|$ and for $c_3|A|^2$ pairs of integers $0 \leq i, j < n/c_2$, we have that*

$$|A' \cap [ic_2, jc_2)| = j - i. \tag{39}$$

It is easy to establish that a set with property (39) has large indexed energy.

Lemma 60. *For every $\delta > 0$, there exists c_0, c_1, N such that if $A \subseteq [1, n]$ with $n > N$ sufficiently large and $|A| = \delta n$, then A has a subset $A' \subseteq A$ with $|A'| \geq c_1|A|$ and $EI(A') \geq c_0|A|^3$.*

Proof of Lemma 59. It suffices to prove that there exists an $A' \subseteq A$ and c_1, c_2, c_3 dependent on δ such that the following holds: $|A'| \geq c_1|A|$, for $c_3|A|$ integers $0 \leq i < n/c_2$,

$$|A' \cap [0, ic_2)| = i.$$

Once this statement is established, then for any pair of integers i, j satisfying the above, we have $|A' \cap [ic_2, jc_2)| = j - i$. This would prove the statement of the lemma.

Denote $A = \{a_1 < a_2 < \ldots < a_{\delta n}\}$. Let $d = \lfloor \frac{2}{\delta} \rfloor$. We may assume $d|n$, if not, replace n with $n \leq n' \leq 2n$ where $d|n'$. Such an n' exists if n is sufficiently large, and the proof will proceed in the same manner with only a slight modification in our constants c_1, c_2, c_3. Let $I_j = [(j-1)d, jd)$ for all $j = 1, \ldots, \frac{n}{d}$. Let $A_j = A \cap I_j$. We pick our subset A' as follows:

- *Step 1:* If $A_1 \neq \emptyset$ then let $X_1 = \{a_1\}$. Else, $X_1 = \emptyset$.
- *Step k:* If $|A_k \cup X_{k-1}| \leq k$, then $X_k := A_k \cup X_{k-1}$. Else, arbitrarily choose $Y \subseteq A_k$ so that $|Y \cup X_{k-1}| = k$ and then let $X_k := Y \cup X_{k-1}$.

It is clear this algorithm ends after $\frac{n}{d}$ steps. Let $A' = X_{\frac{n}{d}}$.

To prove that A' satisfies the conclusion of the lemma, we analyze the algorithm as follows. First, note that $X_1 \subseteq X_2 \subseteq \ldots \subseteq X_{\frac{n}{d}} = A'$ and $|X_i| \leq i$ for all i. Now, the sets X_i for which $|X_i| = i$ we will call good, and the others we will call bad. Note that if X_i is good, then $|A' \cap [0, id)| = i$; hence, showing that lots of X_i are good will prove the lemma. Let $J = \{j_1, j_2, \ldots, j_k\}$ be the set of indices such that X_{j_i} is good. Observe that for indices between j_i and j_{i+1}, we must not have enough elements to

make any of those corresponding sets good. More precisely,

$$|A_{j_i+k}| \leq k - 1 - \sum_{s=1}^{k-1} |A_{j_i+s}|.$$

This implies that

$$\left| \bigcup_{k=1}^{j_{i+1}-j_i-1} A_{j_i+k} \right| \leq j_{i+1} - j_i - 2.$$

So, we must have that

$$\left| \bigcup_{i=1}^{k-1} \bigcup_{s=1}^{j_{i+1}-j_i-1} A_{j_i+s} \right| \leq \sum_{i=1}^{k-1} j_{i+1} - j_i - 2 = j_k - j_1 - 2(k-1) \leq j_k \leq \frac{n}{d}$$

Thus, we have that $\delta n - \frac{n}{d} \geq \delta n/2$ elements of A are distributed over good intervals. Since each interval is of length d, then we must have that k, the number of good intervals, is at least

$$\frac{\delta n}{2d} \geq \frac{n\delta^2}{4}.$$

This in turn gives us a lower bound on $|A'| = j_k \geq k \geq \frac{n\delta^2}{4} = \frac{\delta}{4}|A|$. □

Proof of Lemma 60. Apply Lemma 59 to A to get A', c_1, c_2, c_3 as in the lemma. Let $A' = \{b_1 < b_2 < \ldots < b_m\}$. Let $J = \{j : |A' \cap [0, c_2 j)| = j\}$. We know that $|J| \geq c_2|A|$. Now, let $A'' = \{b_j : j \in J\}$. Since $EI(A') \geq |\{(i,j,k,l) \in J^4 : b_i + b_j = b_k + b_l \text{ and } i + j = k + l\}|$, we will simply work with these quadruples from A''. However, our final set will still be A' since we need to keep the indices of elements the same as they were in A'.

For all of the following, b_j will be assumed to be from A''. Let $t \in \{2, \ldots, 2m\}$. For $t \leq m$, there are $t-1$ pairs $(i,j) \in [1,m] \times [1,m]$ such that $i + j = t$. For $t > m$, there are $2m - (t-1)$ pairs $(i,j) \in [1,m] \times [1,m]$ such that $i + j = t$. Let α_t be defined so that for $t \in \{2, \ldots, 2m\}$ there are $\alpha_t(t-1)$ pairs $(i,j) \in J \times J$ with $i + j = t$ and there are $\alpha_t(2m - (t-1))$ such pairs for $t \in \{m+1, \ldots, 2m\}$. Observe that for such pairs $(i,j) \in J \times J$, we have $b_i + b_j \in [(t-2)d, td)$. Thus, there are only $2d$ values that $b_i + b_j$ can take. For every $i \in [0, 2d-1]$, let t_i denote the number of

pairs $(i,j) \in J \times J$ with $b_i + b_j = (t-2)d + i$. We can bound the indexed energy of A' as follows:

$$EI(A') \geq \sum_t \sum_{i=0}^{2d-1} t_i^2 = \sum_{t=2}^{m} \sum_{i=0}^{2d-1} t_i^2 + \sum_{t=m+1}^{2m} \sum_i t_i^2$$

Using Cauchy-Schwarz, one has

$$\geq \frac{1}{2d} \left(\sum_{t=2}^{m} (\alpha_t(t-1))^2 + \sum_{t=m+1}^{2m} (\alpha_t(2m-t+1))^2 \right)$$

Using Cauchy-Shwarz again,

$$\geq \frac{1}{2d}\frac{1}{m} \left(\left(\sum_{t=2}^{m} \alpha_t(t-1) \right)^2 + \left(\sum_{t=m+1}^{2m} \alpha_t(2m-t+1) \right)^2 \right)$$

Since

$$\sum_{t=2}^{m} \alpha_t(t-1) + \sum_{t=m+1}^{2m} \alpha_t(2m-t+1) = |J|^2$$

one of the sums must be at least $|J|^2/2$. Hence, we have that

$$EI(A') \geq \frac{|J|^4}{2md} = c_0|A|^3$$

for some constant c_0 depending only on δ. □

Now, we are ready to prove Theorem Theorem 57.

Proof of Theorem 57. Let A be a finite subset of integers with $|A+A| \leq c|A|$. All constants c_i in the following depend only on c. Apply Theorem 47 to A to get a set $A' \subseteq A$ with $|A'| \geq c_1|A|$ and an order-preserving Freiman $\phi : A' \to [-c_2|A'|, c_2|A'|]$. We may assume at least one third of the elements are in $[1, c_2|A'|]$ or simply shift A' by $v = c_2|A'|$. Apply Lemma 60 to $\phi(A')$ to conclude that $EI(\phi(A')) \geq c_3|\phi(A')|^3 = c_3|A'|^3$. It is easy to see that $EI(\phi(A')) = EI(A')$ since ϕ is an order-preserving Freiman 2-isomorphism, so the result follows. □

3.2.2 An Extremal Construction

The proof of Theorem 56 follows from the following lemma.

Lemma 61. *Let $n \in \mathbb{N}$, and let $p \in (1,2)$ and denote $p = 1+\epsilon$. Let $A = \{\lfloor a^p \rfloor : 1 \leq a \leq \lfloor n^{1/p} \rfloor\}$. Then, $EI(A) \leq 16\epsilon^{-1} n^2 \log n$.*

Proof of Lemma 61. Let $x, y \in [1, \lfloor n^{1/p} \rfloor]$ with $x+1 < y$. The main part of the argument is to establish the following bound:

$$x^p + y^p - (x+1)^p - (y-1)^p > \frac{\epsilon(y-x)}{2y} \tag{40}$$

For now, assume (40) holds. If $x+y = z+w$, then by convexity, $x^p + y^p \neq z^p + w^p$ unless $z = x$ and $y = w$ or vice versa. However, it may happen that $x+y = z+w$ and $\lfloor x^p \rfloor + \lfloor y^p \rfloor = \lfloor z^p \rfloor + \lfloor w^p \rfloor$. Since $\lfloor a^p \rfloor = a^p - [a^p]$, where $[a^p]$ is the noninteger part of a^p, we must have that if $x+y = z+w$ and

$$\lfloor x^p \rfloor + \lfloor y^p \rfloor = \lfloor z^p \rfloor + \lfloor w^p \rfloor$$

then

$$|x^p + y^p - z^p - w^p| < 2.$$

So, fixing an x and a y, we can bound how many other pairs z and w can have $z+w = x+y$ and $\lfloor z^p \rfloor + \lfloor w^p \rfloor = \lfloor x^p \rfloor + \lfloor y^p \rfloor$. More specifically, we find the largest t such that

$$x^p + y^p - (x+t)^p - (y-t)^p < 2.$$

Using (40), the triangle inequality, and letting $k = y - x$ we get that

$$x^p + y^p - (x+t)^p - (y-t)^p \geq \frac{\epsilon k}{2y} + \frac{\epsilon(k+2)}{2(y-1)} + \ldots + \frac{\epsilon(k+2(t-1))}{2(y-(t-1))}$$

Each term in the sum is greater than or equal to $\frac{\epsilon k}{2y}$, so we get a lower bound of $\frac{t\epsilon k}{2y}$. So, if $t \geq \frac{4y}{\epsilon(y-x)}$, then we cannot have

$$\lfloor x^p \rfloor + \lfloor y^p \rfloor = \lfloor (x+t)^p \rfloor + \lfloor (y-t)^p \rfloor.$$

This allows us to conclude that any quadruple (x, y, z, w) with $x + y = z + w$, with $x < z < w < y$, $z < x < y < w$, $w < y < x < z$, or $y < w < z < x$ we must have that $|z - x| < \frac{4y}{\epsilon(y-x)}$. Accounting for an extra factor of 2 for when $x < w < z < y$ and so on, we can bound the indexed energy of A

$$EI(A) \leq 2 \sum_{y} \sum_{x<y} \frac{4y}{\epsilon(y-x)}$$

Estimating this summation by using the harmonic series gets us that

$$EI(A) \leq \frac{16}{\epsilon} n^2 \log n$$

concluding the proof assuming that (40) holds.

Now, we work to establish (40). First, since $f(x) = x^p$ is convex for $p > 1$, it is easy to establish the following bound for any $k > 1$:

$$p(x+k)^{p-1} > (x+1)^p - x^p > px^{p-1}$$

Assuming $p = 1 + \epsilon < 2$, we have that x^{p-1} is concave. Doing a similar analysis for $g(x) = x^{p-1}$, we get that

$$(p-1)x^{p-2} > (x+k)^{p-1} - x^{p-1} > (p-1)(x+k)^{p-2},$$

Let $k = y - x$, and we have

$$x^p + y^p - (x+1)^p - (y-1)^p =$$

$$= y^p - (y-1)^p - ((x+1)^p - x^p) > p(y-1)^{p-1} - p(x+1)^{p-1}$$

Since $x = y - k$, we have

$$p[(y-1)^{p-1} - (y-k+1)^{p-1} > p[(k-2)(p-1)(y-1)^{p-2}] > \frac{\epsilon k}{2y}$$

where we remind the reader $p = 1 + \epsilon$, $\epsilon \in (0, 1)$. □

Theorem 56 follows by letting $\epsilon = \frac{1}{\log n}$.

Proof of Theorem 56. Let A be as in the above lemma, let $\epsilon = \frac{1}{\log n}$. Then, for n sufficiently large

$$|A| = \lfloor n^{\frac{1}{1+\epsilon}} \rfloor = \left\lfloor n^{\frac{1}{1+\frac{1}{\log n}}} \right\rfloor = \left\lfloor \frac{n}{e} \cdot n^{\frac{1}{1+\log n}} \right\rfloor \geq \left\lfloor \frac{n}{e^2} \right\rfloor \geq \frac{n}{9}.$$

So, $A \subseteq [1, n]$, $|A| = \frac{n}{9}$, and $A + A \subseteq [1, 2n]$. Thus, $|A + A| \leq 2n \leq 18|A|$. Hence,

$$E(A) \geq \frac{|A|^4}{|A + A|} \geq \frac{|A|^3}{18}.$$

By the lemma above, for A sufficiently large,

$$\begin{aligned} EI(A) \leq 16n^2(\log n)^2 \leq 16 \cdot (9|A|)^2(\log 9|A|)^2 &\leq 1296|A|^2(\log 9|A|)^2 \\ &\leq 2000|A|^2(\log |A|)^2. \end{aligned}$$

□

3.3 Further Applications and Conjectures

Since $|(A \times B) + (A \times B)| = |A + A||B + B|$, it is obvious that if $|A + A| \leq K|A|$ and $|B + B| \leq K|A|$, then for any $C \subseteq A \times B$ of size $\delta|A||B|$, one has $|C + C| \ll_K |C|$. However, if $|C| = O(\sqrt{|A||B|})$, one has little control of $|C + C|$. Does there exist a $C \subseteq A \times B$ with $|C| = c\sqrt{|A||B|}$, and $|C + C| \ll_K |C|$? Clearly one could simply take $C = \{(a, b) : a \in A\}$ for a fixed $b \in B$. If we forbid such sets lying on vertical or horizontal lines by additionally requiring that for any distinct $(x, y), (z, w) \in C$ we have $(x - z)(y - w) > 0$, the answer is not as obvious.

For a set $C \subseteq A_1 \times \ldots \times A_k$, call C a **diagonal set** if for any distinct pairs of elements $(x_1, \ldots, x_k), (y_1, \ldots, y_k) \in C$, one has $x_i - y_i > 0$ for all i or $x_i - y_i < 0$ for all i.

Theorem 62. *For any $k, K \in \mathbb{N}$, there exists c_1, c_2 such that the following holds. Let $A_1, \ldots, A_k \subseteq \mathbb{Z}$ be sufficiently large sets of size n such that $|A_i + A_i| \leq K|A_i|$ for all $i = 1, \ldots, k$. Then, there exists a diagonal set $C \subset A_1 \times \ldots \times A_k$ such that $|C + C| \leq c_1|C|$ and $|C| = c_2(|A_1| \ldots |A_k|)^{1/k}$.*

Proof. We may apply the Condensing Lemma to each A_i individually to find constants $c_{1,i}, c_{2,i}$ depending on K such that there exists a subset $A_i' \subseteq A_i$ that is Freiman isomorphic to a set $B_i \subseteq [0, c_{1,k}n]$, and $|A_i'| \geq c_{2,i}n$. Let c_1 be the maximum of $\{c_{1,i} : i = 1, \ldots, k\}$ and let c_2 be the minimum of $\{c_{2,i} : i = 1, \ldots, k\}$. So, we may view all the B_i as being dense in the interval $[0, c_1 n]$. Next, we claim that there exists $t_1, \ldots, t_k \in \mathbb{Z}$ such that

$$\left| \bigcap_{i=1}^{k} (B_i + t_i) \right| \geq \frac{c_2^k}{2^{k-1}} n.$$

We prove this by induction on k. For $k = 1$, it is trivial. For the induction step, let $X, Y \subset [1, n]$ be of size $\delta_1 n$ and $\delta_2 n$ respectively. Then,

$$\sum_{t=-(n-1)}^{n-1} |X + t \cap Y| = |X||Y| = \delta_1 \delta_2 n^2.$$

Hence, there exists a t such that

$$|(X + t) \cap Y| \geq \frac{\delta_1 \delta_2}{2} n.$$

Letting $X := B_k$ and $Y := \cap_{i=1}^{k-1} B_i + t_i$ finishes the inductive argument. Now, let $C' = \cap_{i=1}^{k} B_i + t_i$ for such a set of t_i, $i = 1, \ldots, k$. Denote $C' := \{x_1 < \ldots < x_m\}$. We let C be the following set:

$$C := \{(x_i - t_1, x_i - t_2, \ldots, x_i - t_m) : i = 1, \ldots, m\}.$$

Since $x_i - t_j \in B_j$, we have that $C \subseteq B_1 \times \ldots \times B_k$. Since $x_i - t_j > x_\ell - t_j$ for $i > \ell$, C must be diagonal. Also, $|C| = |C'| \in [\frac{c_2^k}{2^{k-1}} n, n]$. Lastly, it is easy to see that

$$|C + C| = |C' + C'| \leq 2n = \frac{2^k}{c_2^k} |C'|.$$

□

Although the above application is similar in spirit to the indexed energy problem – letting $A \times B := A \times [1, |A|]$ – there are several subtle differences. Mainly, in the indexed energy problem, when we pass to a subset, we are forced to reindex the set

in a very specific way. The following conjecture however would be general enough to imply Theorem 57.

Conjecture 63. *Let $A, B \subseteq \mathbb{Z}$ be sets of size N such that $|A+A|, |B+B| \leq KN$. Then, there exists c_1, c_2 depending only on K such that the following holds. There exists an $A' \subseteq A$ with $|A'| \geq c_1|A|$, and if we denote $A' := \{a'_1 < \ldots < a'_k\}$ and $B := \{b_1 < \ldots < b_n\}$, then*

$$|\{(a'_i, a'_j, a'_k, a'_\ell) : a'_i + a'_j = a'_k + a'_\ell \text{ and } b_i + b_j = b_k + b_\ell\}| \geq c_2|A'|^3.$$

Conjecture 63 is true in the case where $B = [1, N]$ (or any arithmetic progression of size N) since this then becomes the indexed energy result. It would be interesting to know whether the conjecture is even true in the case where B is a generalized arithmetic progression of dimension 2.

Another problem closely related to the indexed energy problem is as follows. Let $A \subseteq \mathbb{Z}$ and let $f : A \to \mathbb{Z}$ be such that $|f(A) + f(A)| \leq c|A|$, and $|A + A| \leq c|A|$. Let

$$E_f(A) := \{(a, b, c, d) : a + b = c + d, f(a) + f(b) = f(c) + f(d)\}.$$

When f is the indexing function, $E_f(A)$ becomes $EI(A)$. What is the relation between $E_f(A)$ and $E(A)$? Does there always exist an $A' \subseteq A$ with $|A'| \gg |A|$, and $E_f(A') \gg |A|^3$? Here, we point out to the reader a subtle but important difference between this problem and the indexed energy problem: when passing to a subset, there is a natural way to reindex a set which is distinctly different than how a function restricted to a subset behaves. Therefore, $E_f(A)$ is not a generalization of $EI(A)$, but instead, it is a different quantity altogether. There is not always an $A' \subseteq A$ with $E_f(A') \gg_K |A|^3$ when $E(A) \geq K|A|^3$. For instance, let f be the indexing function, let A be as in Theorem 56, and since sets are not reindexed

$$E_f(A') \leq EI(A) \ll_K |A|^2 \log |A|.$$

Moreover, $|\{(a+a', f(a)+f(a')) : a, a' \in A\}| \gg |A|^2/\log|A|$. As an openended question, we ask if there are any reasonable conditions that we can impose on f or A to arrive at a different conclusion?

Lastly, we remark that the content of Lemma 59 is making a statement about equidistribution of a set in an interval. This has been a well-studied topic in discrepancy theory; however, we are not aware of it appearing in this specific, combinatorial form – where one is allowed to pass to a subset of the original set, and one only requires that for lots of interval, the subset is well-distributed. We tepidly conjecture a generalization of Lemma 59 to higher dimensions, but it would also be interesting if a counterexample was found.

Conjecture 64. *Let $A \subseteq [1,n] \times [1,n]$ be of size $|A| = \delta n^2$. There exists constants c_1, c_2, c_3 depending only on δ such that the following holds. There exists an $A' \subseteq A$ such that $|A'| \geq c_1|A|$ and for $c_2 n^2$ pairs $0 \leq i, j \leq n/c_3$, $|A' \cap [0, ic_3) \times [0, jc_3)| = ij$.*

REFERENCES

[1] Amirkhanyan, G., Bush, A., Croot, E., and Pryby, C., "Sets of rich lines in general position," *Preprint on arXiv:1310.6707*, 2013.

[2] Balog, A. and Szemerédi, E., "A statistical theorem of set addition," *Combinatorica*, vol. 14, no. 3, pp. 263–268, 1994.

[3] Barvinok, A., *A Course in Convexity*, vol. 54. American Mathematical Soc., 2002.

[4] Bourgain, J. and Chang, M.-C., "On multiple sum and product sets of finite sets of integers," *Comptes Rendus Mathematique*, vol. 337, no. 8, pp. 499–503, 2003.

[5] Brown, T., Jungiċ, V., and Poelstra, A., "On double 3-term arithmetic progressions," *INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY*, vol. 14, p. A43, 2014.

[6] Chang, M.-C., "The Erdős-Szemerédi problem on sum set and product set," *Annals of mathematics*, pp. 939–957, 2003.

[7] Chang, M.-C., "Sum and product of different sets," *Contributions to Discrete Mathematics*, vol. 1, no. 1, 2006.

[8] Chen, Y.-G., "On sums and products of integers," *Proceedings of the American Mathematical Society*, vol. 127, no. 7, pp. 1927–1933, 1999.

[9] Croot, E. and Hart, D., "h-fold sums from a set with few products," *SIAM Journal of Discrete Mathematics*, vol. 24, pp. 505–519, 2010.

[10] Croot, E., Ruzsa, I., and Schoen, T., "Arithmetic progressions in sparse sumsets," *INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY*, vol. 7, no. 2, p. A10, 2007.

[11] Elekes, G., "On the number of sums and products," *Acta Arithmetica*, vol. 81, pp. 365–367, 1997.

[12] Elekes, G., "Sums Versus Products in Number Theory, Algebra and Erdős Geometry," in *Paul Erdős and His Mathematics, II* (Halasz, G., Lovasz, L., Simonovits, M., and Sós, V. T., eds.), vol. 11 of *Bolyai Society Mathematical Studies*, pp. 241–290, Janos Bolyai Mathematical Society and Springer Science+Business Media, 2002.

[13] Elekes, G. and Ruzsa, I., "Few sums, many products," *Studia Sci. Math. Hungar.*, vol. 40, no. 3, pp. 301–308, 2003.

[14] Erdős, P. and Szemerédi, E., "On sums and products of integers," *Studies in Pure Mathematics*, pp. 213–218, 1983.

[15] Ford, K., "Sums and products from a finite set of real numbers," *Ramanujan Journal*, vol. 2, pp. 59–66, 1998.

[16] Ford, K., "Integers with a divisor in $[y, 2y)$," in *Anatomy of Integers* (de Koninck, J.-M., Granville, A., and Luca, F., eds.), vol. 46 of *CRM Proceedings Lecture Notes*, pp. 65–80, American Mathematical Society, 2008.

[17] Fox, J. and Sudakov, B., "Dependent random choice," *Random Structures & Algorithms*, vol. 38, no. 1-2, pp. 68–99, 2011.

[18] Freĭman, G., "Nachala strukturnoĭ teorii slozheniya mnozhestv. Einführung in die Strukturtheorie der Summenmengen.." Kazan': Kazan. Gosudarstv. Ped. Inst; Elabuzh. Gosudarstv. Ped. Inst., Kazan 140 p. (1966)., 1966.

[19] Gowers, W. T., "A new proof of Szemerédi's theorem," *Geometric and Functional Analysis*, vol. 11, no. 3, pp. 465–588, 2001.

[20] Iosevich, A., Roche-Newton, O., and Rudnev, M., "On an application of the Guth-Katz Theorem," *Math. Res. Letters*, vol. 18, no. 4, pp. 691–697, 2011.

[21] Konyagin, S., "h-fold sums from a set with few products," *Moscow J. Combin. and Number Theory*, vol. 4, no. 3, pp. 14–20, 2014.

[22] Konyagin, S. and Shkredov, I. D., "On sum sets of sets, having small product set," *arXiv preprint arXiv:1503.05771*, 2015.

[23] Li, L., "Multifold sums from a set with few products," *arXiv preprint arXiv:1106.6074*, 2011.

[24] Nathanson, M., "On sums and products of integers," *Proceedings of the American Mathematical Society*, vol. 125, pp. 9–16, 1997.

[25] Ruzsa, I. Z., "An application of graph theory to additive number theory," *Scientia, Ser. A*, vol. 3, pp. 97–109, 1989.

[26] Sanders, T., "On the Bogolyubov-Ruzsa lemma," *Anal. PDE*, vol. 5, no. 3, pp. 627–655, 2012.

[27] Solymosi, J., "On distinct consecutive differences," *arXiv preprint math/0503069*, 2005.

[28] Solymosi, J., "On the number of sums and products," *Bull. of the LMS*, vol. 37, no. 04, pp. 491–494, 2005.

[29] Solymosi, J., "Bounding multiplicative energy by the sumset," *Advances in mathematics*, vol. 222, no. 2, pp. 402–408, 2009.

[30] Szemerédi, E. and Trotter, W. T., "Extremal Problems in Discrete Geometry," *Combinatorica*, vol. 3, pp. 381–392, 1983.

[31] Tao, T. and Vu, V., *Additive Combinatorics*. Cambridge University Press, 2010.

[32] Wooley, T., "Multigrade efficient congruencing and Vinogradov's mean value theorem," *arXiv:1310.8447*, 2013.