VILNIUS UNIVERSITY

PAULIUS ŠARKA

ARITHMETIC PROPERTIES OF SPARSE SETS

Doctoral dissertation

Physical sciences, mathematics (01P)

Vilnius, 2013

Doctoral dissertation was written in 2009–2013 at Vilnius University

Scientific supervisor:

prof. habil. dr. Artūras Dubickas
Vilnius University, Physical sciences, Mathematics – 01P

Scientific adviser:

doc. dr. Paulius Drungilas
Vilnius University, Physical sciences, Mathematics – 01P

VILNIAUS UNIVERSITETAS

PAULIUS ŠARKA

RETŲ AIBIŲ ARITMETINĖS SAVYBĖS

Daktaro disertacija

Fiziniai mokslai, matematika (01P)

Vilnius, 2013

Disertacija rengta 2009–2013 metais Vilniaus universitete.

Mokslinis vadovas:

prof. habil. dr. Artūras Dubickas
Vilniaus universitetas, fiziniai mokslai, matematika – 01P

Mokslinis konsultantas:

doc. dr. Paulius Drungilas
Vilniaus universitetas, fiziniai mokslai, matematika – 01P

Contents

Notation

$\mathbb{N}$	the set of positive integers
$\mathbb{Z}$	the set of integers
$\mathbb{P}$	the set of primes
$\mathbb{C}$	the set of complex numbers
$\mathbb{F}_q$	the finite field of order q
$\lfloor x \rfloor$	the largest integer not larger than x (floor)
$\lceil x \rceil$	the smallest integer not smaller that x (ceiling)
$[n]$	the set of positive integers not larger than n, $[1,n] \cap \mathbb{N}$
A	a set A as well as the indicator function of a set A
$A[n]$	intersection $A \cap [n]$
$f * g$	the sum convolution of two functions, $f * g(x) = \sum_i f(i)g(x-i)$
$f \circ g$	the difference convolution of two functions, $f \circ g(x) = \sum_i f(i)g(x+i)$
$\mathsf{E}(A,B)$	the additive energy of sets A and B
$\mathbb{E}\, X$	the average of a random variable X
$A+B$	the sumset $\{a+b, a \in A, b \in B\}$
kA	the k-fold sumset $\{a_1 + ... + a_k, a_i \in A\}$
k_*A	the k-fold sumset with no repetitions $\{a_1 + ... + a_k, a_i \in A, a_i \neq a_j\}$
$f \ll g$	the same as $f = O(g)$
$f \sim g$	the same as $f \ll g$ and $g \ll f$

1 Introduction

In this work we will study various arithmetic properties of subsets integers. Phenomena that interest us arise from the simplest operation possible – *addition*, which carried out between elements of a set or sets leads to interesting and complicated combinatorics. To give a taste of questions we will investigate, we introduce the main objects of this nature. The first one, a *sumset* of two sets A and B, is the set of all possible pairwise sums:

$$A+B=\{a+b,\ a\in A, b\in B\}.$$

The aspect of a sumset that draws the most attention is its *size*. An estimate $|A+B|\geqslant |A|+|B|-1$ for finite sets of integers is an easy exercise, as well as noting that equality can only occur when A and B are arithmetic progressions of the same difference. Taking a small step back and asking to describe sets for which equality does not hold, but that still have *small* sumset, gets us straight to the heart of additive combinatorics. Description of such sets is a famous Freiman theorem proved by Freiman in 1966 and improved a number of times since then.

Closely connected to a sumset is a notion of *additive energy*, which is the number of incident sums of elements of A and B:

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

Relation between two notions is straightforward in two ways – sets with small sumset have large additive energy and sets with small additive energy have large sumset (we will give more precise meaning to the notions of large and small in Chapter 3). Large sumset or energy, on the other hand, does not imply that the counterpart should be small and connection here is much more subtle: a well known Balog-Szemerédi-Gowers theorem states that sets with large additive energy have large subsets with small sumset.

Lastly, we introduce a notion of *representation function* of sets A and B, from which both their sumset and additive energy can be deduced. Representation

function at a given point x is equal to the number of ways x can be expressed as a sum $a+b,\ a\in A, b\in B$. It can be quickly recognized to be nothing else then a convolution of the indicator functions of sets A and B, so overloading the notation we write it as

$$A*B(x)=\sum_i A(x-i)B(i).$$

Famous *Sidon* sets are defined as sets A having all the sums $a+a', a, a'\in A$ different. Another way to describe them is to say that they are the sets with *extremal* property of having a representation function $A*A$ as small as possible – bounded by 2. Finding as dense as possible Sidon sets is much investigated but still open problem.

In the thesis we will solve several problems that are connected to those mentioned above. We describe them in the next section.

1.1 Aims and problems

We give a short summary of the problems considered in this work.

In Chapter 3 we discuss a relation between several notions of unstructured sets, for example those with large sumset and small additive energy. The main question of the chapter is the following extension of a theorem by Alon and Erdős: given finite sets $A, B\subseteq\mathbb{Z}$ of equal sizes $|A|=|B|=n$ and with fixed additive energy $\mathsf{E}(A,B)=|A||B|+E$, what are the sizes of largest subsets $A'\subseteq A$ and $B'\subseteq B$ with all $|A'||B'|$ sums $a+b$, $a\in A', b\in B'$, being different (we call such subsets A', B' *co-Sidon*)? We will answer this question within the logarithmic accuracy for small and large values of additive energy $E\ll n^2$ and $E\gg n^3$, and will extend them (non-optimally, however) to the full range of values of E.

In Chapter 4 we investigate infinite sets $A\subseteq\mathbb{N}$ with bounded *doubling*, that is sets satisfying $|A[n]+A[n]|\ll|A[n]|$, where $A[n]=A\cap[n]$. Full description of such sets is an open problem raised by Sós. We give a partial answer to this question under additional restrictions and investigate the doubling of sets of asymptotically polynomial growth.

In Chapter 5 we consider a generalization of a question concerning maximal Sidon sets. Call a given set B_h set if all h-fold sums of its elements are different.

We look for the upper bound on the size of a B_h set in a d-dimensional box $[n]^d$, and the upper bound of the density of infinite B_h sets in a d-dimensional space $\mathbb{N}^d$, extending known results of Chen, Jia, Graham and Green.

In Chapter 6 we look for slowly growing infinite sets of positive integers with an interesting property, that the *iterated sumset*, a set of all sums of finite subsets, would not contain a square of an integer. We find an example of exponentially growing set and extend it to not contain all powers of integers. We also discuss a subsequent improvement of this result by Dubickas and Stankevičius.

In Chapter 7 we solve a multivariate Cauchy functional equation for multiplicative functions $f : \mathbb{N} \to \mathbb{C}$, limiting the additivity condition on primes: $f(p_1 + p_2 + \cdots + p_k) = f(p_1) + f(p_2) + \cdots + f(p_k), \;\; p_i \in \mathbb{P}$. We show that such equation has essentially unique non-zero solution: $f(n) = n$ for each $n \in \mathbb{N}$. This result for $k = 2$ has been established earlier by Spiro, and $k = 3$ by Fang.

1.2 Methods

Throughout the thesis we use a variety of methods which are quite common in additive combinatorics.

Construction of sets with no large co-Sidon subsets in Chapter 3 relies on the probabilistic method, which is often used to prove the existence of complicated combinatorial objects. In order to apply it we use Chernoff and Kim-Vu concentration inequalities for sums and polynomials of indicator random variables, and a recently established structural description of independent sets in balanced graphs by Kohayakawa, Lee, Rödl, and Samotij.

Lower bounds for growth of the doubling of polynomial sequences in Chapter 4 are established using an effective version of Freiman's theorem (due to Konyagin), which gives an accurate description of finite sets with small sumset.

Most of the upper bounds for size of B_h sets in Chapter 5 are obtained by elementary combinatorial arguments, however, fine improvements in the case when h is large rely heavily on Fourier analytical methods, developed by Green for the one-dimensional case.

Examples in Chapter 6 for sets whose iterated sumset avoids squares are obtained by using elementary number theoretical properties of natural numbers.

In Chapter 7 we use known bounds for binary, ternary and quinary Goldbach conjecture as well as a lemma of Spiro.

1.3 Actuality and novelty

Most of the results presented in the thesis are original and correspond roughly to six publications in mathematical journals [22, 24–26, 64, 68]. Remaining minority are either well known lemmas, which we prove for completeness, or small unpublished observations.

1.4 Acknowledgements

During the years that led to the completion of this thesis I had an honor to meet, be taught by, and work with many wonderful people. I am very happy to thank them here.

I would like to express my gratitude to my supervisor Artūras Dubickas, who supported and encouraged my wanderings into different places of mathematics and showed me the ways of professional mathematician.

I am highly indebted to Javier Cilleruelo, who gave me many ideas and open problems for research and made sure I was very comfortable when I visited him in Madrid. Problems solved in Chapter 5 were suggested by him.

I am also highly indebted to Tomasz Schoen, who very kindly invited me to visit him in Poznan and spent long hours sharing his knowledge and insight. The results of Chapter 3 were obtained together with him.

I am extremely grateful to Vytautas Narmontas, my school teacher, who introduced me to mathematics and spent all his afternoons helping me crack a daily set of Olympiad problems.

A big thanks and a high five to my coauthors Laurence Rackham, Juanjo Rué, Manuel Silva, and Ana Zumalacárregui. I worked with Laurence in Barcelona, with Juanjo and Ana in Madrid, and with Manuel in Lisbon. Those were very fun and productive times!

A big thanks to my mathematical comrades and friends Aivaras Novikas, Tomas Juškevičius, Matas Šileikis and Gražvydas Šemetulskis, with whom I spent

ages discussing all kinds of mathematics with me and sunk many evenings in a glass of beer.

For all the fun and creative atmosphere I am grateful to my department colleagues Hamletas Markšaitis, Romualdas Kašuba, Ramūnas Garunkštis, Mindaugas Bloznelis, Giedrius Alkauskas, Jonas Jankauskas, Justas Kalpokas, Valentas Kurauskas, Jonas Šiurys, and Albertas Zinevičius. I am especially grateful to Paulius Drungilas, who took great care of me and always had a good advice before I even realized I needed one.

Finally, I wish to thank my beloved parents and wife for their never-ending care and support.

2 Literature review

2.1 Sidon, co-Sidon and B_h sets

Recall that a set A is called *Sidon*, if all pairwise sums of its elements $a+a', a, a' \in A$ are different, discounting trivial equalities such as $a + a' = a' + a$. Two generalizations of this notion that we will investigate in this work are B_h sets, those with different h-fold sums $a_1 + \cdots + a_h, a_i \in A$ and *co-Sidon* pairs of sets, that is sets A and B with different sums $a + b, a \in A, b \in B$.

The central question that motivated research on Sidon and, more generally, B_h sets (formulated by Sidon himself in 1932 [76]) is that of *maximal* set: most numerous Sidon set in the finite interval $[n]$ and the densest Sidon set in the set of positive integers. Simple counting arguments immediately give bounds $|A| \ll n^{1/h}$ for a size of a B_h subset of the interval $[n]$ and $|A[n]| \ll n^{1/h}$ for the asymptotic growth of an infinite B_h subset of positive integers. Improving these bounds or constructing matching examples is far from simple.

The best known upper bound for the maximal Sidon set $A \subseteq [n]$ is

$$|A| \leqslant n^{1/2} + n^{1/4} + 1/2,$$

essentially obtained by Erdős and Turán [30] in 1941 and later refined by Lindström [55] and Cilleruelo [15] (Erdős and Turán only reported bound $n^{1/2} + O(n^{1/4})$, but their method can be seen to give the above bound, and Lindström obtained constant 1).

In the more general setting of maximal B_h subset of $[n]$, Lindström [54] improved the elementary bound for $h = 4$:

$$|A| \leqslant 8^{\frac{1}{4}} N^{\frac{1}{4}} + O\left(N^{\frac{1}{8}}\right),$$

and Jia [45] (see also [35]) generalised his argument for all even $h = 2k$:

$$|A| \leqslant k^{\frac{1}{2k}} (k!)^{\frac{1}{k}} N^{\frac{1}{2k}} + O\left(N^{\frac{1}{4k}}\right).$$

For the case of odd $h = 2k - 1$, the best known upper bound was given by Chen and Graham [10, 35]:

$$|A| \leqslant (k!)^{\frac{2}{2k-1}} N^{\frac{1}{2k-1}} + O\left(N^{\frac{1}{4k-2}}\right).$$

Finally, Green used the techniques of Fourier analysis to further improve main term constants in the B_3 and B_4 cases (getting $(\frac{7}{2})^{1/3}$ and $7^{1/4}$ respectively) and in the case when h is sufficiently large.

In Chapter 5 we will extend the results of Jia, Chen, Graham and Green to the d-dimensional case and discuss a particular construction of dense B_h sets by Ruiz and Trujillo [69]. Here we mention that in the case of dense Sidon sets the first example with the optimal constant in the main term was given by Singer [77]. Another interesting (and exceptionally dense) example of a Sidon set in the two dimensional plane was given by Cilleruelo [15]. He constructed an example of the Sidon set in $[n]^2$ of size $n + \log n \log\log\log n$.

Infinite B_h sets

Infinite B_h sets are more complicated than finite ones. Somewhat unexpectedly one cannot construct a Sidon subset $A \subseteq \mathbb{N}$ satisfying $|A[n]| \gg n^{1/2}$, as was showed by Erdős who proved that any Sidon set $A \subseteq \mathbb{N}$ satisfies

$$\liminf_{n\to\infty} |A[n]| \sqrt{\frac{\log n}{n}} < \infty.$$

This result was generalised for d-dimensional Sidon sequences by Cilleruelo [15], who showed that Sidon set $A \subseteq \mathbb{N}^d$ satisfies

$$\liminf_{n\to\infty} |A[n]| \sqrt{\frac{\log n}{n^d}} < \infty$$

and for one dimensional B_{2k} sets by Chen [9], who showed that B_{2k} set $A \subseteq \mathbb{N}$ satisfies

$$\liminf_{n\to\infty} |A[n]| \sqrt[2k]{\frac{\log n}{n}} < \infty.$$

Interestingly, no results of this type are known for h odd.

Constructing dense Sidon subsets of $\mathbb{N}$ is even more challenging. Using greedy algorithm or probabilistic method one can construct a Sidon set A satisfying

$|A[n]| \geqslant n^{1/3}$, which is far from the upper bound. An improvement was obtained by Ruzsa [71], who constructed a Sidon set $A \subseteq \mathbb{N}$ satisfying

$$|A[n]| \geqslant n^{\sqrt{2}-1+o(1)}.$$

The construction of Ruzsa was cleverly modified by Cilleruelo and Tesoro [18] to work in the case $h = 3, 4$ and by Cilleruelo [13] for all h, resulting in B_h subsets $A \subset \mathbb{N}$ of size

$$|A[n]| \geqslant n^{\sqrt{(h-1)^2+1}-(h-1)+o(1)}.$$

Sidon and co-Sidon subsets of other sets

An interesting generalization of the question about the maximal Sidon set in the interval or d-dimensional box is a question about the maximal Sidon set in any given set. The question formulated without any conditions on the initial set was answered by [49] Komlós, Sulyok, and Szemerédi, who proved that the interval is essentially the worst case. That is, given a set A one can always find a Sidon subset of size $\gg \sqrt{|A|}$, and the equivalent bound holds for B_h sets (or any other sets, avoiding solutions to linear equation). That should come as no surprise, as the interval (or box) is the set with the largest additive structure (it has a small sumset and large additive energy).

Erdős and Alon considered the same question with strong restrictions on the additive structure of the initial set [2, 27, 28], more precisely, requiring $A * A$ to be bounded by a constant. One could expect to find a very large Sidon subset in this case, but that is only partially true, as Erdős gave an example of a set A with representation function bounded by 4, which has the largest Sidon subset of size $\ll |A|^{2/3}$. In Chapter 3 we consider a similar extremal question of finding co-Sidon subsets of two sets with given additive energy.

Finding dense co-Sidon subsets is slightly easier than finding dense Sidon sets. It is not hard to find co-Sidon subsets of an interval attaining the maximal possible size, for example, and it is also possible to find (and actually describe all of them) maximal co-Sidon subsets of positive integers – this was done by Benevides, Hulgan, Lemons, Palmer, Riet, and Wheeler [5]. In Chapter 3 we show that one can find larger co-Sidon subsets that the Sidon one in the above problem of Alon and Erdős only when we allow them to be of different sizes.

Finally, we mention a variation of this problem considered by Lewko and Lewko [53] who proved, among other statements, that there exist a set with large sumset and large difference set (within a constant multiple of maximal one) that does not have a large Sidon and, more strongly, large subsets of bounded representation function.

2.2 Sets with large additive structure

Sets with large additive structure are those with, contrary to Sidon sets, many coinciding pairwise sums of their elements. Two most common properties that signify large additive structure are small sumset $|A+A| \ll |A|$ and large additive energy $\mathsf{E}(A,A) \gg |A|^3$. Investigation of such sets has been very extensive and much is known about them. We will only mention two very well known results which we will use in this work as the focus of this thesis lies somewhat in the compliment of this universe. A lot of material on the subject can be found in the books of Tao and Vu [81], Nathanson [60, 61], and Geroldinger and Ruzsa [33].

The central result of additive combinatorics is Freiman's theorem, stating that any set A with small sumset $|A+A| \leqslant C|A|$ is a subset of generalized arithmetic progression

$$A \subseteq \{b_0 + b_1 z_1 + \cdots + b_d z_d \mid z_i = 0, \ldots, \ell_i - 1 \text{ for } i = 1, \ldots, d\},$$

with constant *dimension* d and small *size* $\ell_1 \cdots \ell_d < s|A|$, with constants d, s only depending on C. Originally proved by Freiman [32], the theorem has been reproved (simplifying the arguments) by Ruzsa [70] (also see [6]) and improved a number of times by Chang [7], Schoen [75], Sanders [72] and Konyagin [50]. We give precise formulation of the theorem and state the bounds on d, s in Chapter 4.

An equivalent question about description of infinite sets of positive integers with small sumset was raised by Sós [19], with the notion of small sumset stated as $|A[n] + A[n]| \ll |A[n]|$. In Chapter 4 we give a partial answer to this question.

Large additive energy is a slightly weaker condition on the set, as it does not imply that the set should have a small sumset. Nevertheless, the famous Balog-Szemerédi-Gowers theorem (proved by Balog and Szemerédi [4] and independently

by Gowers [34]) states that such set should have a large structural part, that is a large subset (a constant proportion of the size of original set) with small sumset.

Precise dependence of the doubling C' of the structural subset A' (satisfying $|A' + A'| \leqslant C|A'|$ on the size of additive energy of initial set A (with $\mathsf{E}(A, A) \geqslant c|A|$) is important in many applications of Balog-Szemerédi-Gower theorem. The best known bound has been obtained by Schoen (personal communication) giving $C \leqslant (1/c)^4$.

2.3 Additive properties of number theoretic sets

Additive properties of sets of number theoretic nature is a classical part of number theory. One of the well known problems is a Goldbach's conjecture, raised in 1742 in the correspondence between Goldbach and Euler. Strong version of Goldbach's conjecture states that every even integer greater than 2 can be expressed as a sum of two primes and is among the most challenging conjectures in the whole mathematics.

Weak version of Goldbach's conjecture states that every odd prime greater than 5 can be expressed as a sum of three primes. Its proof has been recently announced by Helfgott in a series of thee preprints [40–42], culminating a century long effort started by Hardy-Littlewood and Vinogradov.

Among the intermediate achievements is a result of Tao [80], who proved that every odd integer larger than 1 can be expressed as a sum of at most 5 primes. We will rely on this result heavily in the Chapter 7, where we solve multivariate Cauchy equation

$$f(p_1 + \cdots + p_k) = f(p_1) + \cdots + f(p_k)$$

for multiplicative functions with above additivity condition defined on the set of primes. This equation was solved for $k = 2$ by Spiro [78] and for $k = 3$ by Fang [31], in both cases obtaining that such equation has essentially unique multiplicative solution. This equation with additivity condition defined on various other sets of number theoretic nature was solved by Chen and Chen [8], Chung [11], Chung and Phong [12], Phong [63] and De Koninck, Kátai and Phong [20].

Another gem in the additive number theory is Waring's problem which, similarly to Goldbach's conjecture, states that every sufficiently large integer can be expressed as a sum of at most s k^{th} powers of integers. The first proof that for each k such finite s indeed does exist was given by Hilbert in 1909 [43], although that fact that every integer is a sum of four squares was already proved by Lagrange in 1770.

Since the proof of Hilbert, the main focus in Waring's problem was finding the smallest required number of summands s for each power k. The structure of squares or any powers of integers is much simpler than that of primes, so Waring's problem has been largely solved. The value of s is known to the accuracy of ± 1, and is equal to $2^k + \lfloor (3/2)^k \rfloor - 2$ provided that $2^k\{(3/2)^k\} + \lfloor (3/2)^k \rfloor < 2^k$ which is conjectured to hold for all $k \in \mathbb{N}$. For more details and historical perspective we point to a wonderful survey by Vaughan and Wooley [82].

In Chapter 6 we investigate the additive structure of the complement of the squares (or powers) of integers. For a set A (finite or infinite) by S_A denote the set of sums of its finite subsets, so called *iterated sumset*. Departure point for the investigation is a question by Erdős in 1986 [29] about the size of the largest subset A of $[n]$, such that S_A would not contain a square. Erdős and later Cilleruelo [14] gave an example of such set of size $\gg n^{1/3}$. Getting the matching upper bound was much more difficult and took an effort of Alon [1], Lipkin [56], Alon and Freiman [3], Sárközy [73] and finally Nguyen and Vu [62] to reduce it to

$$|A| \leqslant n^{1/3} \log^C n$$

for some constant C. In Chapter 6 we solve the same problem in the infinite setting. The first example of infinite subset of integers $A \subseteq \mathbb{N}$ such that S_A would not contain squares was provided by Luca [58]. We give an example of a slower growing set and discuss a subsequent improvement by Dubickas and Stankevičius [23], who gave an example of a set locally matching the Nguyen and Vu lower bound.

A very similar question concerning the largest subset $A \subseteq \mathbb{N}$ such that $A + A$ would not contain any squares was raised by Erdős and Silverman [38]. In this case a very large such set can be found due to properties of quadratic residues modulo some integer. The best known lower bound for the density $d(A)$ of such

set (density $d(A)$ of the infinite set A is defined as $\lim_{n\to\infty} \frac{|A[n]|}{n}$ if it exists) $d(A) \geqslant 11/32$ was obtained by Massias[1], and the best known upper bound $d(A) \leqslant 2/5$ by Schoen [74], improving the results by Lagarias, Odlyzko and Shearer [51, 52].

[1]The paper of Massias *Sur les suites dont les sommes des termes 2 a 2 ne sent pas des carres* (which is cited as *to be published* in some works) does not seem to exist.

3 Large co-Sidon subsets of sets with a given additive energy

In the introduction we discussed a relation between the sumset $A+B$ of two sets A and B and their additive energy $\mathsf{E}(A,B)$. In this chapter we we will include a third measure of (joint) structure of two sets – the maximal value of representation function $A*B$.

For now we will only concern ourselves with the notion of unstructured sets of equal sizes $|A|=|B|=n$. We give four following descriptions of unstructuredness:

I A, B have large sumset $|A+B| \geqslant cn^2$,

II A, B have small additive energy $\mathsf{E}(A,B) \leqslant Cn^2$,

III A, B have bounded representation function $A*B(x) \leqslant C$,

IV A, B are co-Sidon.

These four conditions imply one another in the order $IV \implies III \implies II \implies I$. First two implications follow immediately from the definitions, while the third one is the lemma 5.9. Reverse implications do not hold. Sets with bounded representation function not necessarily are co-Sidon, and sets with small additive energy can have a representation function with few very large values, for example take A a Sidon set and $B=-A$. Lastly, sets with large sumset can each have large structural part, hence large additive energy.

The question that interests us (and inspired by Balog-Szemerédi-Gowers theorem) is whether these reverse implications do hold in a strong way. That is, maybe whenever A, B satisfy condition N it is possible to find large subsets $A' \subset A, B' \subset B$ that satisfy condition $N+1$? We are unfortunately unable to answer this question in any of three cases, but we will deal with the implication $IV \implies II$ later in the chapter. However, these questions can be answered somewhat easily in the symmetric case $A=B$ which we now consider.

Symmetric case

The problem we stated is considerably easier in the symmetric case, that is when $A = B$ and we look for one subset A' with corresponding properties. We start from the implication $IV \implies III$. The theorem that reverse implication does not hold for large subsets was already mentioned in the introduction and we formulate it here in a succinct form:

Theorem 3.1 (Erdős, Erdős-Alon)**.** *Given $C \geqslant 4$ and a set A of size n, satisfying*

$$A * A(x) < C,$$

for each $x \in \mathbb{Z}$, the largest guaranteed Sidon subset $A' \subseteq A$ is of size

$$|A'| = cn^{2/3}.$$

Sketch of the proof. The upper bound for the size of maximal Sidon set is proved by giving an example:

$$A = \{4^i + 4^I, i \in [n], I \in [n+1, n^2+n]\}.$$

It is very easy to see that this set has representation function bounded by 4, so one only needs to prove that no subset larger than cn^2 is Sidon. The prove is simple and elegant, so we sketch it here.

Let A' be a subset of A and count the number of pairs of elements (x, y) from A' such that x and y have the same second coordinate (power) I. By Cauchy-Schwarz inequality the number of such pairs is at least $\binom{|A'|/n^2}{2}n^2$. If this number is larger than $\binom{n}{2}$, by pigeon hole principle two of such pairs will have the same set of first coordinates (i, i'), thus forming a quadruple $(4^I+4^i, 4^I+4^{i'}, 4^J+4^i, 4^J+4^{i'})$, satisfying forbiden equation $x + y = x' + y'$. Comparing these bounds gives the required inequality.

The lower bound is proved using a simple double counting argument, which we also use for the case of different sets in the proof of Theorem 3.4, so we do not repeat it here. □

We continue with the next implication $III \implies II$. Reverse implication does not hold for large subsets in this case as well, but we do not have asymptotically matching lower and upper bounds in this case.

Theorem 3.2. *Given $C \geqslant \frac{5}{3}$ and a set A of size n, satisfying*

$$\mathsf{E}(A, A) \leqslant Cn^2,$$

the largest guaranteed subset $A' \subseteq A$ with $A' * A' < C'$ is of size

$$cn^{2/3} \leqslant |A'| \leqslant c'n^{3/4}.$$

Proof. For the upper bound take A a union of n arithmetic progressions A_i of length n (for simplicity we will construct a set of size n^2), such that each pair of progressions is co-Sidon (for example $A_i = \{kn^i,\ k = 1, \dots, n\}$). Then additive energy of A will be equal to the sum of non-trivial energies of progressions plus the trivial part (by trivial we mean a part of additive energy coming from equalities $a + b = a + b, a \in A, b \in b$):

$$\mathsf{E}(A, A) = |A|^2 + \sum_i (\mathsf{E}(A_i, A_i) - n^2) \leqslant \frac{5}{3} n^4.$$

Let $A' \subseteq A$ have representation function $A' * A'$ bounded by C. Then for any progression A_i the intersection $A' \cap A_i$ is a so-called $B_2[C]$ set contained in a progression of length n, so its size is less than $c'n^{1/2}$, where c' only depends on C (the exact dependence is unknown (but asymptotically quadratic $c' \approx C^{1/2}$, see [17] and references therein). From this we conclude that $|A'| \leqslant c'n^{3/2} = c'|A'|^{3/4}$, as required.

The lower bound is a special case of Theorem 3.4. □

Finally we prove that the implication $II \implies I$ also does not have an inverse for large subsets.

Theorem 3.3. *Given $c \leqslant \frac{1}{2}$ and a set A of size n, satisfying*

$$|A + A| \geqslant cn^2,$$

the largest guaranteed subset $A' \subseteq A$ with $\mathsf{E}(A', A') \leqslant Cn^2$ is of size

$$|A'| = c'n^{1/2}.$$

Proof. For the upper bound take A a union of arithmetic progressions $[n]$ and $[n, 2n, \dots, n^2]$ (for simplicity we will construct an example of size $2n$). Then the sumset $A+A$ is equal to $[n+1, \dots, n^2+n]$ and is of the stated size. Let $A' \subseteq A$ be a subset of A, and denote a larger of the two sets $A' \cap [n], A' \cap [n, \dots, n^2]$ by $A'_{\frac{1}{2}}$. The sumset $A'_{\frac{1}{2}} + A'_{\frac{1}{2}}$ is supported on a set of size $2n-1$, so by Cauchy-Schwarz inequality (see Lemma 5.9) we have

$$\mathsf{E}(A'_{\frac{1}{2}}, A'_{\frac{1}{2}}) \geqslant |A'_{\frac{1}{2}}|^4/(2n-1).$$

Using this observation we get that if $\mathsf{E}(A', A') \leqslant C|A'|^2$, then $|A'_{\frac{1}{2}}|^4 \leqslant C(2n-1)|A'|^2$, or $|A'| \leqslant c'n^{1/2}$ for a corresponding value of c'.

For the lower bound we note that by Komlós, Sulyok and Szemerédi theorem [49] we can find a Sidon set of size of order $n^{1/2}$ in any set of size n. □

3.1 The problem and results

We now formulate our main question of the chapter and give a partial answer. Note that it is more general than stated in the introduction, as it considers full range of additive energy, which in the case of sets of equal sizes $|A| = |B| = n$ is (see Lemmas 3.17 and 3.18)

$$n^2 \leqslant \mathsf{E}(A, B) \leqslant 2n^3/3 + n/3.$$

Question. Let A, B be sets of integers of equal sizes $|A| = |B| = n$ and fixed additive energy $\mathsf{E}(A, B) = |A||B| + E$. What is the largest pair of co-Sidon subsets $A' \subseteq A, B' \subseteq B$?

The partial answer is given in the following four theorems. The strategy of proofs depend on the size of additive energy, so we considered two cases for lower bound and two cases for upper bound. These bounds match when non-trivial additive energy of A, B is either small ($E \ll n^2$) or very large ($E \gg n^3$).

Theorem 3.4. *Let A, B be any finite sets of integers such that*

$$\mathsf{E}(A, B) - |A||B| = E \neq 0.$$

For all integers k, ℓ satisfying $1 \leqslant k \leqslant |A|/2$, $1 \leqslant \ell \leqslant |B|$ and

$$k\ell^2 \leqslant \frac{|A|^2|B|^2}{2E}, \tag{3.1}$$

there exists a pair of co-Sidon subsets $A' \subseteq A$, $B' \subseteq B$ with $|A'| = k$, $|B'| = \ell$.

In particular, in the case of small energy $E \ll n^2$, and both subsets A', B' of equal size $k = \ell$, one can always find a co-Sidon pair A', B' satisfying $|A'|, |B'| \gg n^{2/3}$, so that their sumset $A' + B'$ has the size $|A'||B'| \gg n^{4/3}$, similarly as in Theorem 3.1. If, however, one of the two subsets is allowed to be larger than the other, one can get a co-Sidon pair A', B' of sizes $|A'| \gg n$ and $|B'| \gg n^{1/2}$ with a larger sumset $|A' + B'| = |A'||B'| \gg n^{3/2}$.

Theorem 3.4 is far from optimal when the additive energy (and so E) is very large. It can then by replaced by the following result (which is similar to one of Komlós, Sulyok and Szemerédi [49]):

Theorem 3.5. *Let A, B be any finite sets of integers of sizes $|A| = |B| = n$, where $n \geqslant 10^6$. Then for all positive integers k, ℓ satisfying*

$$k\ell \leqslant n/12800, \tag{3.2}$$

there exists a pair of co-Sidon subsets $A' \subseteq A$, $B' \subseteq B$ with $|A'| = k$, $|B'| = \ell$.

Theorems 3.4 and 3.5 have their counterparts showing that they are almost optimal at the extreme ends of additive energy:

Theorem 3.6. *For any sufficiently large integer n and any integer E satisfying $n \leqslant E \leqslant 2n^3/3$ there exist two sets A, B of sizes $|A| = |B| = n$ and additive energy $\mathsf{E}(A, B) = |A||B| + E(1 + o(1))$ such that for all integers k, ℓ satisfying $k \geqslant 2\ell$ and*

$$k\ell^2 \geqslant 40n^2 \log n \Big(1 + \frac{3n^2}{E} \log n\Big), \tag{3.3}$$

no subsets $A' \subseteq A, B' \subseteq B$ with $|A'| = k, |B'| = \ell$ are co-Sidon.

Comparing (3.3) with (3.1) (which becomes $k\ell^2 \leqslant n^4/2E$ for $|A| = |B| = n$) we see that for $n \leqslant E \ll n^2$ there is just an extra factor $\log^2 n$ on the right hand side of (3.3).

Theorem 3.7. *For any sufficiently large integer n and any integer E satisfying $n^2 \leqslant E \leqslant 2n^3/3$ there exist sets A, B of sizes $|A| = |B| = n$ and additive energy $\mathsf{E}(A, B) = |A||B| + E(1 + o(1))$ such that for all integers k, ℓ satisfying $k \geqslant \ell$ and*

$$k\ell \geqslant \frac{4n^4}{3E}, \tag{3.4}$$

no subsets $A' \subseteq A, B' \subseteq B$ with $|A'| = k, |B'| = \ell$ are co-Sidon.

This time from (3.4) we see that the inequality (3.2) of Theorem 3.5 is optimal (up to the constant) in the range $n^3 \ll E \leqslant 2n^3/3$.

Finally, for sets with nearly maximal sumset Theorem 3.4 can be improved. Observe first that if $A, B \subseteq \mathbb{Z}$ satisfy

$$|A+B| = |A||B| - s$$

for some non-negative integer $s < |A| + |B|$ then one can remove in total no more than s elements from A and B so that the remaining sets will be co-Sidon. Indeed, for such sets we have

$$\sum_{x \in A+B} (A * B(x) - 1) = s,$$

so after removing at most s elements from A and B one can assure that $A*B(x) \leqslant 1$ for all $x \in \mathbb{Z}$. For larger values of s we employ a different strategy which gives the following result:

Theorem 3.8. *Let A, B be any finite subsets of integers with*

$$|A+B| = |A||B| - s$$

for some s in the range $|B|/4 \leqslant s \leqslant |A||B|/4$. Then there exists a pair of co-Sidon sets $A' \subseteq A$ and $B' \subseteq B$ satisfying

$$|A'| \geqslant |A||B|/4s - 1 \quad \textit{and} \quad |B'| \geqslant |B|/2.$$

A remark on combinatorial background

Before proving the theorems we discuss a related problem, obtained from the formulated above but with no constraints coming from the arithmetics of integers.

Non-trivial additive energy has the following interpretation in terms of graphs. Consider a graph with vertices $A \times B$ and edges connecting two distinct vertices (a, b), (a', b') whenever $a + b = a' + b'$. If for some $x \in A + B$ we have $A * B(x) = k \geqslant 2$ then there are exactly $k(k-1)/2$ edges connecting all the vertices (a, b) for which $a + b = x$. Otherwise, if $A * B(x) = k \in \{0, 1\}$, there are no edges

corresponding to x. Let $\mathsf{e}(A,B)$ be the total number of such edges in the graph with vertices in $A \times B$. Assume that the sum $\sum_{x\in\mathbb{Z}} A*B(x)^2$ contains s nonzero terms $k_1, \cdots, k_s$. Then the additive energy can be written as follows

$$\mathsf{E}(A,B) = \sum_{x\in\mathbb{Z}} A*B(x)^2 = \sum_{j=1}^{s} k_j^2 = 2\sum_{j=1}^{s} \frac{k_j(k_j-1)}{2} + \sum_{j=1}^{s} k_j = 2\mathsf{e}(A,B) + |A||B|.$$

Therefore,

$$2\mathsf{e}(A,B) = \mathsf{E}(A,B) - |A||B| = E. \tag{3.5}$$

Furthermore, the above graph is a union of cliques, as $a_1 + b_1 = a_2 + b_2$ and $a_2 + b_2 = a_3 + b_3$ implies $a_1 + b_1 = a_3 + b_3$. The number of cliques in the graph is nothing else than the size of sumset $|A + B|$.

Forgetting that the graph was constructing from two sets and replacing a graph with a table (which is more natural representation of union of cliques) we get a following version of our main problem.

Question. Given an $n \times n$ table, we color each cell and fix the number E of monochromatic pairs of cells. What is the largest rainbow subtable (that is subtable with all cells of different colors) we can find?

If we would like to ask the same question but give a condition on the size of sumset rather than energy, the equivalent reformulation would be with condition on the total number of colors in the table. Such problems do not seem to be considered before, although a similar situation when only two colors are used in the initial table $|A| \times |B|$ and one looks for a largest subtable colored with one color have been investigated in [79].

The remaining of the chapter is organized as follows. In the next section we prove Theorems 3.4 and 3.8. (These proofs are direct and do not involve any auxiliary results.) In Section 3.3 we prove Theorem 3.5. Then (in Section 3.4) we give several auxiliary lemmas which will be used in the proofs of Theorems 3.6 and 3.7. The proofs of these two theorems will be completed in Section 3.5.

3.2 Proof of Theorems 3.4 and 3.8

Proof of Theorem 3.4. Take positive integers $k \leqslant |A|/2, \ell \leqslant |B|$ satisfying $k\ell^2 \leqslant \frac{|A|^2|B|^2}{2E}$ and construct a graph on $A \times B$ as described earlier (i.e. connect (a,b)

with (a', b') if $a + b = a' + b'$). Consider a spanned subgraph of the graph $A \times B$ with vertices $A_1 \times B_1$, where $A_1 \subseteq A$, $|A_1| = 2k$, $B_1 \subseteq B$, $|B_1| = \ell$. There are exactly

$$\binom{|A|}{2k}\binom{|B|}{\ell}$$

of such subgraphs. Suppose that each of them contains at least $k + 1$ edges and count the total number of edges over all subgraphs. It is easy to see that each edge of the graph with vertices $A \times B$ is counted exactly

$$\binom{|A| - 2}{2k - 2}\binom{|B| - 2}{\ell - 2}$$

times. Since in view of (3.5) there are $\mathsf{e}(A, B) = E/2$ edges in the graph $A \times B$, we must have

$$(k + 1)\binom{|A|}{2k}\binom{|B|}{\ell} \leqslant \binom{|A| - 2}{2k - 2}\binom{|B| - 2}{\ell - 2}\frac{E}{2}.$$

This yields

$$k < k + 1 \leqslant \frac{2k(2k - 1)}{|A|(|A| - 1)}\frac{\ell(\ell - 1)}{|B|(|B| - 1)}\frac{E}{2} \leqslant \frac{2k^2\ell^2 E}{|A|^2|B|^2},$$

and hence

$$k\ell^2 > \frac{|A|^2|B|^2}{2E},$$

contrary to our assumption. This proves that there exists a subgraph with vertices $A_1 \times B_1$ satisfying $|A_1| = 2k$, $|B_1| = \ell$ which contains $\mathsf{e}(A_1, B_1) = k_1$ edges, where $0 \leqslant k_1 \leqslant k$.

To complete the proof for each of the remaining k_1 edges we do the following. Take an edge connecting the vertices (a, b), (a', b') and remove the element a' from the set A_1. This decreases $|A_1|$ by 1 and the number of edges in the subgraph by at least 1. In this way step by step we remove at most k_1 of the elements of the set A_1 so that the remaining subgraph with vertices at $A_2 \times B_1$ will have no edges, and thus the pair A_2, B_1 will be co-Sidon. As

$$|A_2| \geqslant |A_1| - k_1 = 2k - k_1 \geqslant k$$

and $|B_1| = \ell$, any k element subset of A_2 and the set B_1 is a pair of co-Sidon subsets with required cardinalities. $\square$

Proof of Theorem 3.8. Set $C := \{x \in \mathbb{Z} :\ A * B(x) \geqslant 2\}$ and observe that

$$\begin{aligned} 2|C| \leqslant \sum_{x\in C} A * B(x) &= \sum_{x\in A+B} A * B(x) - \sum_{x\in (A+B)\setminus C} A * B(x) \\ &= |A||B| - (|A+B| - |C|) \\ &= s + |C|. \end{aligned}$$

It follows that $|C| \leqslant s$ and

$$\sum_{a\in A} |(C-a)\cap B| = \sum_{x\in C} A * B(x) \leqslant 2s.$$

Then for each $k \leqslant |A|$ there exists a k element subset $A' \subseteq A$ such that

$$\sum_{a\in A'} |(C-a)\cap B| \leqslant k\frac{2s}{|A|}.$$

Taking $k = \lfloor |A||B|/4s \rfloor$ we see that the right hand side of the above formula does not exceed $|B|/2$. Therefore, selecting $B' = B \setminus \cup_{a\in A'}(C-a)$ we find that

$$|B'| \geqslant |B| - \sum_{a\in A'} |(C-a)\cap B| \geqslant |B| - |B|/2 = |B|/2$$

and $(A'+B')\cap C = \emptyset$, so A', B' are co-Sidon subsets A, B with required cardinalities. □

3.3 Proof of Theorem 3.5

In this section we assume that $n \geqslant 12800$ (otherwise no k, ℓ satisfying the condition of Theorem 3.5 exist), although most of the statements hold for smaller n as well. The proof of Theorem 3.5 follows the ideas of [49]. Call a pair of maps

$$\varphi : A \to \mathbb{Z} \cup \{\emptyset\} \quad \text{and} \quad \psi : B \to \mathbb{Z} \cup \{\emptyset\}$$

A, B-preserving if for all $a, a' \in A$ and $b, b' \in B$ we have

- $\varphi(a) = \varphi(a') \in \mathbb{Z} \implies a = a'$,
- $\psi(b) = \psi(b') \in \mathbb{Z} \implies b = b'$,
- $a+b = a'+b' \implies \varphi(a)+\psi(b) = \varphi(a')+\psi(b')$ or $\emptyset \in \{\varphi(a), \varphi(a'), \psi(b), \psi(b')\}$.

Denote $\varphi_{\mathbb{Z}}(A) = \varphi(A) \cap \mathbb{Z}$ and $\psi_{\mathbb{Z}}(B) = \psi(B) \cap \mathbb{Z}$. Observe that if for A, B-preserving maps φ, ψ the subsets $X \subseteq \varphi_{\mathbb{Z}}(A), Y \subseteq \psi_{\mathbb{Z}}(B)$ are co-Sidon, then $\varphi^{-1}(X) \subseteq A, \psi^{-1}(Y) \subseteq B$ are also co-Sidon. Then, in order to find large co-Sidon subsets of A, B, it is sufficient to map sets A, B efficiently to a set (an interval in this case) that is known to contain large co-Sidon subsets. This mapping is done in four steps (Lemmas 3.9, 3.10, 3.11 and 3.12) similar to those in [49].

Lemma 3.9. *For any $A, B \subseteq \mathbb{Z}$, $|A| = |B| = n$, there exists a pair of A, B-preserving maps φ, ψ satisfying $\varphi(A) \subseteq [1, 4^{2n}]$, $\psi(B) \subseteq [1, 4^{2n}]$.*

Proof. Consider a union $A \cup B$ which satisfies $|A \cup B| \leqslant 2n$ and use the fact that for any k element set there exist a Freiman isomorphic subset of $[1, 4^k]$ (see Exercise 21, p. 128 in [33]). Recall that sets X and Y are Freiman isomorphic if there exists a bijective map (*Freiman's isomorphism*) $f : X \to Y$ such that $x_1 + x_2 = x_3 + x_4 \iff f(x_1) + f(x_2) = f(x_3) + f(x_4)$. In particular, Freiman's isomorphism restricted to A and B forms a pair of A, B-preserving maps. □

Lemma 3.10. *For any $A, B \subseteq [1, k]$, $|A| = |B| = n$, there exists a pair of A, B-preserving maps φ, ψ satisfying*

$$\varphi_{\mathbb{Z}}(A) \subseteq [1, 2n^2 \log^2 k], \quad \psi_{\mathbb{Z}}(B) \subseteq [1, 2n^2 \log^2 k]$$

and

$$|\varphi_{\mathbb{Z}}(A)| \geqslant n/2, \quad |\psi_{\mathbb{Z}}(B)| \geqslant n/2.$$

Proof. If $k < n^2$ the statement is trivial, so assume the contrary. Observe that the number of prime divisors of all differences $a - a'$, where $a > a'$, $a, a' \in A$, and $b - b'$, where $b > b'$, $b, b' \in B$, is less than $n^2 \log k$, because there are at most $n(n-1)$ of such differences, each does not exceed $k - 1$ and so it has at most $\log k$ prime divisors for $k \geqslant 8$. There are more than $n^2 \log k$ primes in the interval $[1, 2n^2 \log^2 k]$, so take one, $q > 2$, which does not divide any of these differences.

Each element of A when reduced modulo q belongs to one of the two intervals: $[1, q/2)$ or $(q/2, q]$, thus naturally dividing A into two parts. Define φ by mapping the larger of these two parts to the corresponding least positive residues modulo

q and the smaller one to $\emptyset$. Define ψ analogously. Clearly, this pair of maps φ, ψ satisfies the conditions of the lemma. $\square$

Lemma 3.11. *For any $A, B \subseteq [1, n^5]$, $|A| = |B| = n$, there exists a pair of A, B-preserving maps φ, ψ satisfying*

$$\varphi_{\mathbb{Z}}(A) \subseteq [1, n^{3/2}], \quad \varphi_{\mathbb{Z}}(B) \subseteq [1, n^{3/2}]$$

and

$$|\varphi_{\mathbb{Z}}(A)| \geqslant 9n/20, \quad |\varphi_{\mathbb{Z}}(B)| \geqslant 9n/20.$$

Proof. Let us index the elements of A and B arbitrarily. For a prime number p define $f_p(i,j)$, $1 \leqslant i < j \leqslant n$, as follows:

$$f_p(i,j) = \begin{cases} 1 & \text{if} \quad p|(a_i - a_j) \text{ or } p|(b_i - b_j), \\ 0 & \text{otherwise.} \end{cases}$$

Since

$$\sum_p f_p(i,j) \leqslant 2\log(n^5) = 10 \log n,$$

we have

$$\sum_p \sum_{(i,j)} f_p(i,j) \leqslant \frac{n(n-1)}{2} 10 \log n < 5n^2 \log n.$$

It follows that there exists prime $q \leqslant n^{3/2}$ such that

$$\sum_{(i,j)} f_q(i,j) \leqslant \frac{5n^2 \log n}{\pi(n^{3/2})} < n/10, \tag{3.6}$$

where the above inequality follows from the estimate $\pi(x) > x/\log x$ which holds for all $x \geqslant 17$ (see [67]) and the assumption $n \geqslant 10^6$.

Remove all the elements a, a' from A for which $q|(a-a')$ (by (3.6), there are less than $n/10$ of them). As in the proof of the previous lemma, split the remaining elements into two groups according to which interval, $[1, q/2)$ or $(q/2, q]$, their residues modulo q belong to. Again define φ by mapping the larger part (which contains at least $(n - n/10)/2 = 9n/20$ elements) to corresponding residues and the remaining elements (including removed ones) to $\emptyset$. Define ψ analogously. $\square$

Lemma 3.12. *For any $A, B \subseteq [1, n^2/8]$, $|A| = |B| = n$, there exists a pair of A, B-preserving maps φ, ψ satisfying*

$$\varphi_{\mathbb{Z}}(A) \subseteq [1, 8n], \quad \varphi_{\mathbb{Z}}(B) \subseteq [1, 8n]$$

and

$$|\varphi_{\mathbb{Z}}(A)| \geqslant n/8, \quad |\varphi_{\mathbb{Z}}(B)| \geqslant n/8.$$

Proof. Index elements of A and B arbitrarily and choose an odd prime $p \in [n/4, n/2]$. As above, split all the elements of A into two groups according to which interval, $[1, p/2)$ or $(p/2, p]$, their residues modulo p belong to. Choose the larger group and do the same for B (remaining elements will be mapped to $\emptyset$).

Let $q \in [4n + 1, 8n]$ be another prime. We will chose φ and ψ of the form

$$\varphi(a_k) = (\omega p h_k + r_k) \bmod q,$$
$$\psi(b_k) = (\omega p g_k + s_k) \bmod q,$$

where by mod q we mean a map to the least positive residue modulo q. Here r_k, s_k are the least positive residues modulo p so that $a_k = ph_k + r_k$, $b_k = pg_k + s_k$ ($\omega \in [1, q-1]$ will be chosen below). Define $f_\omega(i, j)$ for $w \in [1, q-1]$, $1 \leqslant j < i \leqslant n$, as

$$f_\omega(i,j) = \begin{cases} 1 & \text{if} \quad q|(\omega p h_i + r_i - \omega p h_j - r_j) \text{ or } q|(\omega p g_i + s_i - \omega p g_j - s_j), \\ 0 & \text{otherwise.} \end{cases}$$

Since for every pair (i, j) we have

$$\sum_{\omega \in [1, q-1]} f_\omega(i,j) \leqslant 2,$$

and so

$$\sum_{\omega \in [1, q-1]} \sum_{(i,j)} f_\omega(i,j) \leqslant n^2,$$

there exists $\omega_0 \in [1, q-1]$ for which

$$\sum_{(i,j)} f_{\omega_0}(i,j) \leqslant \frac{n^2}{q-1} \leqslant \frac{n}{4}.$$

Choose $\omega = \omega_0$ and consider φ, ψ as define above. For them to be A, B-preserving, first remove no more than $n/4$ elements from the remaining parts of

A and B, so that all the remaining differences $\varphi(a) - \varphi(a')$ and $\psi(b) - \psi(b')$ were not divisible by q. Then divide the remaining parts of A and B again, according to which interval, $[1, q/2)$ or $(q/2, q]$, residues of $\varphi(a_k), \psi(b_k)$ modulo q belong to, and choose the larger parts. Map them as defined above and map the remaining elements to $\emptyset$. □

Proof of Theorem 3.5. Starting from any sets $A, B \subset \mathbb{Z}$, $|A| = |B| = n$, we will apply Lemmas 3.9, 3.10, 3.11 and 3.12 one after another and obtain the sum preserving maps $\varphi^{(i)}, \psi^{(i)}, i = 1, 2, 3, 4$. The first two steps give the sets

$$A_1 := \varphi^{(1)}(A) \subset [1, 4^{2n}], \quad B_1 := \psi^{(1)}(B) \subset [1, 4^{2n}],$$

$$|A_1| = |B_1| = n$$

and

$$A_2 := \varphi_{\mathbb{Z}}^{(2)}(A_1) \subset [1, 8n^4 \log^2 4], \quad B_2 := \psi_{\mathbb{Z}}^{(2)}(B_1) \subset [1, 8n^4 \log^2 4],$$

$$|A_2|, |B_2| \geqslant n/2.$$

Next, take arbitrary $\lfloor n/2 \rfloor$ elements of A_2, arbitrary $\lfloor n/2 \rfloor$ elements of B_2 and apply Lemma 3.11 to those sets (without changing notation). Since the role of n is played by $\lfloor n/2 \rfloor$, we obtain

$$A_3 := \varphi_{\mathbb{Z}}^{(3)}(A_2) \subset [1, n^{3/2}], \quad B_3 := \psi_{\mathbb{Z}}^{(3)}(B_2) \subset [1, n^{3/2}],$$

$$|A_3|, |B_3| \geqslant 9\lfloor n/2 \rfloor/20 > n/5.$$

Finally, let us take arbitrary $\lceil n/5 \rceil$ elements of A_3, arbitrary $\lceil n/5 \rceil$ elements of B_3 and apply Lemma 3.12 to those sets (again without changing the notation A_3, B_3). Now the role of n is played by $\lceil n/5 \rceil$, so we obtain

$$A_4 := \varphi_{\mathbb{Z}}^{(4)}(A_3) \subset [1, 2n], \quad B_4 := \psi_{\mathbb{Z}}^{(4)}(B_3) \subset [1, 2n],$$

$$|A_4|, |B_4| \geqslant \lceil n/5 \rceil/8 \geqslant n/40.$$

Let k, ℓ be two integers satisfying $k\ell \leqslant n/12800$. Take X to be the set $[1, 160k]$ and Y to be the arithmetic progression $160ks$, $s = 1, 2, \ldots, 160\ell$. Then the largest element of Y is $160k \cdot 160\ell \leqslant 160^2 n/12800 = 2n$. It is easy to see that X, Y are co-Sidon subsets of $[1, 2n]$ of sizes $160k, 160\ell$, respectively.

Note that

$$\sum_{t\in[-2n,2n]} |A_4 \cap (X+t)| = |A_4||X|,$$

so there exists a shift $t_0 \in [-2n, 2n]$ of X such that

$$|A_4 \cap (X+t)| \geqslant \frac{|A_4||X|}{4n} \geqslant \frac{n|X|}{40\cdot 4n} = \frac{160k}{160} = k.$$

Analogously, there exists a shift $r \in [-2n, 2n]$ of B_4 such that

$$|B_4 \cap (Y+r)| \geqslant \frac{|B_4||Y|}{4n} \geqslant \frac{|B_4||Y|}{4n} \geqslant \frac{n|Y|}{40\cdot 4n} = \frac{160\ell}{160} = \ell.$$

From this we get that A_4, B_4 contain two co-Sidon subsets of sizes k, ℓ, thus so do A and B. □

3.4 Auxiliary lemmas

The following concentration inequalities of Kim-Vu [47] and Chernoff [44, p. 26] will be used in a probabilistic construction of the example below. We note that weaker inequalities would suffice, but we use these as they are easier to apply.

Lemma 3.13. *Let t_i, $i \in [1,n]$, be independent indicator random variables, and let Y be a polynomial in t_i of degree k with positive coefficients. For a set $I \subseteq [1,n]$ let Y_I be a partial derivative of Y with respect to the variables t_i, where $i \in I$. Define M_i to be the maximal value of $\mathbb{E}(Y_I)$ over all $I \subseteq [1,n]$ of cardinality i. Set*

$$M' = \max_{1\leqslant i\leqslant k} M_i \quad \text{and} \quad M = \max\{\mathbb{E}(Y), M'\}.$$

Then for any $\lambda > 1$ we have

$$\mathbb{P}\left[|Y - \mathbb{E}(Y)| > (MM')^{1/2}(8\lambda)^k\sqrt{k!}\right] < 2e^{2-\lambda}n^{k-1}.$$

Lemma 3.14. *Let X_i, $i \in [1,n]$, be independent, equally distributed indicator random variables and $X = \sum_{i=1}^n X_i$. Then for any $t \geqslant 0$ we have*

$$\mathbb{P}[X \geqslant \mathbb{E}(X) + t] \leqslant e^{-\frac{t^2}{2(\mathbb{E}(X)+t/3)}}. \tag{3.7}$$

We shall also need the following lemma.

Lemma 3.15. *Take two random sets A and B from $[1,m]$, by selecting elements for each set from $[1,m]$ independently with probability $p = n/m$, where $m^{1/3} \leqslant n \leqslant m$ and m is large enough. Then*

$$n \leqslant |A|, |B| \leqslant n + O(n^{1/2})$$

with probability greater than $1/9$,

$$\mathsf{E}(A,B) = |A||B| + (2/3 + o(1))n^4/m$$

with probability $1 - o(1)$, and

$$\max_{x \neq 0} \delta_B(x) \leqslant 2n^2/m + 3\log m$$

with probability greater than $9/10$.

Proof. Let $\mathbb{I}_j^A, \mathbb{I}_i^B$ be the indicator functions of $j \in A$ and $i \in B$. Then

$$|A| = \sum_{i \in [1,m]} \mathbb{I}_i^A \quad \text{and} \quad \mathbb{E}\,|A| = n.$$

Selecting $t = 2n^{1/2}$ in Chernoff's inequality (3.7), we obtain

$$\mathbb{P}\left[|A| \geqslant n + 2n^{1/2}\right] \leqslant e^{-2+o(1)} < 1/6.$$

On the other hand, the median of $|A|$ is equal to n, so

$$\mathbb{P}\left[n \leqslant |A| \leqslant n + 2n^{1/2}\right] > 1/2 - 1/6 = 1/3.$$

Analogously, we get the same inequality for B and, since A and B were chosen independently, this proves the first part.

For the second part, using (3.5) we write

$$\begin{aligned} 2\mathsf{e}(A,B) &= \mathsf{E}(A,B) - |A||B| \\ &= \sum_{x \in [2,2m]} \big(A * B(x^2) - A * B(x)\big) = \sum_{x \in [2,2m]} \sum_{i \neq j} \mathbb{I}_i^A \mathbb{I}_{x-i}^B \mathbb{I}_j^A \mathbb{I}_{x-j}^B. \end{aligned}$$

When $i \neq j$ all the indicators $\mathbb{I}_i^A, \mathbb{I}_{x-i}^B, \mathbb{I}_j^A, \mathbb{I}_{x-j}^B$ are independent (note that only pairs corresponding to the same set A or B could be dependent, but such dependences are ruled out by condition $i \neq j$) with expectations equal to p, so for

$2 \leqslant x \leqslant m+1$ we have

$$\mathbb{E}\left(\sum_{i\neq j} \mathbb{I}_i^A \mathbb{I}_{x-i}^B \mathbb{I}_j^A \mathbb{I}_{x-j}^B\right) = \sum_{i\neq j} \mathbb{E}\left(\mathbb{I}_i^A\right)\mathbb{E}\left(\mathbb{I}_{x-i}^B\right)\mathbb{E}\left(\mathbb{I}_j^A\right)\mathbb{E}\left(\mathbb{I}_{x-j}^B\right)$$
$$= (x-1)(x-2)p^4.$$

From symmetry $\mathbb{E}\left(A * B(x)\right) = \mathbb{E}\left(A * B(2m+2-x)\right)$ we then get

$$\mathbb{E}\left(2\mathsf{e}(A,B)\right) = 2\left(\sum_{x\in[2,m]} (x-1)(x-2)p^4\right) + m(m-1)p^4$$
$$= p^4(2m^3/3 - m^2 + m/3)$$
$$= (2/3 + o(1))n^4/m.$$

In order to use Lemma 3.13 we will bound the averages of the partial derivatives of $2\mathsf{e}(A,B)$. Extend $\mathbb{I}_x^A$ and $\mathbb{I}_x^B$ to be equal to 0 when $x \notin [1,m]$ and without loss of generality write

$$M_1 = \max_j \mathbb{E}\left(2\mathsf{e}(A,B)'_{\mathbb{I}_j^A}\right) = \max_j \mathbb{E}\left(\sum_{x,i\neq j} \mathbb{I}_i^A \mathbb{I}_{x-i}^A \mathbb{I}_{x-j}^B\right),$$

which is no more than m^2p^3, as there are no more than m^2 values of (x,i) for which $\mathbb{E}\left(\mathbb{I}_i^A \mathbb{I}_{x-i}^A \mathbb{I}_{x-j}^B\right)$ is equal to p^3 rather than zero. Analogously, we get

$$M_2 = \max\left\{\max_{i\neq j} \mathbb{E}\left(\sum_x \mathbb{I}_{x-i}^A \mathbb{I}_{x-j}^B\right), \max_{j,x} \mathbb{E}\left(\sum_{i\neq j} \mathbb{I}_i^A \mathbb{I}_{x-i}^A\right)\right\} \leqslant mp^2,$$
$$M_3 = \max_{i\neq j,x} \mathbb{E}\,\mathbb{I}_{x-i}^A \leqslant p,$$
$$M_4 = 1.$$

This yields $M' \leqslant m^2p^3 = n^3/m$ and $(MM')^{1/2} \leqslant n^{7/2}/m$, so selecting $\lambda = n^{1/16}$ in Lemma 3.13 we arrive to the required concentration.

For the third part, we use the inequality

$$\mathbb{P}[\delta_B(1) \geqslant a] \geqslant \mathbb{P}[\delta_B(x) \geqslant a],$$

which is valid for all $a > 0$ and $x \neq 0$ as 1 has the largest number of representations as a difference of two elements of the interval $[1,m]$. We then bound the probability in question as follows:

$$\mathbb{P}[\max_{x\neq 0} \delta_B(x) \geqslant 2n^2/m + 3\log m] \leqslant \sum_{\substack{x\in[-m+1,m-1]\\ x\neq 0}} \mathbb{P}[\delta_B(x) \geqslant 2n^2/m + 3\log m]$$
$$\leqslant 2m\mathbb{P}[\delta_B(1) \geqslant 2n^2/m + 3\log m]$$
$$\leqslant 2m\mathbb{P}[\delta_B(1) \geqslant 2(m-1)p^2 + 3\log m].$$

In order to estimate the latter probability we write

$$\delta_B(1) = \sum_{x \leqslant m/2} \mathbb{I}_{2x}\mathbb{I}_{2x-1} + \sum_{x \leqslant m/2} \mathbb{I}_{2x+1}\mathbb{I}_{2x} := \Sigma_1 + \Sigma_2.$$

Note that both Σ_1, Σ_2 are sums of independent indicator random variables and

$$\mathbb{E}(\Sigma_1) = \lfloor m/2 \rfloor p^2, \quad \mathbb{E}(\Sigma_2) = (\lceil m/2 \rceil - 1)p^2.$$

Hence

$$\mathbb{E}(\Sigma_1) + \mathbb{E}(\Sigma_2) = (m-1)p^2$$

and $2m\mathbb{P}[\delta_B(1) \geqslant 2(m-1)p^2 + 3\log m]$ does not exceed the sum

$$2m\mathbb{P}[\Sigma_1 \geqslant 2\mathbb{E}(\Sigma_1) + 1.5\log m] + 2m\mathbb{P}[\Sigma_2 \geqslant 2\mathbb{E}(\Sigma_2) + 1.5\log m]. \tag{3.8}$$

Applying Lemma 3.14 with $t = \mathbb{E}(\Sigma_1) + 1.5\log m$ we find that

$$\frac{t^2}{2\mathbb{E}(\Sigma_1) + 2t/3} = \frac{\mathbb{E}(\Sigma_1)^2 + 3\mathbb{E}(\Sigma_1)\log m + 2.25\log^2 m}{8\mathbb{E}(\Sigma_1)/3 + \log m} > \frac{9\log m}{8}.$$

Hence

$$\mathbb{P}[\Sigma_1 \geqslant 2\mathbb{E}(\Sigma_1) + 1.5\log m] < m^{-9/8} < 1/40m$$

for m large enough. Similarly, applying Lemma 3.14 with $t = \mathbb{E}(\Sigma_2) + 1.5\log m$, we find that

$$\mathbb{P}[\Sigma_2 \geqslant 2\mathbb{E}(\Sigma_2) + 1.5\log m] < 1/40m$$

for m large enough. Hence the sum (3.8) does not exceed $2m/40m + 2m/40m = 0.1$, and the result follows. □

The following two lemmas are not necessary for the proofs of the results of this chapter. Rather they establish the stated bounds on the additive energy.

Lemma 3.16. *For a set of integers A, the maximal possible sum of ℓ largest values of $A \circ A$ is equal to*

$$\begin{cases} |A|^2, & \text{if } |A| < \frac{\ell+1}{2}, \\ |A|\ell - \lfloor \frac{\ell^2}{4} \rfloor, & \text{if } |A| \geqslant \frac{\ell+1}{2}. \end{cases}$$

If $\ell \geqslant 2$, then the second bound is only obtained when A is an arithmetic progression.

Proof. Note that the sum of ℓ largest values of $A \circ A$ is always smaller than the sum of all values, which is equal to $|A|^2$. This bound can be obtained if ℓ is larger than $|A + A|$, which is possible when $|A| < \frac{\ell+1}{2}$.

For the case $|A| \geqslant \frac{\ell+1}{2}$ note that adding additional element to A increases each value of $A \circ A$ by at most one, so the sum of ℓ largest values increases by at most ℓ. This implies that for $|A| \geqslant \frac{\ell+1}{2}$ the sum of ℓ largest values will not be larger than $\lfloor \frac{\ell+1}{2} \rfloor^2 + (|A| - \lfloor \frac{\ell+1}{2} \rfloor)\ell$, which simplifies to the expression in the statement of the lemma.

In order to see for which sets this bound can be obtain, one considers a procedure in which A is constructed by adding elements one by one. First $\lfloor \frac{\ell+1}{2} \rfloor$ elements (denote their set by A_0) have to have their difference set supported on the set of cardinality ℓ, which is only possible when A_0 is an arithmetic progression. Indeed, for even ℓ this condition is equivalent to $|A_0 + A_0| \leqslant 2|A_0|$, and, since size of difference set is always odd, to $|A_0 + A_0| \leqslant 2|A_0| - 1$ (which is known to imply that A_0 is an arithmetic progression). The case ℓ odd is handled in the same manner. Next, addition of every subsequent element has to increase each of ℓ largest values of $A \circ A$. Since we start from an arithmetic progression, the second largest value of $A \circ A$ is attained at the difference of progression, so we have to place new element at that distance from the old ones. The only two ways to achieve this result in a longer progression. □

Lemma 3.17. *Let A, B be finite sets of integers with $|A| \geqslant |B|$. Then*

$$\mathsf{E}(A, B) \leqslant |A||B|^2 - \frac{|B|^3}{3} + \frac{|B|}{3}.$$

If $|B| \geqslant 2$, the equality is obtained iff A and B are arithmetic progressions with equal differences.

Proof. We fix the size $|A|$ and argue by induction on the size of B. For $|A| = 1$ and $|B| = 1$ the statement is true (an equality holds), so assume that $|A| \geqslant 2$, and that the statement is known for $|B| = \ell \geqslant 1, \ell < |A|$. We will prove it for $|B| = \ell + 1$.

Take one element, say b_0, from B and denote the remaining set by B_0. By induction hypothesis $\mathsf{E}(A, B_0) \leqslant |A|\ell^2 - \frac{\ell^3}{3} + \frac{\ell}{3}$. The difference $\mathsf{E}(A, B) - \mathsf{E}(A, B_0)$ is equal to the number of solutions to equation $a - a' = b - b'$, where $a, a' \in A$,

$b, b' \in B$ and at least one of b, b' is equal to b_0. There are $|B_0| + |B_0| + 1 = 2\ell + 1$ possibilities for the right hand side and all the differences are distinct, hence the number of solutions is as large as the sum of $2\ell + 1$ largest values of $A \circ A$. As $|A| \geqslant \frac{2\ell+1+1}{2}$, the second case of lemma 3.16 bounds this by $|A|(2\ell + 1) - \ell^2 - \ell$, which is precisely the difference $(|A|(\ell + 1)^2 - \frac{(\ell+1)^3}{3} + \frac{\ell+1}{3}) - (|A|\ell^2 - \frac{\ell^3}{3} + \frac{\ell}{3})$.

Let us check now when does the equality hold. First of all, since $2\ell + 1 \geqslant 2$, lemma 3.16 gives us that A has to be an arithmetic progression. Having this we know that the largest $2\ell + 1$ values of $A \circ A$ are attained at the multiples of the difference d, that is on the set $\{id, i \in [-\ell, \ell]\}$. Clearly, for set of differences $b - b'$, where $b, b' \in B$ and one of b, b' is equal to b_0, to coincide with $\{id, i \in [-\ell, \ell]\}$, the set B has to be an arithmetic progression of the difference d itself. □

Lemma 3.18. *Let A, B be finite sets of integers. Then*

$$\mathsf{E}(A, B) \geqslant |A||B|.$$

Proof.

$$\begin{aligned}\mathsf{E}(A, B) &= |\{a + b = a' + b' \mid a, a' \in A, b, b' \in B\}| \\ &\geqslant |\{a + b = a + b \mid a \in A, b \in B\}| = |A||B|.\end{aligned}$$

□

3.5 Proofs of Theorems 3.6 and 3.7

For simplicity throughout this section we will omit the floor and ceiling signs in binomial coefficients. In order to prove Theorem 3.6 we will use the following lemma from [48]:

Lemma 3.19. *Let G be a graph with N vertices, $q \in \mathbb{N}$ and let $0 \leqslant \beta \leqslant 1$ and R be real numbers satisfying*

$$R \geqslant e^{-\beta q} N. \tag{3.9}$$

Suppose the number of edges $e(U)$ induced in G by any set $U \subseteq V(G)$ with $|U| \geqslant R$ satisfies

$$e(U) \geqslant \beta \binom{|U|}{2}. \tag{3.10}$$

Then, for all integers $r \geqslant 0$, the number of independent sets of cardinality $q + r$ in G is at most

$$\binom{N}{q}\binom{R}{r}.$$

We will use this lemma to count some specific pairs of co-Sidon sets. A set X is called Δ-*random* if $\max_{x\in\mathbb{Z}} \delta_X(x) \leqslant \Delta$.

Lemma 3.20. *Let k, ℓ, m be integers satisfying $k, \ell < m$ and*

$$k \geqslant 4\Delta m\ell^{-2}\log\ell.$$

Then the number of co-Sidon sets A' and B' in the interval $[1, m]$ with $|A'| = k$, $|B'| = \ell$ and B' being Δ-random is less than

$$\binom{m}{\ell}\binom{m}{4\Delta m\ell^{-2}\log\ell}\binom{4m/\ell}{k - 4\Delta m\ell^{-2}\log\ell}.$$

Proof. Let B' be any Δ-random ℓ element subset of $[1, m]$ (note that there are at most $\binom{m}{\ell}$ of such sets). We will bound the number of subsets $A' \subseteq [1, m]$ of size k that are co-Sidon with B'.

Let G be a graph with vertex set $[1, m]$ and $a, a' \in [1, m]$ are connected by an edge whenever $a - a' \in B' - B'$. Note that each independent vertex set in G corresponds to a co-Sidon pair A', B' with $A' \subseteq [1, m]$, and vice versa.

We will use Lemma 3.19 to bound the number of independent vertex sets in G of size k. For this we put

$$\begin{aligned} R &:= 4m/\ell, \\ \beta &:= \ell^2/4\Delta m, \\ q &:= 4\Delta m\ell^{-2}\log\ell, \\ r &:= k - q. \end{aligned}$$

Since $\beta q = \log\ell$ and $N = m$, one can easily check that (3.9) holds. We will show that (3.10) holds as well. This will imply the statement of the lemma, as one can see easily by substituting the parameter values chosen above.

Let U be any subgraph of G with $|V(U)| \geqslant R$, then

$$\begin{aligned} 2e(U) &= |\{u_1 - u_2 \in B' - B', u_1, u_2 \in U, u_1 \neq u_2\}| \\ &\geqslant \frac{1}{\Delta}|\{u_1 - u_2 = b_1 - b_2, u_1, u_2 \in U, b_1, b_2 \in B', u_1 \neq u_2\}| \\ &= \frac{1}{\Delta}(\mathsf{E}(U, B') - |U||B'|). \end{aligned}$$

Cauchy-Schwarz inequality implies $\mathsf{E}(U, B') \geqslant \frac{|U|^2|B'|^2}{|U+B'|} \geqslant \frac{|U|^2|B'|^2}{2m}$ (see, e.g., p. 63 in [81]), so $e(U) \geqslant \frac{|U|\ell}{2\Delta}\left(\frac{|U|\ell}{2m} - 1\right)$. Noting that $|U| \geqslant R = 4m/\ell$ implies $|U|\ell \geqslant 2m + |U|\ell/2$ we get the required bound:

$$\frac{|U|\ell}{2\Delta}\left(\frac{|U|\ell}{2m} - 1\right) \geqslant \frac{\ell^2}{4\Delta m}\frac{|U|^2}{2} > \beta\binom{|U|}{2}.$$

This proves (3.10) and completes the proof of the lemma. □

Proof of Theorem 3.6. Take $m = 2n^4/3E$ and two random sets A and B from $[1, m]$, by selecting the elements for each from $[1, m]$ independently with probability $p = n/m$. As for such m we have $m^{1/3} \leqslant n \leqslant m$, Lemma 3.15 implies that with probability greater than, say, $1/100$ we have $n \leqslant |A|, |B| \leqslant n + O(n^{1/2})$, $\mathsf{E}(A, B) = |A||B| + E(1 + o(1))$ and B Δ-random for $\Delta = 3E/n^2 + 9\log n$ (call this event I).

Now for each ℓ define

$$k_\ell := \lceil 20m\Delta\ell^{-2}\log\ell\rceil. \tag{3.11}$$

We shall prove that with probability $1 - o(1)$ randomly taken sets A and B will not contain co-Sidon subsets A', B' of sizes k_ℓ, ℓ for the values $\ell \in L$ that satisfy $n \geqslant k_\ell \geqslant 2\ell$ with additionally property that B' is Δ-random (call this event II). Since $1/100+1-o(1) > 1$ for n large enough, we will have that both events happen with positive probability, thus there will exist sets A, B satisfying conditions of event I and not having co-Sidon subsets of required size (as due to event I set B will only have Δ-random subsets).

From Lemma 3.20 it follows that number of co-Sidon sets $A', B' \subseteq [1, m]$ with $|A'| = k_\ell, |B'| = \ell$ and B' being Δ-random is less than

$$\binom{m}{\ell}\binom{m}{k_\ell/5}\binom{4m/\ell}{4k_\ell/5}.$$

From the union bound, the probability that A, B contain co-Sidon pair of sizes k_ℓ, ℓ ($\ell \in L$) can be bounded by

$$\sum_{\ell \in L} \binom{m}{\ell} \binom{m}{k_\ell/5} \binom{4m/\ell}{4k_\ell/5} (n/m)^{k_\ell+\ell}.$$

Using the inequality $\binom{s}{t} < (se/t)^t$ we can estimate each summand from above by

$$n^{\ell+k_\ell} \ell^{-\ell-4k_\ell/5} k_\ell^{-k_\ell} e^{\ell+k_\ell} 5^{k_\ell}. \tag{3.12}$$

As $n \leqslant (m\Delta)^{1/2} \leqslant \ell k_\ell^{1/2}$ (see (3.11)) and $\ell \leqslant k_\ell/2$, we obtain

$$\begin{aligned} n^{\ell+k_\ell} \ell^{-\ell-4k_\ell/5} k_\ell^{-k_\ell} &\leqslant \ell^{\ell+k_\ell} k_\ell^{\ell/2+k_\ell/2} \ell^{-\ell-4k_\ell/5} k_\ell^{-k_\ell} \\ &= \ell^{k_\ell/5} k_\ell^{\ell/2-k_\ell/2} \\ &< k_\ell^{k_\ell/5+k_\ell/4-k_\ell/2} = k_\ell^{-k_\ell/20}. \end{aligned}$$

Combining this with (3.12) and $k_\ell \geqslant n^{2/3}$, $|L| \leqslant n$ (which are implied by (3.11) and $n \geqslant k_\ell \geqslant 2\ell$) we bound the sum by

$$n k_\ell^{-k_\ell/20} e^{\ell+k_\ell} 5^{k_\ell} < n k_\ell^{-k_\ell/20} 23^{k_\ell} < n^{1-n^{2/3}/31},$$

which tends to zero with $n \to \infty$.

Finally, note that we can make A and B to be exactly of sizes n (instead of $n \leqslant |A|, |B| \leqslant n + O(n^{1/2})$) by removing extraneous elements. This will not increase the sizes of the largest co-Sidon pair and by selecting which elements to remove we can assure that this will not effect the additive energy. As (3.11) is implied by the condition of the theorem, we will have shown that required sets exist with positive probability which proves the theorem.

Indeed, denote by $e(x)$ ($x \in A$ or $x \in B$) the number of equations $a+b = a'+b'$ (where $a, a' \in A, b, b' \in B$ and $a \neq a', b \neq b'$) that x participates in and write

$$E = \mathsf{E}(A, B) - |A||B| = \frac{1}{2} \sum_{x \in A} e(x).$$

From here it follows that there exists a subset A' of size $|A| - n = O(n^{1/2})$ such that

$$\frac{1}{2} \sum_{x \in A'} e(x) \leqslant \frac{|A| - n}{|A|} E = o(E).$$

Then remove A' from A and do the same for B. □

Proof of Theorem 3.7. As above take $m = 2n^4/3E$ and two random sets A and B from $[1, m]$, by selecting the elements for each from $[1, m]$ independently with probability $p = n/m$. Again, with probability greater than, say, $1/100$ we have $n \leqslant |A|, |B| \leqslant n + O(n^{1/2})$ and $\mathsf{E}(A, B) = |A||B| + E(1 + o(1))$.

Now simply note that no subsets of $[1, m]$ (and hence no subsets of A and B) of sizes k, ℓ with $k\ell \geqslant 2m$ are co-Sidon (see [5] for the proof of this simple statement), so to prove the existence of required sets, it remains to remove extraneous elements from A and B as is done above. $\square$

4 Sparse infinite sets with small sumset

Celebrated Freiman's theorem describes all finite sets with small sumset. Sós asked [19] (Problem 4.2) whether it is possible to describe infinite sets with small sumset:

Question. Let $A \subseteq \mathbb{N}$ be an infinite subset of positive integers, and suppose that

$$|A[n] + A[n]| \ll |A[n]|.$$

What can be said about the set A?

One can easily see that sets with positive lower density, that is sets satisfying $\frac{|A[n]|}{n} \gg 1$, do have small sumset in the above sense. In this chapter we consider the remaining case of *sparse* sets, those with lower density equal to zero.

In the first theorem of the chapter we consider sparse sets with bounded *jumps*, that is with bounded quotients of subsequent elements of A: $\frac{a_{i+i}}{a_i} \ll 1$. We show that such sets cannot have small sumset. In the second theorem we impose stronger conditions on the growth of the set A and get a more precise estimate on the growth of $\frac{|A[n]+A[n]|}{|A[n]|}$.

To show the strength of above theorems we also give two contrasting examples. First one is of the sparse set with small sumset (but unbounded jumps, of course) and the second with the growth of the sumset nearly matching the growth of the second theorem.

Before we formulate and prove the exact results, we would like to mention one result of this taste which was obtained by Nash [59], answering a question of Erdős. He showed that set satisfying $\lim_{n\to\infty} \frac{|A[n]|}{n} = 0$ and $A + A = \mathbb{N}$ cannot have a small sumset. Note that the condition of the growth implies that the set A has to be sparse, and the condition on the sumset implies that it has bounded jumps. Sets satisfying condition $A + A = \mathbb{N}$[1] are called *a basis* or *an additive basis*

[1] actually $\mathbb{N} \setminus \{1\}$ as in this thesis we take $\mathbb{N} = \{1, 2, 3, \dots\}$

of positive integers and are well investigated. We do not discuss them in this work and direct an interested reader to a classical book *Sequences* by Halberstam and Roth [39].

4.1 The problem and results

We start from the first theorem we already mentioned:

Theorem 4.1. *Let $A \subseteq \mathbb{N}$ be an infinite subset of positive integers satisfying*

$$\liminf_{n\to\infty} \frac{|A[n]|}{n} = 0 \quad \text{and} \quad \frac{a_{n+1}}{a_n} \ll 1.$$

Then $\frac{|A[n]+A[n]|}{|A[n]|}$ is unbounded.

In contrast to this theorem we give an example of a sparse set with small sumset. Such an example is not hard to come by, so we give an example with specific conditions on its growth which will serve as a contrast to the Theorem 4.3 as well:

Theorem 4.2. *For any numbers σ and ε satisfying $0 < \sigma < \sigma + \varepsilon < 1$, there exist two constants $N = N(\sigma, \varepsilon)$, $\mu = \mu(\sigma, \varepsilon)$ and a set $A \subset \mathbb{N}$ such that*

$$n^{\sigma} \leqslant |A[n]| \leqslant n^{\sigma+\varepsilon}$$

for each $n \geqslant N$ and

$$\frac{|A[n] + A[n]|}{|A[n]|} < \mu$$

for each $n \geqslant 1$.

We now consider sets with a stronger condition than bounded jumps, that is sets of *polynomial growth*. For these sets we give an estimate on their sumset depending on their growth. Note that in this case we are interested in the size of $(A + A)[n]$ rather than $A[n] + A[n]$:

Theorem 4.3. *Let A be an infinite subset of $\mathbb{N}$ such that*

$$0 < \liminf_{n\to\infty} |A[n]|n^{-\sigma} \leqslant \limsup_{n\to\infty} |A[n]|n^{-\sigma} < \infty. \tag{4.1}$$

for some $0 < \sigma < 1$. Then there is a positive constant $c(\sigma)$ such that

$$\frac{|(A + A)[n]|}{|A[n]|} > c(\sigma)\frac{\log n}{(\log\log n)^3 \log\log\log n(\log\log\log\log n)^3}. \tag{4.2}$$

for each sufficiently large n.

The constant $c(\sigma)$ is given explicitly in (4.11). The obtained estimate is almost best possible:

Theorem 4.4. *For any $\varepsilon > 0$, there exists a set $A \subset \mathbb{N}$ satisfying (4.1) such that*

$$\frac{|(A+A)[n]|}{|A[n]|} < \varepsilon \log n \tag{4.3}$$

for each sufficiently large n.

We conjecture that the lower bound in (4.2) should be of the order $\log n$, in which case it would be sharp up to a constant in view of Theorem 4.4. This bound would follow from the conjectural optimal bounds in Freiman's theorem, i.e., if the bound (4.5) below (due to Konyagin [50]) could be replaced by $d(\alpha), C(\alpha) < c\alpha$. The latter bound in Freiman's theorem was conjectured by Ruzsa [70] (see also [60]), although the description of set would likely be a convex set of given volume and dimension, rather than generalized arithmetical progression, so some modifications to the proof would be necessary.

In the next section we shall remind the reader Freiman's theorem and give a simple auxiliary lemma. In Sections 3, 4 and 5 we prove the stated theorems.

4.2 Effective Freiman's theorem

Let $A \subset \mathbb{N}$ be a finite set. Assume that $\alpha \geqslant 1$ is a real number such that

$$|A+A| \leqslant \alpha |A|. \tag{4.4}$$

Freiman's theorem then asserts that there are constants $d(\alpha)$ and $C(\alpha)$ such that A is contained in a generalized arithmetical progression

$$\{b_0 + b_1 z_1 + \cdots + b_d z_d \mid z_i = 0, \ldots, \ell_i - 1 \text{ for } i = 1, \ldots, d\},$$

where $d \leqslant d(\alpha)$ and

$$\ell_1 \ell_2 \ldots \ell_d \leqslant |A| e^{C(\alpha)}.$$

Originally proved by Freiman [32], the theorem has been reproved (simplifying the arguments) by Ruzsa [70] (also see [6]) and improved a number of times by

Chang, Schoen, Sanders and Konyagin, as listed below:

$$\begin{aligned}
&d(\alpha), C(\alpha) < c\alpha^2(\log\alpha)^2, \quad [7]\\
&d(\alpha), C(\alpha) < \alpha^{1+c\log^{-1/2}\alpha}, \quad [75]\\
&d(\alpha), C(\alpha) < c\alpha\log^{c'}\alpha, \quad [72]\\
&d(\alpha), C(\alpha) < c\alpha(\log\alpha)^3\log\log\alpha(\log\log\log\alpha)^3. \quad [50]
\end{aligned} \tag{4.5}$$

We will use the Konyagin's to get an estimate in Theorem 4.3. We will also need the following lemma:

Lemma 4.5. *Let A be an infinite subset of $\mathbb{N}$ satisfying* (4.1). *Then there is a positive constant κ such that*

(*a*) $|A[2n]| \leqslant \kappa 2^\sigma |A[n]|$ *for each sufficiently large n, say $n \geqslant n_0$,*

(*b*) *each interval $(n, \kappa^{1/\sigma}n]$, where $n \geqslant n_0$, contains an element of A.*

Proof. By (4.1), there are two positive constants $c_1 < c_2$ and some real number $n_0 \geqslant 1$ such that

$$c_1 n^\sigma \leqslant |A[n]| \leqslant c_2 n^\sigma \tag{4.6}$$

for each $n \geqslant n_0$. Set

$$\kappa := c_2/c_1. \tag{4.7}$$

Then, for $n \geqslant n_0$, using (4.6), we obtain

$$|A[2n]| \leqslant c_2(2n)^\sigma = \kappa 2^\sigma c_1 n^\sigma \leqslant \kappa 2^\sigma |A[n]|.$$

To prove (*b*) assume that $A \cap (n, \kappa^{1/\sigma}n] = \emptyset$. Then $\kappa^{1/\sigma}n \notin A$, so there is a positive number ε such that the set $A \cap (n, \kappa^{1/\sigma}(n+\varepsilon)]$ is empty. This, combined with (4.6) and (4.7), implies

$$c_2 n^\sigma \geqslant |A[n]| = |A[\kappa^{1/\sigma}(n+\varepsilon)]| \geqslant c_1(\kappa^{1/\sigma}(n+\varepsilon))^\sigma = c_2(n+\varepsilon)^\sigma$$

for $n \geqslant n_0$, a contradiction. □

4.3 Proofs of Theorem 4.1 and Theorem 4.3

We will use a non-effective version of Freiman's theorem for the proof of Theorem 4.1, as the result we are aiming for is qualitative.

Proof of Theorem 4.1. We argue by contradiction and assume that there exists a set $A = \{a_1, a_2, \dots\}$ satisfying conditions

$$\liminf_{n\to\infty} \frac{|A[n]|}{n} = 0, \tag{4.8}$$

$$\forall i \in \mathbb{N} \quad \frac{a_{i+1}}{a_i} \leqslant c_1, \tag{4.9}$$

$$\frac{|A[n] + A[n]|}{|A[n]|} \leqslant c_2. \tag{4.10}$$

From Freiman's theorem it follows that for each n there exists a generalized arithmetic progression

$$G = \{b_0 + b_1 z_1 + \dots + b_d z_d \mid z_i \in [\ell_i]\},$$

with $A[n] \subseteq G$, $d \leqslant d(c_1, c_2)$ and $\ell_1 \dots \ell_d = s|A[n]| \leqslant s(c_1, c_2)|A[n]|$. Without loss of generality assume that $b_1 \leqslant b_2 \leqslant \dots \leqslant b_d$ and note that b_0 and b_1 cannot be very large as G has to contain first two elements of A. More precisely, $b_0 \leqslant a_1$ and $b_1 \leqslant a_2 - a_1$. We are not interested in precise constants, so write $b_0 \ll 1$ and $b_1 \ll 1$ (as $n \to \infty$).

Let a_{-1} be the largest element of $A[n]$. We now show that either $b_2 \ll \ell_1$ or $a_{-1} \ll \ell_1$. Take an element $a_k \in A[n]$ such that $a_{k-1} \leqslant b_0 + b_1\ell_1 < a_k$ (if such element does not exist we are done). It has to satisfy $a_k \in G$, which implies $a_k \geqslant b_2 + b_0$. On the other hand, from condition 4.9 we know that $a_k \ll a_{k-1} \leqslant b_0 + b_1\ell_1$. Putting the inequalities together we get $b_2 \ll b_1\ell_1 \ll \ell_1$ as required.

Similarly we can prove that $\min\{a_{-1}, b_i\} \ll \ell_1 \cdots \ell_{i-1}$ for all $i \leqslant d$: argue by induction and take $a_k \in A[n]$ satisfying $a_{k-1} \leqslant b_0 + \ell_1 b_1 + \dots + \ell_{i-1} b_{i-1} < a_k$. From here we get $a_k \geqslant b_0 + b_i$ and $a_k \ll a_{k-1} \ll l_{i-1} b_{i-1} \ll \ell_{i-1} \cdots \ell_1$.

Using above inequalities and size estimate in the Freiman's theorem we can bound the size of a_{-1}. If at some induction step we obtained that $a_{-1} \ll \ell_1 \cdots \ell_k$, then we know that $a_{-1} \ll \ell_1 \cdots \ell_d$. Else the largest element of G (and so a_{-1}) is smaller than $b_0 + b_1\ell_1 + \dots + b_d\ell_d \ll \ell_1 \cdots \ell_d$ and hence in both cases we get that $a_{-1} \ll \ell_1 \cdots \ell_d \ll |A[n]|$.

Condition 4.9, on the other hand, implies that a_{-1} has to be greater or equal to n/c_1 (as the next element of A is greater than n), which gives $|A[n]| \gg n$ – a contradiction to 4.8. □

The proof of Theorem 4.3 follows essentially the same line but is more precise to get a quantitative bound.

Proof of Theorem 4.3. Fix a small constant $\delta > 0$ and take a positive constant

$$c(\sigma) := \sqrt{\frac{\sigma(1-\sigma)}{c(\sigma+\log\kappa)\kappa^2 4^{\sigma-1}}} - \delta, \tag{4.11}$$

where c is given in (4.5) and κ is given in (4.7).

Assume that $A \subset \mathbb{N}$ is a set satisfying (4.1) for which (4.2) does not hold. Then, for each $\varepsilon > 0$, there are infinitely many positive integers n for which

$$\frac{|(A+A)[n]|}{|A[2n]|} \leqslant (c(\sigma)+\varepsilon)\frac{\log n}{(\log\log n)^3 \log\log\log n (\log\log\log\log n)^3}.$$

Since the sumset $A[n]+A[n]$ is contained in $(A+A)[2n]$ we have

$$|A[n]+A[n]| \leqslant |(A+A)[2n]|.$$

So, by Lemma 4.5 (a),

$$\frac{|(A+A)[2n]|}{|A[2n]|} \geqslant \frac{|A[n]+A[n]|}{\kappa 2^\sigma |A[n]|}$$

for each $n \geqslant n_0$. It follows that

$$|A[n]+A[n]| \leqslant \kappa 2^\sigma (c(\sigma)+\varepsilon)\frac{\log n}{(\log\log n)^3 \log\log\log n (\log\log\log\log n)^3}|A[n]| \tag{4.12}$$

for infinitely many positive integers n.

Fix one of those n, where $n > n_1$ and n_1 will be chosen later. By Freiman's theorem bound (4.5), where $A := A \cap [1,n]$ and

$$\alpha := \kappa 2^\sigma (c(\sigma)+\varepsilon)\frac{\log n}{(\log\log n)^3 \log\log\log n (\log\log\log\log n)^3} > 1 \tag{4.13}$$

in (4.4), inequality (4.12) implies that the set $A[n]$ is contained in a d-dimensional arithmetical progression

$$P := \{b_0 + b_1 z_1 + \cdots + b_d z_d \mid z_i = 0, \ldots, \ell_i - 1 \text{ for } i = 1, \ldots, d\},$$

where $b_0 = b_0(n) \geqslant 0$, $b_1 = b_1(n), \ldots, b_d = b_d(n)$ are positive integers, and

$$\begin{aligned} d &\leqslant c\alpha(\log\alpha)^3 \log\log\alpha(\log\log\log\alpha)^3, \\ \ell_1\ell_2\ldots\ell_d &\leqslant |A[n]| e^{c\alpha(\log\alpha)^3\log\log\alpha(\log\log\log\alpha)^3}. \end{aligned} \tag{4.14}$$

Among all progressions with this property we choose one with the smallest possible d. Assume without loss of generality that $1 \leqslant b_1 \leqslant b_2 \leqslant \ldots \leqslant b_d$. By the minimality of d, we have $\ell_1, \ldots, \ell_d \geqslant 2$.

Let $s = s(n) \in \{1, \ldots, d\}$ be the smallest positive integer for which

$$b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_s - 1)b_s \geqslant n_0,$$

n_0 being the constant of Lemma 4.5. Clearly, $b_1, \ldots, b_{s-1}, \ell_1, \ldots, \ell_{s-1} < n_0 + 1$, because

$$b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_{s-1} - 1)b_{s-1} < n_0.$$

Since $A[n] \subseteq P$, for the set $A = \{a_1 < a_2 < a_3 < \ldots\}$, we have $b_0 \leqslant a_1$ and $b_0 + b_1 \leqslant a_2$. So both $b_0 = b_0(n)$ and $b_1 = b_1(n)$ cannot tend to infinity with n. Thus $b_0, b_1, \ldots, b_{s-1}, \ell_1, \ldots, \ell_{s-1}$ (and, in addition, b_1 if $s = 1$) are bounded from above by an absolute constant $c_4 = c_4(A)$.

We claim that $b_s = b_s(n)$ is also bounded by an absolute constant. Indeed, for $s = 1$, this is already proved. Suppose that $s \geqslant 2$ and $b_s(n) > n_0$ (otherwise there is nothing to prove). Then the number $b_0 + (j_1 - 1)b_1 + \cdots + (j_{s-1} - 1)b_{s-1}$ for each collection $j_1, \ldots, j_{s-1}$ satisfying $1 \leqslant j_i \leqslant l_i$ $(i = 1, \ldots, s-1)$ is smaller than n_0 and so smaller than $b_0 + b_s$. Hence

$$b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_{s-1} - 1)b_{s-1} \leqslant a_{\ell_1 \ell_2 \ldots \ell_{s-1}}$$

and

$$b_0 + b_s \leqslant a_{\ell_1 \ell_2 \ldots \ell_{s-1}+1}.$$

It follows that $b_s = b_s(n) \leqslant a_r$ with $r := (\lfloor n_0 \rfloor + 2)^{\lfloor n_0 \rfloor + 1} + 1$, because

$$\ell_1, \ldots, \ell_{s-1} < n_0 + 1 < \lfloor n_0 \rfloor + 2$$

and

$$s - 1 \leqslant \ell_1 - 1 + \cdots + \ell_{s-1} - 1 < n_0 < \lfloor n_0 \rfloor + 1.$$

Put $c_3 := \max(c_4, a_r)$.

Suppose first that $s = 1$. Then, by Lemma 4.5 (b), the interval $(b_0 + (\ell_1 - 1)b_1, \kappa^{1/\sigma}(b_0 + (\ell_1 - 1)b_1)]$ contains an element of A. Thus

$$b_0 + b_2 \leqslant \kappa^{1/\sigma}(b_0 + (\ell_1 - 1)b_1) \leqslant \kappa^{1/\sigma} c_3 \ell_1.$$

This yields $b_2 \leqslant \kappa^{1/\sigma} c_3 \ell_1$. By the same argument, using the inequalities $b_0 + (\ell_1 - 1)b_1 \leqslant c_3\ell_1$, $\kappa^{1/\sigma} \geqslant 1$ and Lemma 4.5 (b), we obtain

$$b_0+b_3 \leqslant \kappa^{1/\sigma}(b_0+(\ell_1-1)b_1+(\ell_2-1)b_2) \leqslant \kappa^{1/\sigma}(c_3\ell_1+(\ell_2-1)\kappa^{1/\sigma}c_3\ell_1) \leqslant \kappa^{2/\sigma}c_3\ell_2\ell_1.$$

Hence $b_3 \leqslant \kappa^{2/\sigma} c_3\ell_2\ell_1$ and so on, i.e.,

$$b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_k - 1)b_k \leqslant \kappa^{k/\sigma} c_3 \ell_k \dots \ell_1 \tag{4.15}$$

for $k = 1, \dots, d-1$, giving

$$b_k \leqslant \kappa^{(k-1)/\sigma} c_3 \ell_{k-1} \dots \ell_1 \tag{4.16}$$

for every $k = 2, 3, \dots, d$.

In case $s \geqslant 2$, we start with the interval

$$(b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_s - 1)b_s, \kappa^{1/\sigma}(b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_s - 1)b_s)].$$

At the first step, we get

$$\begin{aligned} b_0 + b_{s+1} &\leqslant \kappa^{1/\sigma}(b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_{s-1} - 1)b_{s-1} + (\ell_s - 1)b_s) \\ &\leqslant \kappa^{1/\sigma}(n_0 + (\ell_s - 1)b_s) \leqslant \kappa^{1/\sigma}(c_3 + (\ell_s - 1)c_3) = \kappa^{1/\sigma} c_3 \ell_s, \end{aligned}$$

because $n_0, b_s \leqslant c_3$. This yields $b_{s+1} \leqslant \kappa^{1/\sigma} c_3 \ell_s$. Continuing as above, we use less steps in our inductive argument, so the inequalities (4.15) and (4.16) are also true for $k = s, \dots, d-1$ and for $k = s+1, \dots, d$, respectively.

Now, from (4.15) with $k = d-1$ and (4.16) with $k = d$, we deduce that the largest element of the progression P is

$$\begin{aligned} b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_d - 1)b_d &\leqslant \kappa^{(d-1)/\sigma} c_3 \ell_{d-1} \dots \ell_1 + \kappa^{(d-1)/\sigma} c_3 \ell_{d-1} \dots \ell_1 (\ell_d - 1) \\ &= \kappa^{(d-1)/\sigma} c_3 \ell_d \dots \ell_1. \end{aligned}$$

Recall that $A[n] \subseteq P$ and the interval $(\kappa^{-1/\sigma} n, n]$ contains an element of A for $n \geqslant \kappa^{1/\sigma} n_0$, by Lemma 4.5 (b). Hence the upper bound on the largest element of P gives

$$\begin{aligned} \kappa^{-1/\sigma} n &\leqslant b_0 + (\ell_1 - 1)b_1 + \cdots + (\ell_d - 1)b_d \\ &\leqslant \kappa^{(d-1)/\sigma} c_3 \ell_d \dots \ell_1 \\ &\leqslant \kappa^{(d-1)/\sigma} c_3 |A[n]| e^{c\alpha(\log \alpha)^3 \log\log\alpha (\log\log\log\alpha)^3}, \end{aligned}$$

by (4.14). Thus

$$n \leqslant \kappa^{d/\sigma} c_3 |A[n]| e^{c\alpha(\log\alpha)^3 \log\log\alpha(\log\log\log\alpha)^3}.$$

Using the upper bounds on d, $d \leqslant c\alpha(\log\alpha)^3 \log\log\alpha(\log\log\log\alpha)^3$, and on $A[n]$, $|A[n]| \leqslant c_2 n^\sigma$ for $n \geqslant n_0$, we further get

$$\log n \leqslant \log|A[n]| + \log c_3 + c\Big(1 + \frac{\log\kappa}{\sigma}\Big)\alpha(\log\alpha)^3 \log\log\alpha(\log\log\log\alpha)^3$$

$$\leqslant \sigma \log n + \log(c_2 c_3) + c\Big(1 + \frac{\log\kappa}{\sigma}\Big)\alpha(\log\alpha)^3 \log\log\alpha(\log\log\log\alpha)^3.$$

Employing (4.13), for each sufficiently large n, say $n > n_2$, we have $\alpha > 1$ and

$$\alpha(\log\alpha)^3 \log\log\alpha(\log\log\log\alpha)^3 < \kappa 2^{\sigma-1}(c(\sigma) + 2\varepsilon)\log n.$$

Hence

$$1 - \sigma < \frac{\log(c_2 c_3)}{\log n} + c\Big(1 + \frac{\log\kappa}{\sigma}\Big)(\kappa 2^{\sigma-1}(c(\sigma) + 2\varepsilon))^2. \tag{4.17}$$

Evidently, $\log(c_2 c_3)/\log n \to 0$ as $n \to \infty$. Thus, by the choice of $c(\sigma)$ in (4.11), we see that inequality (4.17) does not hold provided that ε is small enough and n is large enough, say $n > n_3$. Selecting $n_1 := \max(\kappa^{1/\sigma} n_0, n_2, n_3)$, we have a contradiction. This completes the proof of the theorem. □

4.4 Proof of Theorem 4.4

Fix a large integer N and select

$$A := \cup_{j=0}^{\infty} B_j,$$

where

$$B_j := \{N^j, N^j + 1, \ldots, N^j + \lfloor N^{\sigma j} \rfloor - 1\}.$$

For each $k \in \mathbb{N}$ and each real number $n \in [N^{k-1}, N^k)$, using $|B_j| = [N^{\sigma j}] \leqslant N^{\sigma j}$, we obtain

$$|A[n]| \leqslant \sum_{j=0}^{k-1} |B_j| \leqslant \sum_{j=0}^{k-1} N^{\sigma j} = \frac{N^{k\sigma} - 1}{N^\sigma - 1} < N^{(k-1)\sigma} \frac{N^\sigma}{N^\sigma - 1} \leqslant n^\sigma \frac{N^\sigma}{N^\sigma - 1}.$$

Similarly, for $k \geqslant 2$ and $n \in [N^{k-1}, N^k)$, we have

$$|A[n]| \geqslant \sum_{j=0}^{k-2} |B_j| \geqslant |B_{k-2}| > N^{(k-2)\sigma} - 1 \geqslant N^{k\sigma} \frac{1}{N^2} > n^\sigma \frac{1}{N^2}$$

for N large enough, so A satisfies (4.1) for each fixed large integer N.

It is sufficient to prove (4.3) for each sufficiently large integer n, so in all what follows we assume that n is an integer. By the choice of A, it is clear that, for every $k \in \mathbb{N}$,

$$|(A+A)[N^k]| \leqslant \sum_{0 \leqslant i \leqslant j \leqslant k-1} |B_i + B_j| \leqslant \sum_{0 \leqslant i \leqslant j \leqslant k-1} (|B_i| + |B_j|) \tag{4.18}$$
$$= k \sum_{j=0}^{k-1} |B_j| = k(|A[N^k]| - 1).$$

Observe that, for $n \in [N^{k-1}, N^{k-1} + [N^{\sigma(k-1)}] - 1]$,

$$\frac{|(A+A)[n]|}{|A[n]|} = \frac{|(A+A)[n]|}{|A[N^{k-1}]| + n - N^{k-1}}$$
$$\leqslant \frac{|(A+A)[N^{k-1}]| + n - N^{k-1}}{|A[N^{k-1}]| + n - N^{k-1}} \leqslant \frac{|(A+A)[N^{k-1}]|}{|A[N^{k-1}]|},$$

because $|(A+A)[N^{k-1}]| > |A[N^{k-1}]|$ for each sufficiently large N. So, by (4.18),

$$\frac{|(A+A)[n]|}{|A[n]|} \leqslant \frac{|(A+A)[N^{k-1}]|}{|A[N^{k-1}]|} \leqslant k - 1 \leqslant \frac{\log x}{\log N}.$$

Alternatively, if $n \in (N^{k-1} + [N^{\sigma(k-1)}] - 1, N^k]$, then, by (4.18) again,

$$\frac{|(A+A)[n]|}{|A[n]|} = \frac{|(A+A)[n]|}{|A[N^k]| - 1} \leqslant \frac{|(A+A)[N^k]|}{|A[N^k]| - 1} \leqslant k \leqslant 1 + \frac{\log n}{\log N}.$$

In both cases, for any $\varepsilon > 0$, taking some integer $N > N(\varepsilon) := [e^{1/\varepsilon}] + 1$ we deduce that

$$\frac{|(A+A)[n]|}{|A[n]|} \leqslant 1 + \frac{\log n}{\log N} < \varepsilon \log n$$

for all sufficiently large n.

4.5 Proof of Theorem 4.2

For $\gamma > 1$ and $0 < \beta < 1$, let us take $A := \cup_{j=0}^{\infty} C_j$, where

$$C_j := \{\lfloor 2^{\gamma^j} \rfloor, \lfloor 2^{\gamma^j} \rfloor + 1, \ldots, \lfloor 2^{\gamma^j} \rfloor + \lfloor 2^{\gamma^j \beta} \rfloor - 1\}.$$

Take $k \in \mathbb{N}$ for which $\lfloor 2^{\gamma^{k-1}} \rfloor \leqslant n < \lfloor 2^{\gamma^k} \rfloor$. Then

$$|A[n]| \leqslant \sum_{j=0}^{k-1} |C_j| = \lfloor 2^{\gamma^{k-1}\beta} \rfloor + \lfloor 2^{\gamma^{k-2}\beta} \rfloor + \cdots + 1 \leqslant 2^{\gamma^{k-1}\beta} + 2^{\gamma^{k-2}\beta} + \cdots + 1.$$

This is smaller than $2^{\gamma^k\beta}$ for k large enough. As $2^{\gamma^{k-2}} < n$, we get

$$|A[n]| < 2^{\gamma^k\beta} < n^{\beta\gamma^2}.$$

Similarly, for $k \geqslant 2$, we obtain

$$|A[n]| \geqslant |C_{k-2}| = \lfloor 2^{\gamma^{k-2}\beta} \rfloor \geqslant n^{\beta\gamma^{-3}}$$

provided that n is large enough. It follows that A satisfies $n^\sigma \leqslant |A[n]| \leqslant n^{\sigma+\varepsilon}$ for a suitable choice of γ and β, e.g., $\beta^5 = \sigma^2(\sigma+\varepsilon)^3$ and $\gamma^5 = 1 + \varepsilon/\sigma$ and n large enough.

On the other hand, observe that, for every $k \in \mathbb{N}$,

$$\lfloor 2^{\gamma^k} \rfloor + \lfloor 2^{\gamma^k\beta} \rfloor \geqslant \lfloor 2^{\gamma^k} \rfloor + \lfloor 2^{\gamma^j} \rfloor$$

whenever $k - j \geqslant t := \lfloor \log(\beta^{-1})/\log\gamma \rfloor + 1$. Hence

$$\cup_{j=0}^{k-t}(C_k + C_j) \subseteq (C_k + C_0) \cup (C_k + C_{k-t}).$$

It follows that

$$|A[\lfloor 2^{\gamma^k} \rfloor] + A[\lfloor 2^{\gamma^k} \rfloor]| \leqslant \sum_{i=0}^{k-1} |\cup_{j=0}^{i} (C_i + C_j)|$$

$$\leqslant \sum_{i=0}^{t} |\cup_{j=0}^{i} (C_i + C_j)| + \sum_{i=t+1}^{k-1} |(C_i + C_0) \cup (C_i + C_{i-t}) \cup_{j=i-t+1}^{i} (C_i + C_j)|.$$

Since $|C_j| \leqslant |C_i|$ whenever $j \leqslant i$, the first sum is bounded from above by

$$\sum_{i=0}^{t}\sum_{j=0}^{i} |C_i + C_j| \leqslant \sum_{i=0}^{t}\sum_{j=0}^{i} 2|C_i| = \sum_{i=0}^{t} 2(i+1)|C_i| = (2t+2)\sum_{i=0}^{t} |C_i|.$$

The second sum is bounded from above by

$$\sum_{i=t+1}^{k-1} (|C_i| + |C_0| + |C_i| + |C_{i-t}| + 2t|C_i|) \leqslant (2t+4)\sum_{i=t+1}^{k-1} |C_i|.$$

Adding these two inequalities we derive that

$$|A[\lfloor 2^{\gamma^k} \rfloor] + A[\lfloor 2^{\gamma^k} \rfloor]| \leqslant (2t+4)\sum_{i=0}^{k-1} |C_i| \leqslant (2t+5)(|A[\lfloor 2^{\gamma^k} \rfloor]|) - 1) \qquad (4.19)$$

for k large enough, because the quotient $|A[\lfloor 2^{\gamma^k} \rfloor]|/\sum_{i=0}^{k-1} |C_i|$ tends to 1 as k tends to infinity.

Note that (4.19) is an analogue of (4.18). Arguing as in proof of Theorem 4.4 one easily derives that for n satisfying $\lfloor 2^{\gamma^{k-1}} \rfloor \leqslant n < \lfloor 2^{\gamma^k} \rfloor$ the quotient $|A[n] + A[n]|/|A[n]|$ is bounded from above by the larger of

$$\frac{|A[\lfloor 2^{\gamma^{k-1}} \rfloor] + A[\lfloor 2^{\gamma^{k-1}} \rfloor]|}{|A[\lfloor 2^{\gamma^{k-1}} \rfloor]|}$$

and

$$\frac{|A[\lfloor 2^{\gamma^{k}} \rfloor] + A[\lfloor 2^{\gamma^{k}} \rfloor]|}{(|A[\lfloor 2^{\gamma^{k}} \rfloor]| - 1)}$$

whenever n is large enough. Thus $|A[n]+A[n]|/|A[n]| \leqslant 2t+5$ for each sufficiently large n, by (4.19). This clearly implies that $|A[n] + A[n]|/|A[n]|$ is bounded from above by an absolute constant $\mu = \mu(\sigma, \varepsilon)$ for each $n \geqslant 1$.

5 B_h Sequences in Higher Dimensions

We start this chapter with a version of a simple and elegant theorem proved by Ruiz and Trujillo [69], the main idea of which goes back to Erdős and Turán [30]:

Theorem 5.1. *Let $\mathbb{F}_p$ be a finite field with $p \geqslant h$. Then the set*

$$A := \left\{(x, x^2, \ldots, x^h), x \in \mathbb{F}_p\right\}$$

is a B_h set on $(\mathbb{F}_p^h, +)$.

As it is not unusual, proving that a given set is Sidon (or B_h in this case) is simpler than coming up with the set in the first place[1], so we will only discuss the result itself.

The first thing to notice is that this theorem gives correct asymptotic size (up to constant) for a maximal size of B_h in a group $(\mathbb{F}_p^h, +)$ and, via appropriate map, in the interval $[n]$ of integers. Indeed, for any B_h set A in a finite group G all the $\binom{h}{|A|}$ sums of h different elements have to be different, so $|A| \ll |G|^{1/h}$, which is the size of a set A in the Theorem 5.1.

Secondly, while construction is very elegant and dense, it is not clear, if it is the densest possible for all h, except the Sidon case $h = 2$. In this case by calculating number of differences between different elements of A, which for a Sidon set also have to be distinct, we get that $|A|(|A| - 1) \leqslant |G| - 1$, which gives that the $|A| = p$ is indeed the maximal cardinality of a Sidon set in $G = (\mathbb{F}_p^2, +)$. In other cases $h > 2$ more clever counting arguments can reduce the constant obtained in counting different sums, but they do not give constant equal to 1. Obtaining a good upper bound will be our main concern in this chapter.

[1] Curiously, essentially this theorem rephrased as a system of equations was given as a problem #2 in Vilnius University Mathematical Olympiad in 2008, five years before the manuscript of Ruiz and Trujillo.

Finally, this construction is a nice illustration of importance of *dimension* in the study of B_h sets. Most, if not all, very dense constructions of B_h sequences make use of it, see [69], [15] and [16] for more examples.

5.1 The problem and results

We turn our attention from finite groups to a set of positive integers – a setting, in which the original question of Sidon was formulated. Our first question is its following generalisation: let A be an infinite B_h subset (or sequence) of $\mathbb{N}^d$, and let $|A[n]^d|$ denote the cardinality of $A \cap [n]^d$. Is it possible that $|A[n]^d| \gg n^{d/2}$? We are only able to give an answer in the case of even h, and it is negative, as in the one-dimensional case:

Theorem 5.2. *If $A \subseteq \mathbb{N}^d$ is a B_{2k} sequence, then*

$$\liminf_{n\to\infty} |A[n]^d| \frac{\log^{1/2k} n}{n^{d/k}} < \infty.$$

While no dense infinite B_h sequences for even h exist, it is possible (as we saw in the introduction of this chapter) to construct a finite B_h sets, that is subsets of $[n]^d$. An upper bound on how large such sets can be is naturally of interest, and our next two theorems extend known results (which were mentioned in Section 2.1) to the d-dimensional case.

Theorem 5.3. *If $A \subseteq [n]^d$ is a B_{2k} set, then*

$$|A| \leqslant (k!)^{\frac{1}{k}} k^{\frac{d}{2k}} n^{\frac{d}{2k}} + O\big(n^{\frac{d^2}{2k(d+1)}}\big).$$

Theorem 5.4. *If $A \subseteq [n]^d$ is a B_{2k-1} set, then*

$$|A| \leqslant (k!)^{\frac{2}{2k-1}} k^{\frac{d-1}{2k-1}} n^{\frac{d}{2k-1}} + O\big(n^{\frac{d^2}{(d+1)(2k-1)}})\big).$$

Finally, for large h it is possible to improve above bounds using Fourier analytical techniques developed in [37]:

Theorem 5.5. *If $A \subseteq [n]^d$ is a B_{2k} set and k is large enough, then*

$$|A| \leqslant (\pi d)^{\frac{d}{4k}} (1+\epsilon(k)) k^{\frac{d}{4k}} (k!)^{\frac{1}{k}} n^{\frac{d}{2k}} + O\big(n^{\frac{d^2}{2k(d+1)}}\big).$$

Theorem 5.6. *If $A \subseteq [n]^d$ is a B_{2k-1} set and k is large enough, then*

$$|A| \leqslant (\pi d)^{\frac{d}{2(2k-1)}}(1+\epsilon(k))k^{\frac{d-2}{2(2k-1)}}(k!)^{\frac{2}{2k-1}}n^{\frac{d}{2k-1}} + O\big(n^{\frac{d^2}{(2k-1)(d+1)}}\big).$$

We continue this chapter with the Preliminaries section, where we prove several lemmas used later. In the subsequent sections we prove above theorems.

5.2 Preliminaries

We start from the following very simple lemma:

Lemma 5.7. *Let functions $f, g : \mathbb{Z} \to \mathbb{R}$ be finitely supported. Then*

$$\sum_x f * g(x)^2 = \sum_x f \circ g(x)^2 = \sum_x f \circ f(x) g \circ g(x).$$

Proof. All three sums are equal to summation of $f(i)g(j)f(i')g(j')$ over quadruples (i,j,i',j') satisfying $i+j = i'+j'$ and correspond to different rearrangements of the equality as $i - j' = i' - j$ and $i - i' = j - j'$. □

It implies the following very useful estimate:

Lemma 5.8. *For any subsets A, B of some additive semigroup G we have*

$$\sum_x A \circ A(x) B \circ B(x) \geqslant \frac{|A|^2|B|^2}{|A+B|}. \tag{5.1}$$

We deduce it from a more general following result, by taking f, g to be indicators of A and B:

Lemma 5.9. *Let functions $f, g : G \to \mathbb{R}$ be defined on some additive semigroup and finitely supported, then*

$$\sum_x f \circ f(x) g \circ g(x) \geqslant \frac{(\sum_x f(x))^2 (\sum_x g(x))^2}{|\operatorname{supp}(f*g)|}.$$

Proof. Start from the equality $\sum_x f*g(x)^2 = \sum_x f\circ f(x) g\circ g(x)$ of Lemma 5.7 and use Cauchy-Schwarz inequality:

$$\sum_x f*g(x)^2 \geqslant \frac{(\sum_x f*g(x))^2}{|\operatorname{supp}(f*g)|} = \frac{(\sum_x f(x))^2(\sum_x g(x))^2}{|\operatorname{supp}(f*g)|}.$$

□

This general result can be used to prove a well known van der Corput lemma, which was used by Lindström [55], [54] to get a best known bound for a maximal densities of Sidon and B_4 sets. Even though this is not a very suitable place for such a detour, we cannot resist stating and giving a short proof of this lemma:

Lemma 5.10 (van der Corput)**.** *Let function $f : \mathbb{Z} \to \mathbb{R}$ be supported on the interval $[-n, n]$, then for any positive integer ℓ we have*

$$\Big(\sum_x f(x)\Big)^2 \leqslant \frac{2n+\ell}{\ell^2} \sum_{i=-\ell}^{\ell} (\ell - |i|) \sum_x f(x) f(x+i).$$

Proof. Start from inequality of lemma 5.9 and take g to be an indicator function of an interval $[\ell]$. Then $g\circ g(x)$ is zero outside $[-\ell, \ell]$ and equal to $i - |\ell|$ for any $i \in [-\ell, \ell]$, so the left hand side becomes

$$\sum_x f\circ f(x) g\circ g(x) = \sum_{i=-\ell}^{\ell} (\ell - |i|) f\circ f(i) = \sum_{i=-\ell}^{\ell} (\ell - |i|) \sum_x f(x) f(x+i).$$

It remains to note that for such g we have $\sum_x g(x) = \ell$ and $\operatorname{supp}(f*g) \subseteq [-n+1, n+\ell]$. □

If one wishes to apply this lemma to get an upper bound for density of Sidon set in $[n]$, it is sufficient to take an indicator $f(x) = A(x)$ and note that $\sum_x f(x) f(x+i) \leqslant 1$ for all i. It remains to chose an optimal ℓ which is easily done. It was noted by Cilleruelo [15] that correct choice gives a bound $|A| \leqslant n^{1/2} + n^{1/4} + 1/2$ – an improvement of $1/2$ over the original Lindsröm bound.

We return to our present matters and prove two more lemmas, which will enable us to use lemma 5.8 in a more general case. For any $x = x_1 + \cdots + x_r \in rA$, we let $\overline{x}$ be the multiset (i.e. set with multiple entries) $\{x_1, \dots, x_r\}$. For a B_h-set $A \subseteq [n]^d$ we define

$$D_j(z; r) \;=\; \{(x, y) : x - y = z, x, y \in jA, |\overline{x} \cap \overline{y}| = r\},$$

and write $d_j(z;r)$ for its cardinality.

Lemma 5.11. *Let $A \subseteq [n]^d$.*

(i) If A is a B_{2k} sequence, for $1 \leqslant j \leqslant k$,

$$d_j(z;0) \leqslant 1;$$

(ii) If A is B_{2k} sequence, for $1 \leqslant r \leqslant k$,

$$\sum_{z \in \mathbb{Z}^d} d_k(z;r) \leqslant |A|^{2k-r}.$$

Proof.

(i) If $(x,y),(x',y') \in D_j(z;0)$ then we have $x + y' = x' + y$. Since A is a B_h sequence, the two representations correspond to different permutations of the same h elements and as $\overline{x} \cap \overline{y} = \overline{x'} \cap \overline{y'} = \emptyset$, then $x = x'$ and $y = y'$.

(ii) There are at most $|A|^r$ possible values for $\overline{x} \cap \overline{y}$ (where the intersection is taken with multiplicities), so

$$d_k(z;r) \leq |A|^r d_{k-r}(z;0).$$

Then

$$\begin{aligned} \sum_{z \in \mathbb{Z}^d} d_k(z;r) &\leqslant |A|^r \sum_{z \in \mathbb{Z}^d} d_{k-r}(z;0) \\ &\leqslant |A|^r |(k-r)A|^2 \qquad \text{(using (i))} \\ &\leqslant |A|^{2k-r}. \end{aligned}$$

□

Similarly for a B_h-sequence $A \subseteq [n]^d$ we define

$$\begin{aligned} D_j^*(z;r) &= \{(x,y) : x - y = z, x, y \in j_*A, |\overline{x} \cap \overline{y}| = r\}, \\ D_j^*(z;r;a) &= \{(x,y) \in D_j^*(z,r) : a \in \overline{x}\} \end{aligned}$$

and write $d_j^*(z;r)$ and $d_j^*(z;r;a)$ for their respective cardinalities.

Lemma 5.12. *Let $A \subseteq [n]^d$.*

(i) If A is a B_{2k-1} sequence, for $1 \leqslant j \leqslant k-1$,

$$d_j^*(z;0) \leqslant 1;$$

(ii) If A is a B_{2k-1} sequence,

$$d_k^*(z;0) \leqslant \frac{|A|}{k}.$$

(iii) If A is a B_{2k-1} sequence, for $1 \leqslant r \leqslant k$,

$$\sum_{z\in\mathbb{Z}^d} d_k^*(z;r) \leqslant |A|^{2k-r}.$$

Proof.

(i) We may use the same proof as in (i) previous lemma.

(ii) We show that $d_k^*(z;0;a) \leqslant 1$. Assume not. Then there exists $x = x_1 + \ldots + x_k, x' = x'_1 + \ldots + x'_k, y = y_1 + \ldots + y_k, y' = y'_1 + \ldots + y'_k \in k_*A$ such that $x - y = x' - y' = z$. In addition, without loss of generality, we may assume $x_k = x'_k = a$. Hence we have

$$x_1 + \ldots + x_{k-1} + y'_1 + \ldots y'_k = x'_1 + \ldots + x'_{k-1} + y_1 + \ldots y_k.$$

Once again, since A is a B_{2k-1} sequence, the two representations correspond to different permutations of the same $2k-1$ elements and as $\overline{x}\cap\overline{y} = \overline{x}\cap\overline{y} = \emptyset$ we must have $x = x'$ and $y = y'$, giving a contradiction.

Notice that

$$\sum_{a\in A} d_k^*(z;0;a) = kd_k^*(z;0)$$

and the statement of the lemma follows.

(iii) We may use the same proof as in (ii) in previous lemma.

□

5.3 Proof of Theorem 5.2.

We fix a large enough positive integer n and set $u = \lfloor n^{1/(2k-1)} \rfloor$. For any d-dimensional vector $\vec{i}$ use the L_∞ norm defined as follows:

$$|\vec{i}|_\infty = |(i_1, i_2, ..., i_d)|_\infty = \max_{1\leqslant k\leqslant d}\{|i_k|\}.$$

For any d-dimensional set B denote

$$B_{\vec{i}} = B \cap \bigotimes_{j=1}^{d}((i_j - 1)kn, i_j kn].$$

We set

$$\begin{aligned} A' &= A \cap [1, un]^d, \\ C &= kA', \\ c_{\vec{i}} &= |C_{\vec{i}}|, \\ \Delta_j &= \sum_{|\vec{i}|_\infty = j} c_{\vec{i}}, \\ \tau(n) &= \min_{n \leqslant m \leqslant un} \frac{|A[m]^d|}{m^{d/2k}}. \end{aligned}$$

Lemma 5.13.

$$\tau(n)^{2k} n^d \log n = O\left(\sum_{\vec{i} \in [1,u]^d} c_{\vec{i}}^2\right).$$

Proof. Note that

$$\begin{aligned} \left(\sum_{\vec{i}\in[1,u]^d} \frac{c_{\vec{i}}}{|\vec{i}|_\infty^{d/2}}\right)^2 &\leqslant \left(\sum_{\vec{i}\in[1,u]^d} \frac{1}{|\vec{i}|_\infty^{d}}\right)\left(\sum_{\vec{i}\in[1,u]^d} c_{\vec{i}}^2\right) \\ &\leqslant \left(\sum_{i=1}^{u} \frac{di^{d-1}}{i^d}\right)\left(\sum_{\vec{i}\in[1,u]^d} c_{\vec{i}}^2\right) \\ &\leqslant O\left(\log n \sum_{\vec{i}\in[1,u]^d} c_{\vec{i}}^2\right). \end{aligned} \tag{5.2}$$

On the other hand, for any positive i ($1 \leqslant i \leqslant u$) we have,

$$|C[ikn]^d| \geqslant c|A[in]^d|^k,$$

where $c > 0$ is an absolute constant depending only on k, and

$$\begin{aligned} |A[in]^d|^k &= \left(\frac{|A[in]^d|}{(in)^{d/2k}}\right)^k (in)^{d/2} \\ &\geqslant \tau(n)^k (in)^{d/2}. \end{aligned}$$

Hence, for absolute constants c_1, c_2, c_3 depending on d and k,

$$\begin{aligned}
\sum_{\vec{i}\in[1,u]^d} \frac{c_{\vec{i}}}{|\vec{i}|_\infty^{d/2}} &= \sum_{i=1}^{u} \frac{\Delta_i}{i^{d/2}} \\
&= \sum_{i=1}^{u} \left(\frac{1}{i^{d/2}} - \frac{1}{(i+1)^{d/2}} \right) \sum_{j=1}^{i} \Delta_j + \frac{1}{(u+1)^{d/2}} \sum_{j=1}^{u} \Delta_j \\
&\geqslant c_1 \sum_{i=1}^{u} \frac{|C[ikn]^d|}{i^{d/2+1}} \\
&\geqslant c_2 \sum_{i=1}^{u} \frac{\tau(n)^k (in)^{d/2}}{i^{d/2+1}} \\
&= c_2 \tau(n)^k n^{d/2} \sum_{i=1}^{u} \frac{1}{i} \\
&\geqslant c_3 \tau(n)^k n^{d/2} \log n. \qquad (5.3)
\end{aligned}$$

Combining inequalities (5.2) and (5.3), Lemma 5.13 follows. □

Lemma 5.14.

$$\sum_{\vec{i}\in[1,u]^d} c_{\vec{i}}^2 = O(n^d).$$

Proof. We have

$$\begin{aligned}
\sum_{\vec{i}\in[1,u]^d} c_{\vec{i}}^2 &\leq \sum_{r=0}^{k} \sum_{|z|_\infty \leq kn} d_k(z;r) \\
&= \sum_{|z|_\infty \leqslant kn} d_k(z;0) + \sum_{r=1}^{k} \sum_{|z|_\infty \leqslant kn} d_k(z;r) \\
&\leqslant \sum_{|z|_\infty \leqslant kn} 1 + \sum_{r=1}^{k} |A'|^{2k-r} \quad \text{(using Lemma 5.11 (i) and (iv))} \\
&= (2kn)^d + O\left((un)^{d(1-1/(2k))}\right) \\
&= O(n^d).
\end{aligned}$$

□

We are now able to prove Theorem 5.2:

Proof of Theorem 5.2. From Lemmas 5.13 and 5.14 we have $\tau(n)^{2k} \log n = O(1)$.

Hence,

$$\begin{aligned}\liminf_{n\to\infty} |A[n]^d| \sqrt[2k]{\frac{\log n}{n^d}} &= \lim_{n\to\infty} \inf_{n\leqslant m\leqslant un} |A[m]^d| \sqrt[2k]{\frac{\log m}{m^d}} \\ &\leqslant \lim_{n\to\infty} \inf_{n\leqslant m\leqslant un} \frac{|A[m]^d|}{m^{d/2k}} \sqrt[2k]{\log un} \\ &\leqslant 2 \lim_{n\to\infty} \tau(n) \sqrt[2k]{\log n} < \infty.\end{aligned}$$

□

5.4 Proofs of Theorems 5.3 and 5.4

We start from the case $h = 2k$ and first prove the following lemma:

Lemma 5.15. *For a finite B_{2k} set $A \subseteq \mathbb{Z}^d$*

$$\sum_x kA \circ kA(x)[\ell]^d \circ [\ell]^d(x) \leqslant \ell^{2d} + O(\ell^d |A|^{2k-1}).$$

Proof.

$$\begin{aligned}&\sum_x kA \circ kA(x)[\ell]^d \circ [\ell]^d(x) \\ &= \sum_x [\ell]^d \circ [\ell]^d(x) \sum_{r=0}^{k} d_k(x;r) \\ &= \sum_x [\ell]^d \circ [\ell]^d(x) d_k(x;0) + \sum_{r=1}^{k} \sum_x [\ell]^d \circ [\ell]^d(x) d_k(x;r) \\ &\leqslant \ell^{2d} + O(\ell^d |A|^{2k-1}). \quad \text{(using Lemma 5.11 (i) and (ii))}\end{aligned}$$

□

Proof of Theorem 5.3. We will use Lemma 5.8 with (semi)group $\mathbb{Z}^d$, and sets kA and $[\ell]^d$ (where the positive integer ℓ will be chosen later). Note that

$$\begin{aligned}|kA| &\geqslant \frac{1}{k!}|A|^k, \\ |[\ell]^d| &= \ell^d, \\ |kA + [\ell]^d| &\leqslant (kn + \ell)^d.\end{aligned}$$

Thus, using Lemmas 5.15 and 5.8, we have (after simplification)

$$\frac{|A|^{2k}\ell^d}{k!^2(kn+\ell)^d} \leqslant \ell^d + O\left(|A|^{2k-1}\right),$$

or

$$\begin{aligned} |A|^{2k} &\leqslant k!^2(kn+\ell)^d + O\left(\left(\frac{kn}{\ell}+1\right)^d |A|^{2k-1}\right) \\ &\leqslant k!^2(kn+\ell)^d + O\left(\left(\frac{kn}{\ell}+1\right)^d n^{\frac{(2k-1)d}{2k}}\right). \end{aligned}$$

To minimise the error term we need $\left(\frac{n}{\ell}\right)^d n^{\frac{(2k-1)d}{2k}} = \ell n^{d-1}$, so we take $\ell = n^{1-\frac{d}{(d+1)2k}}$ giving

$$\begin{aligned} |A|^{2k} &\leqslant k!^2 k^d n^d + O(n^{d-\frac{d}{(d+1)2k}}) \\ &\leqslant k!^2 k^d n^d (1 + O(n^{-\frac{d}{(d+1)2k}})). \end{aligned}$$

Taking $2k^{\text{th}}$ roots ends the proof. □

We now continue with the case $h = 2k-1$ and again start from a lemma:

Lemma 5.16. *For a finite B_{2k-1} set $A \subseteq \mathbb{Z}^d$ we have*

$$\sum_x k_* A \circ k_* A(x)[\ell]^d \circ [\ell]^d(x) \leqslant \frac{|A|}{k}\ell^{2d} + O(\ell^d |A|^{2k-1}).$$

Proof. The proof follows the same course as that of Lemma 5.15 except using Lemma 5.12 (i), (ii) and (iii) in the final step. □

Proof of Theorem 5.4. As before we make use of Lemma 5.8, taking sets k_*A and $[\ell]^d$ in the (semi)group $\mathbb{Z}^d$. We have

$$\begin{aligned} |k_* A| &\geqslant \frac{1}{k!}|A|^k(1 - \frac{c}{|A|}), \\ |[\ell]^d| &= \ell^d, \\ |k_* A + [\ell]^d| &\leqslant (kn+\ell)^d, \end{aligned}$$

where constant c depends only on k. Now Lemmas 5.16 and 5.8 give

$$\frac{(1-\frac{c}{|A|})^2 |A|^{2k}\ell^{2d}}{(k!)^2(kn+\ell)^d} \leqslant \ell^{2d}\frac{|A|}{k} + O(|A|^{2k-1}\ell^d),$$

or

$$\frac{|A|^{2k}\ell^{2d}}{(k!)^2(kn+\ell)^d} \leqslant \ell^{2d}\frac{|A|}{k} + O(|A|^{2k-1}\ell^d)$$

thus

$$\begin{aligned} |A|^{2k-1} &\leqslant \frac{(k!)^2(kn+\ell)^d}{k} + O\left(\left(\frac{kn}{\ell}+1\right)^d |A|^{2k-2}\right) \\ &\leqslant \frac{(k!)^2(kn+\ell)^d}{k} + O\left(\left(\frac{kn}{\ell}+1\right)^d n^{d\frac{2k-2}{2k-1}}\right). \end{aligned}$$

To minimise the error term we need $n^{d-1}\ell = n^d n^{d(2k-2)/(2k-1)}$ so we take $\ell = n^{1-\frac{d}{(d+1)(2k-1)}}$ which gives

$$\begin{aligned} |A|^{2k-1} &\leqslant (k!)^2 n^d k^{d-1} + O(n^{d-\frac{d}{(d+1)(2k-1)}}) \\ &\leqslant (k!)^2 n^d k^{d-1}(1 + O(n^{-\frac{d}{(d+1)(2k-1)}})). \end{aligned}$$

Taking $2k-1^{\text{th}}$ roots gives the result. □

Finite B_h sequences for large h

We start from a set of fairly standard Fourier analytic prerequisites. Let $\mathbb{Z}_n^d$ be a d-fold direct product of a ring of integers modulo n. For any two vectors (elements of the product ring) $a = (a_1, a_2, \ldots, a_d)$ and $b = (b_1, b_2, \ldots, b_d)$ define scalar product as

$$a \cdot b = \sum_{i=1}^{d} a_i b_i.$$

Ring $\mathbb{Z}_n^d$ with such scalar product is a Hilbert space, so for any function $f : \mathbb{Z}_N^d \to \mathbb{C}$ we define a Fourier transform

$$\hat{f}(r) = \sum_{x \in \mathbb{Z}_N^d} f(x) e^{\frac{2\pi i r \cdot x}{N}}, \quad \forall r \in \mathbb{Z}_n^d.$$

We extend our definition of difference convolution to complex valued functions f, g as

$$(f \circ g)(x) = \sum_i f(i)\overline{g(x+i)},$$

and assume right associativity for repeated convolutions, that is

$$f_1 \circ f_2 \circ \cdots \circ f_k = f_1 \circ (f_2 \circ \cdots \circ (f_{k-1} \circ f_k)).$$

We shall denote $f^{\circ 2k}(x) = (\underbrace{f \circ f \circ \cdots \circ f}_{2k \text{ times}})(x)$ and note that for an indicator function of a set A, $A^{\circ 2k}(x)$ is the number of ordered representations of $x = a_1 + \cdots + a_k - a_{k+1} - \cdots - a_{2k}$ for $a_1, a_2, \ldots, a_{2k} \in A$. We shall use the following two well-known identities:

Lemma 5.17 (Parseval's Identity)**.** *If $f, g : \mathbb{Z}_n^d \to \mathbb{C}$ are two functions then*

$$n^d \sum_{x \in \mathbb{Z}_n^d} f(x)\overline{g(x)} = \sum_{r \in \mathbb{Z}_n^d} \hat{f}(r)\overline{\hat{g}(r)}.$$

Lemma 5.18. *If $f, g : \mathbb{Z}_n^d \to \mathbb{C}$ are two functions then*

$$\widehat{(f \circ g)}(r) = \hat{f}(r)\overline{\hat{g}(r)}.$$

We are now ready to give proofs to the Theorems 5.5 and 5.6.

Proof of Theorem 5.5. We regard A as a subset of $\mathbb{Z}_{kn+v}^d$ where v is sufficiently small compared to n, so that $A^{\circ 2k}(x)$ remains the same for $x \in [-v, v]^d$ as it was when we regarded A as a subset of $\mathbb{Z}^d$. Similarly as in the earlier proofs, we will consider an indicator $[\ell]^d$, where ℓ we be chosen asymptotically much smaller than v.

Notice that, for all $x \in [-v, v]^d$, $A^{\circ 2k}(x) \leqslant (k!)^2 kA \circ kA(x)$, hence arguing as in the proof of Lemma 5.15 we obtain

$$\begin{aligned}\sum_{x \in \mathbb{Z}_{kn+v}^d} A^{\circ 2k}(x)([\ell]^d \circ [\ell]^d)(x) &= \sum_{x \in [-\ell, \ell]^d} A^{\circ 2k}(x)([\ell]^d \circ [\ell]^d)(x) \\ &\leqslant (k!)^2 \ell^{2d} + O\left(|A|^{2k-1}\ell^d\right). \end{aligned} \tag{5.4}$$

Parseval's identity (Lemma 5.17) and Lemma 5.18 give

$$\begin{aligned}\sum_{x \in \mathbb{Z}_{kn+v}^d} A^{\circ 2k}(x)[\ell]^d \circ [\ell]^d(x) &= \frac{1}{(kn+v)^d} \sum_{r \in \mathbb{Z}_{kn+v}^d} \widehat{A^{\circ 2k}}(x)\overline{\widehat{[\ell]^d \circ [\ell]^d}(x)} \\ &= \frac{1}{(kn+v)^d} \sum_{r \in \mathbb{Z}_{kn+v}^d} |\hat{A}(r)|^{2k} |\hat{[\ell]}^d(r)|^2 \\ &\geqslant \frac{1}{(kn+v)^d} \sum_{|r_1|+\cdots+|r_d| \leqslant k/2} |\hat{A}(r)|^{2k} |\hat{[\ell]}^d(r)|^2. \end{aligned} \tag{5.5}$$

Claim 5.19. $|\hat{[\ell]}^d(r)| \geqslant \ell^d - \frac{2\pi|r_1+r_2+\cdots+r_d|\ell^{d+1}}{kn}$.

$$\begin{aligned}
|\ell^d - [\hat{\ell}]^d(r)| &\leqslant \sum_{x\in[\ell]^d} \left|1 - e^{\frac{2\pi i r\cdot x}{kn+v}}\right| \\
&= \sum_{x\in[\ell]^d} \left|1 - \cos\left(\frac{2\pi r\cdot x}{kn+v}\right) - i\sin\left(\frac{2\pi r\cdot x}{kn+v}\right)\right| \\
&\leqslant \ell^d\left(\frac{2\pi(|r_1|+|r_2|+\cdots+|r_d|)(\ell-1)}{kn+v}\right) \\
&\leqslant \frac{2\pi(|r_1|+|r_2|+\cdots+|r_d|)\ell^{d+1}}{kn},
\end{aligned}$$

proving Claim 5.19.

Claim 5.20. $\displaystyle\sum_{|r_1|+\cdots+|r_d|\leqslant k/2} |\hat{A}(r)|^{2k} \geqslant |A|^{2k}\left(\frac{k}{\pi d}\right)^{\frac{d}{2}}(1-\epsilon(k)).$

Note that the set

$$\{x_1r_1+\cdots+x_dr_d : |r_1|+\cdots+|r_d|\leqslant k/2, x\in[n]^d\}$$

is contained in an interval of length $\frac{k}{2}n$. Therefore for such r, vectors in the complex plane corresponding to elements of A in Fourier transform will not cancel each other. Furthermore, we can expect elements of A to be more-or-less distributed in the whole of $[n]^d$, thus rotating by $n/2$ in each dimension should almost align the sum of the these vectors with the real axis.

$$\begin{aligned}
|\hat{A}(r)|^{2k} &= \left|\sum_{x\in\mathbb{Z}^d_{kn+v}} A(x)e^{2\pi i\frac{x_1r_1+\cdots+x_dr_d}{kn+v}}\right|^{2k} \\
&= \left|\sum_{x\in\mathbb{Z}^d_{kn+v}} A(x)e^{2\pi i\frac{(x_1-n/2)r_1+\cdots+(x_d-n/2)r_d}{kn+v}}\right|^{2k} \\
&\geqslant \left|\sum_{x\in\mathbb{Z}^d_{kn+v}} A(x)\cos\left(\frac{\pi(r_1+\cdots+r_d)}{k}\right)\right|^{2k}.
\end{aligned}$$

Since $|r_1|+\cdots+|r_d|\leqslant k/2$, this is greater or equal than

$$|A|^{2k}\left|1-\frac{\pi^2(r_1+\cdots+r_d)^2}{2k^2}\right|^{2k}.$$

Now we can give a bound for the sum:

$$\begin{aligned}\sum_{|r_1|+\cdots+|r_d|\leqslant k/2}|\hat{A}(r)|^{2k} &\geqslant |A|^{2k}\sum_{|r_1|+\cdots+|r_d|\leqslant k/2}\left|1-\frac{\pi^2(r_1+\cdots+r_d)^2}{2k^2}\right|^{2k}\\ &\geqslant |A|^{2k}\sum_{|r_1|+\cdots+|r_d|\leqslant k^{5/8}}\left|1-\frac{\pi^2(r_1+\cdots+r_d)^2}{2k^2}\right|^{2k}.\end{aligned}$$

Since k is large, this is greater or equal than

$$|A|^{2k}\sum_{|r_1|+\cdots+|r_d|\leqslant k^{5/8}}\left|1-\frac{\pi^4(r_1+\cdots+r_d)^4}{4k^4}\right|^{2k}e^{\frac{-\pi^2(r_1+\cdots+r_d)^2}{k}}.$$

In the last step we used inequality $1-s\geqslant e^{-s}(1-s^2)$, which is true for $s\leqslant 1$. Note that, under restrictions $|r_1|+\cdots+|r_d|\leqslant k^{5/8}$, we have

$$\left|1-\frac{\pi^4(r_1+\cdots+r_d)^4}{4k^4}\right|^{2k}\to 1$$

as $k\to\infty$. The remaining sum can be rearranged using the Cauchy-Schwarz inequality:

$$\begin{aligned}\sum_{|r_1|+\cdots+|r_d|\leqslant k^{5/8}}e^{\frac{-\pi^2(r_1+\cdots+r_d)^2}{k}} &\geqslant \sum_{|r_i|\leqslant\frac{k^{5/8}}{d}}e^{\frac{-d\pi^2(r_1^2+\cdots+r_d^2)}{k}}\\ &= \prod_{i=1}^{d}\sum_{|r_i|\leqslant\frac{k^{5/8}}{d}}e^{\frac{-\pi^2 dr_i^2}{k}}.\end{aligned}$$

Now the claim follows from the fact

$$\sum_{|r_i|\leqslant\frac{k^{5/8}}{d}}e^{\frac{-\pi^2 dr_i^2}{k}}\to\int_{-\infty}^{\infty}e^{\frac{-\pi^2 dt^2}{k}}dt=\left(\frac{k}{\pi d}\right)^{1/2}.$$

Combining equations (5.4) and (5.5) with Claims 5.19 and 5.20, we obtain

$$\begin{aligned}(k!)^2\ell^{2d}+O\left(|A|^{2k-1}\ell^d\right) &\geqslant \frac{\ell^{2d}}{(kn+v)^d}\left(1-\frac{\pi\ell d}{n}\right)^2\sum_{|r_1|+|r_2|+\cdots+|r_d|\leqslant\frac{k}{2}}|\hat{A}(r)|^{2k}\\ &\geqslant \frac{\ell^{2d}}{(kn+v)^d}\left(1-\frac{\pi\ell d}{n}\right)^2|A|^{2k}\left(\frac{k}{\pi d}\right)^{\frac{d}{2}}(1-\epsilon(k)).\end{aligned}$$

So, using trivial bound,

$$|A|^{2k}\leqslant\frac{(k!)^2(kn+v)^d+O\left(n^{d(2-\frac{1}{2k})}\ell^{-d}\right)}{\frac{\ell^d}{(kn+v)^d}\left(1-\frac{\pi\ell d}{n}\right)\left(\frac{k}{\pi d}\right)^{\frac{d}{2}}(1-\epsilon(k))}.$$

We can minimise the error term by choosing $\ell = v = n^{1-\frac{d}{2k(d+1)}}$ which, using Taylor's expansions, gives

$$|A|^{2k} \leqslant (\pi d)^{\frac{d}{2}}(1+\epsilon(k))k^{\frac{d}{2}}(k!)^2 n^d \left(1 + O\left(n^{-\frac{d}{2k(d+1)}}\right)\right).$$

Taking $2k^{\text{th}}$ roots gives the result. □

Proof of Theorem 5.6. This uses essentially the same proof except arguing as in Lemma 5.16 to obtain the equivalent of equation (5.4):

$$\sum_{x \in \mathbb{Z}^d_{kn+v}} A^{*2k}(x)([\ell]^d * [\ell]^d)(x) \leqslant |A|k!(k-1)!\ \ell^{2d} + O\left(|A|^{2k-1}\ell^d\right).$$

□

6 Infinite sets with powerless iterated sumset

Let $A \subseteq \mathbb{N}$ be a finite or infinite set of positive integers. By its *iterated sumset* or *set of subset sums* we will call a set $S_A = \{\sum_{a \in A'} a \mid A' \subseteq A, |A'| < \infty\}$. In this chapter we will consider a specific version of the following general question:

Question. Given $B \subseteq \mathbb{N}$, does there exist an infinite set $A \subseteq \mathbb{N}$ with iterated sumset S_A not containing any elements from B? If it does, how dense can it be?

It should be of no surprise that existence and density of such set A depends on the density of B. In fact, we will see in Section 6.1 that for A to exist it is sufficient that B has zero lower density (i.e. $\underline{d}(B) = \liminf_{n\to\infty} |B[n]|/n = 0$). If, on the other hand, one considers B with positive lower density, it soon becomes clear that divisibility properties of elements of B become more important than how numerous it is. To illustrate this we give two examples.

The first one is the set of all odd positive integers $B = \{1, 3, 5, 7, \dots\}$ whose density $d(B)$ is $1/2$. For such B required set A does exist and the densest one is the set of all even positive integers $\{2, 4, 6, 8, \dots\}$ with density $d(A) = 1/2$. The second example is the set of all even positive integers $B = \{2, 4, 6, 8, \dots\}$, for which no infinite sequence A as required exists. Moreover, if B is the set of all positive integers divisible by m, where $m \in \mathbb{N}$ is large, then the density $d(B) = 1/m$ is small. However, by a simple argument modulo m it is easy to see that there is no infinite set $A \subset \mathbb{N}$ (and even no set A with $\geqslant m$ distinct positive integers) with the property that its distinct elements always sum to a number lying outside B.

This observation allows us to make few further observations considering the existence of A: if B can be divided into two parts $B = B_p \cup B_{p'}$ with all elements in $B_{p'}$ not divisible by some prime p and lower density of B_p equal to zero, than we can find the set $A \subseteq p\mathbb{N}$ by using theorem below. On the other hand, if set

B contains all but finitely many multiples of some prime, the required set A will not exist. It is possible, however, to construct infinite subsets of positive integers that do not satisfy any of these conditions and it is still unknown how to deal with them.

The problem of finding densest set A for a given B is much harder. One could prove some general bounds (using a greedy algorithm for example) but they are likely very far from optimal. In this chapter we will concentrate on a specific cases of B being squares of integers and any powers of integers. Finite versions of these problems were considered before as we discussed in the Section 2.3. The infinite version of these problems were first considered by Luca [58]. In the case of perfect squares $B = \{1, 4, 9, \dots\}$ he found an example $A = \{2^{2^n} + 1, n \in \mathbb{N}\}$, which has double exponential growth. In the case of all powers $B = \{1, 4, 8, 9, \dots\}$ he gave an example $\{2^{p_1 p_2 \cdots p_n} + 1, n \geqslant n_0\}$, where p_k is the kth prime and n_0 is a large enough constant. This example is even more sparse.

We will give (in Sections 6.2 and 6.3) denser examples of sets A in both cases. They still are of exponential growth and have been improved by Dubickas and Stankevičius [23]. We discuss their example in Section 6.4.

6.1 Sets with asymptotic density zero

We start from the existence of A for sets B with zero lower density. We prove two slightly more general statements.

Theorem 6.1. *Let $m \in \mathbb{N}$ and let $B = \{b_1 < b_2 < b_3 < \dots\}$ be an infinite sequence of positive integers satisfying $\limsup_{n\to\infty}(b_{n+1} - mb_n) = \infty$. Then there exists an infinite sequence of positive integers A such that every sum over some elements of A, at most m of which are equal, is not in B.*

Proof. Take the smallest positive integer ℓ such that $b_{\ell+1} - b_\ell \geqslant 2$, and set $a_1 := b_\ell + 1$. Then $a_1 \notin B$. Suppose we already have a finite set $\{a_1 < a_2 < \cdots < a_k\}$ such that all possible $(m+1)^k - 1$ nonzero sums $\delta_1 a_1 + \cdots + \delta_k a_k$, where $\delta_1, \dots, \delta_k \in \{0, 1, \dots, m\}$, do not belong to B. Put $a_{k+1} := b_l + 1$, where l is the smallest positive integer for which $b_{l+1} - mb_l \geqslant 1 + m + m(a_1 + \cdots + a_k)$ and $b_l \geqslant a_k$. Such an l exists, because $\limsup_{n\to\infty}(b_{n+1} - mb_n) = \infty$.

Clearly, $b_l \geqslant a_k$ implies that $a_{k+1} > a_k$. In order to complete the proof of the theorem (by induction) it suffices to show that no sum of the form $\delta_1 a_1 + \cdots + \delta_k a_k + \delta_{k+1} a_{k+1}$, where $\delta_1, \ldots, \delta_{k+1} \in \{0, 1, \ldots, m\}$, lies in B. If $\delta_{k+1} = 0$, this follows by our assumption, so suppose that $\delta_{k+1} \geqslant 1$. Then $\delta_1 a_1 + \cdots + \delta_k a_k + \delta_{k+1} a_{k+1}$ is greater than $a_{k+1} - 1 = b_l$ and smaller than

$$1 + m(a_1 + \cdots + a_k + a_{k+1}) \leqslant b_{l+1} - mb_l - m + ma_{k+1} = b_{l+1} - mb_l - m + m(b_l + 1) = b_{l+1},$$

so it is not in B, as claimed. □

For $m \geqslant 2$, it can very often happen that $b_{n+1} < mb_n$ for every $n \in \mathbb{N}$, even if set B has zero lower density. For such a set B Theorem 6.1 is not applicable and a slight modification in the proof is required.

Theorem 6.2. *Let $m \in \mathbb{N}$ and let B be an infinite sequence of positive integers with zero lower asymptotic density. Then there exists an infinite sequence of positive integers A such that every sum over some elements of A, at most m of which are equal, is not in B.*

Proof. Once again, take the smallest positive integer ℓ such that $b_{\ell+1} - b_\ell \geqslant 2$, and put $a_1 := b_\ell + 1$. Then $a_1 \notin B$. Suppose we already have a finite set $\{a_1 < a_2 < \cdots < a_k\}$ such that all possible $(m+1)^k - 1$ nonzero sums $\delta_1 a_1 + \cdots + \delta_k a_k$, where $\delta_1, \ldots, \delta_k \in \{0, 1, \ldots, m\}$, do not belong to B. It suffices to prove that there exists an integer a_{k+1} greater than a_k such that, for every $i \in \{1, \ldots, m\}$, the sum $ia_{k+1} + \delta_k a_k + \cdots + \delta_1 a_1$, where $\delta_1, \ldots, \delta_k \in \{0, 1, \ldots, m\}$, is not in B.

Suppose that $B = \{b_1 < b_2 < b_3 < \ldots\}$. For any $h \in \mathbb{N}$, the set $\{hb_1 < hb_2 < hb_3 < \ldots\}$ will be denoted by hB. Put $B_i := \frac{m!}{i} B$ for $i = 1, 2, \ldots, m$. Since $\underline{d}(B_i) = 0$ for each $i = 1, \ldots, m$, we have $\underline{d}(B_1 \cup \cdots \cup B_m) = 0$. Thus, for any $v > m!(mS + 1)$, where $S := a_1 + \cdots + a_k$, there is an integer $u > m! a_k$ such that the interval $[u, u + v]$ is free of the elements of the set $B_1 \cup \cdots \cup B_m$.

Put $a_{k+1} := \lfloor u/m! \rfloor + 1$. Clearly, $a_{k+1} > a_k$. Furthermore, for any $i \in \{1, \ldots, m\}$, no element of B_i lies in $[u, u + v]$. Thus there is a nonnegative integer $j = j(i)$ such that $m! b_j / i < u$ and $m! b_{j+1} / i > u + v$. (Here, for convenience of notation, we assume that $b_0 = 0$.) Hence $ia_{k+1} > iu/m! > b_j$ and

$$ia_{k+1} + mS < ia_{k+1} + imS \leqslant i(u/m! + 1 + mS) < i(u + v)/m! < b_{j+1}.$$

In particular, these inequalities imply that, for each $i \in \{1, \dots, m\}$, the sum $ia_{k+1} + \delta_k a_k + \cdots + \delta_1 a_1$, where $\delta_1, \dots, \delta_k \in \{0, 1, \dots, m\}$, is between $b_{j(i)} + 1$ and $b_{j(i)+1} - 1$, hence it is not in B. This completes the proof of the theorem. □

As we already mentioned above theorems are sharp in some sense, as for any $\varepsilon > 0$ there exists a set with $\underline{d}(B) < \varepsilon$ and no corresponding A.

On the other hand, there exists a set B with $\underline{d}(B) = 1$ and an existing infinite set A. To construct such set start from $A = \{2^{2^i}, i \in \mathbb{N}\}$ and take $B = \mathbb{N} \setminus S_A$. Clearly B has density 1.

6.2 Infinite sets whose elements do not sum to a square

Sets of exponential growth with iterated sumset avoiding squares or powers are not hard to come by, for example one can take 2^{2n-1}, $n = 1, 2, \dots$. Any sum of its distinct elements

$$2^{2n_1-1} + \cdots + 2^{2n_l-1} = 2^{2n_1-1}(1 + 4^{n_2-n_1} + \cdots + 4^{n_l-n_1}),$$

where $1 \leqslant n_1 < \cdots < n_l$, is not a perfect square, because it is divisible by 2^{2n_1-1}, but not divisible by 2^{2n_1}.

Smaller, but still of exponential growth, is the sequence $2 \cdot 3^n$, $n = 0, 1, 2, \dots$. No sum of its distinct elements is a perfect square, because

$$2(3^{n_1} + \cdots + 3^{n_l}) = 2 \cdot 3^{n_1}(1 + 3^{n_2-n_1} + \cdots + 3^{n_l-n_1}) = h^2$$

implies that n_1 is even, so $2(1 + 3^{n_2-n_1} + \cdots + 3^{n_l-n_1})$ must be a square too. However, this number is of the form $3k + 2$ with integer k, so it is not a perfect square.

A natural way to generate an infinite sequence whose distinct elements do not sum to square is to start with $c_1 = 2$. Then, for each $n \in \mathbb{N}$, take the smallest positive integer c_{n+1} such that no sum of the form $c_{n+1} + \delta_n c_n + \cdots + \delta_1 c_1$, where $\delta_1, \dots, \delta_n \in \{0, 1\}$, is a perfect square. Clearly, $c_2 = 3$, $c_3 = 5$. Then, as $6+3 = 3^2$, $7 + 2 = 3^2$, $8 + 5 + 3 = 4^2$, $9 = 3^2$, we obtain that $c_4 = 10$, and so on. In the following table we give the first 18 elements of this sequence, which were computed by Andrius Stankevičius:

n	c_n	$\log c_n$	n	c_n	$\log c_n$
1	2	0.6931	10	2030	7.6157
2	3	1.0986	11	3225	8.0786
3	5	1.6094	12	8295	9.0234
4	10	2.3025	13	15850	9.6709
5	27	3.2958	14	80642	11.2977
6	38	3.6375	15	378295	12.8434
7	120	4.7874	16	1049868	13.8641
8	258	5.5529	17	3031570	14.9245
9	907	6.8101	18	12565348	16.3464

Here, the values of $\log c_n$ are truncated at the fourth decimal place. At the first glance, they suggest that the limit $\liminf_{n\to\infty} n^{-1}\log c_n$ is positive. If so, then the sequence c_n, $n = 1, 2, 3, \ldots$, is of exponential growth too. It seems that the sequence c_n, $n = 1, 2, 3, \ldots$, i.e.,

$$2, 3, 5, 10, 27, 38, 120, 258, 907, 2030, 3225, 8295, 15850, 80642, 378295, 1049868, \ldots$$

was not studied before. At least, it is not given in N.J.A. Sloane's on-line encyclopedia of integer sequences http://www.research.att.com/njas/sequences/. We thus raise the following problem:

Question. Is $\liminf_{n\to\infty} n^{-1}\log c_n$ positive or zero?

In the opposite direction, one can easily show that $c_n < 4^n$ for each $n \geqslant 1$. Here is the proof of this inequality by Cilleruelo (private communication). Suppose that $c_n < 4^n$. If $c_{n+1} \leqslant c_n + 4^n$, then $c_{n+1} < 4^n + 4^n < 4^{n+1}$. Otherwise, for each $j = 1, 2, \ldots, 4^n$, there exists a set $I = I_j \subseteq \{1, 2, \ldots, n\}$ such that $c_n + j + S(I) = s_j^2$, where $S(I) := \sum_{i\in I} c_i$ and $s_j \in \mathbb{N}$. There are 2^n different subsets I of $\{1, 2, \ldots, n\}$, so the set $\{4^n - 2^n, \ldots, 4^n - 1, 4^n\}$ with $2^n + 1$ elements contains some two indices $j < j'$ for which the corresponding subsets I (and so the values for $S(I)$) are equal. Subtracting $c_n + j + S(I) = s_j^2$ from $c_n + j' + S(I) = s_{j'}^2$, we deduce that $j' - j = (s_{j'} - s_j)(s_{j'} + s_j)$. Since $j' - j \leqslant 2^n$, we have $s_{j'} + s_j \leqslant 2^n$, i.e., $s_{j'} \leqslant 2^n - 1$. Hence

$$4^n - 2^n < j' < c_n + j' + S(I) = s_{j'}^2 \leqslant (2^n - 1)^2 = 4^n - 2^{n+1} + 1,$$

a contradiction.

Of course, $c_n < 4^n$ implies that $\limsup_{n\to\infty} n^{-1}\log c_n < \log 4$. Our next theorem shows that, for any fixed positive ε, there is a sequence $A = \{a_1 < a_2 < a_3 < \dots\}$ whose distinct elements do not sum to a square and whose growth is small in the sense that $\limsup_{n\to\infty} n^{-1}\log a_n < \varepsilon$.

Theorem 6.3. *For any $\varepsilon > 0$ there is a positive constant $K = K(\varepsilon)$ and an infinite sequence $A = \{a_1 < a_2 < a_3 < \dots\} \subset \mathbb{N}$ satisfying $a_n < K(1+\varepsilon)^n$ for each $n \in \mathbb{N}$ such that the sum of any number of distinct elements of A is not a perfect square.*

Proof. Fix a prime number p to be chosen later and consider the following infinite set

$$A := \{gp^{2m} + p^{2m-1} : g \in \{0, 1, \dots, p-2\},\ m \in \mathbb{N}\}.$$

Each element of A in base p can be written as $\overline{g100\dots0}$ with $2m-1$ zeros, where the 'digit' g is allowed to be zero. So all the elements of A are distinct.

First, we will show that the sum of any distinct elements of A is not a perfect square. Assume that there exists a sum S which is a perfect square. Suppose that for every $t = 1, 2, \dots, l$ the sum S contains $s_t > 0$ elements of the form $gp^{2m_t} + p^{2m_t-1}$, where $g \in \{0, 1, \dots, p-2\}$ and $1 \leqslant m_1 < m_2 < \dots < m_l$. Clearly, $s_t \leqslant p-1$. Let us write S in the form

$$S = s_1 p^{2m_1-1} + h_1 p^{2m_1} + s_2 p^{2m_2-1} + h_2 p^{2m_2} + \dots + s_l p^{2m_l-1} + h_l p^{2m_l}$$

$$= p^{2m_1-1}(s_1 + h_1 p + \dots + s_l p^{2m_l-2m_1} + h_l p^{2m_l-2m_1+1}) = p^{2m_1-1}(s_1 + pH).$$

Now, since $s_1 \in \{1, \dots, p-1\}$ and since H is an integer, we see that S is divisible by p^{2m_1-1}, but not by p^{2m_1}, so it is not a perfect square.

It remains to estimate the size of the nth element a_n of A. Write n in the form $n = (p-1)(m-1) + r$, where $r \in \{1, \dots, p-2, p-1\}$ and $m \geqslant 1$ is an integer. Suppose that the elements of A are divided into consecutive equal blocks with $p-1$ elements in each block. Then all the elements of the mth block are of the form $\overline{g100\dots0}$ (with $2m-1$ zeros), where $g = 0, 1, \dots, p-2$. Hence the nth element of A, where $n = (p-1)(m-1) + r$, is precisely the rth element of the mth block, i.e., $a_n = a_{(p-1)(m-1)+r} = (r-1)p^{2m} + p^{2m-1}$. It follows that

$$a_n \leqslant (p-2)p^{2m} + p^{2m-1} < p^{2m+1} = p^{2(n-r)/(p-1)+3} < p^{2n/(p-1)+3} = p^3 e^{(2n\log p)/(p-1)}.$$

Clearly, $(2\log p)/(p-1) \to 0$ as $p \to \infty$. Thus, for any $\varepsilon > 0$, there exists a prime number p such that $e^{(2\log p)/(p-1)} < 1+\varepsilon$. Take the smallest such a prime $p = p(\varepsilon)$. Setting $K(\varepsilon) := p(\varepsilon)^3$, we obtain that $a_n < K(\varepsilon)(1+\varepsilon)^n$ for each $n \in \mathbb{N}$. □

6.3 Infinite sets whose elements do not sum to a power

Observe that distinct elements of the sequence $2 \cdot 6^n$, $n = 0, 1, 2, \ldots$, cannot sum to a perfect power. Indeed,

$$S = 2(6^{n_1} + \cdots + 6^{n_l}) = 2^{n_1+1}3^{n_1}(1 + 6^{n_2-n_1} + \cdots + 6^{n_l-n_1}),$$

where $0 \leqslant n_1 < \cdots < n_l$, is not a perfect power, because $n_1 + 1$ and n_1 are exact powers of 2 and 3 in the prime decomposition of S. So if $S > 1$ were a kth power, where k is a prime number (which can be assumed without loss of generality), then both $n_1 + 1$ and n_1 must be divisible by k, a contradiction.

This example is already 'better' than the example $a^{p_1p_2\cdots p_n} + 1$, $n = n_0, n_0 + 1, \ldots$, given in [58] not only because it is completely explicit, but also because the sequence $2 \cdot 6^n$, $n = 0, 1, 2, \ldots$, grows slower.

As above, we can also consider the sequence $2, 3, 10, 18, \ldots$, starting with $e_1 = 2$, whose each 'next' element $e_{n+1} > e_n$, where $n \geqslant 1$, is the smallest positive integer preserving the property that no sum of the form $\delta_1 e_1 + \cdots + \delta_n e_n + e_{n+1}$, where $\delta_1, \ldots, \delta_n \in \{0, 1\}$, is a perfect power. By an argument which is slightly more complicated than the one given for c_n, one can prove again that $e_n < 4^n$ for n large enough.

However, our aim is to prove the existence of the sequence whose nth element is bounded from above by $K(\varepsilon)(1+\varepsilon)^n$ for $n \in \mathbb{N}$. For this, we shall generalize Theorem 3 as follows:

Theorem 6.4. *Let U be the set of positive integers of the form $q_1^{\alpha_1} \ldots q_k^{\alpha_k}$, where $q_1, \ldots, q_k$ are some fixed prime numbers and $\alpha_1, \ldots, \alpha_k$ run through all nonnegative integers. Then, for any $\varepsilon > 0$, there is a positive constant $K = K(\varepsilon, U)$ and an infinite sequence $A = \{a_1 < a_2 < a_3 < \ldots\} \subset \mathbb{N}$ satisfying $a_n < K(1+\varepsilon)^n$ for*

$n \in \mathbb{N}$ *such that the sum of any number of distinct elements of* A *is not equal to* uv^s *with positive integers* u, v, s *such that* $u \in U$ *and* $s \geqslant 2$.

In particular, Theorem 4 with $U = \{1\}$ implies a more general version of Theorem 3 with 'perfect square' replaced by 'perfect power'.

Proof. Fix two prime numbers p and q satisfying $p < q < 2p$. Here, the prime number p will be chosen later, whereas, by Bertrand's postulate, the interval $(p, 2p)$ always contains at least one prime number, so we can take q to be any of those primes. Consider the following infinite set

$$A := \{gp^{m+1}q^m + p^m q^{m-1} : g \in \{1, \dots, p-1\},\ m \in \mathbb{N}\}.$$

The inequality $p^{m+2}q^{m+1} + p^{m+1}q^m > (p-1)p^{m+1}q^m + p^m q^{m-1}$ implies that all the elements of A are distinct. Also, as above, by dividing the sequence A into consecutive equal blocks with $p-1$ elements each, we find that

$$a_n = rp^{m+1}q^m + p^m q^{m-1}$$

for $n = (p-1)(m-1) + r$, where $m \in \mathbb{N}$ and $r \in \{1, \dots, p-2, p-1\}$.

Assume that there exists a sum S of some distinct a_n which is of the form uv^s. Without loss of generality we may assume that $s \geqslant 2$ is a prime number. Suppose that for every $t = 1, 2, \dots, l$ the sum S contains $s_t > 0$ elements of the form $gp^{m_t+1}q^{m_t} + p^{m_t}q^{m_t-1}$, where $g \in \{1, \dots, p-1\}$ and $1 \leqslant m_1 < m_2 < \cdots < m_l$. Clearly, $s_t \leqslant p-1$, so, in particular, $1 \leqslant s_1 \leqslant p-1$. Then, as above, $S = p^{m_1}q^{m_1-1}(s_1 + pqH)$ with an integer H. If $q > p > q_k$, then $p, q \notin U$, so the equality $uv^s = p^{m_1}q^{m_1-1}(s_1 + pqH)$ implies that $s|m_1$ and $s|(m_1 - 1)$, a contradiction.

Using $a_n = rp^{m+1}q^m + p^m q^{m-1}$, where $n = (p-1)(m-1) + r$ and $p < q < 2p$, we find that

$$a_n < (p-1)q^{2m+1} + q^{2m-1} < q^{2m+2} < (2p)^{2(n-r)/(p-1)+4} < (2p)^4 e^{(2n\log(2p))/(p-1)}.$$

For any $\varepsilon > 0$, there exists a positive number p_ε such that $e^{(2\log(2p))/(p-1)} < 1+\varepsilon$ for each $p > p_\varepsilon$. Take the smallest prime number $p = p(\varepsilon)$ greater than $\max\{p_\varepsilon, q_k\}$, and put $K(\varepsilon, q_k) = K(\varepsilon, U) := 2p(\varepsilon)^4$. Then $a_n < K(\varepsilon, U)(1+\varepsilon)^n$ for each $n \in \mathbb{N}$, as claimed. □

6.4 Other constructions

We start from the exposition of two constructions that solve the finite version of the problem. The first construction was given by Erdős [29] and it is as follows:

Example 6.5. *Let p be a prime of order $n^{2/3}$ and k the largest integer such that $kp \leqslant n$ and $1 + \cdots + k < p$. Then the set $A = \{p, 2p, \ldots, kp\}$ has iterated sumset with no powers.*

The set A given in the example is a subset of $[n]$ and is of size $\sim n^{1/3}$. To see that that its iterated sumset indeed avoids powers it is sufficient to notice that the sum of any subset of A is divisible by p but not divisible by p^2. Second example is given by Cilleruelo [14]:

Example 6.6. *Let p be the largest prime smaller than $n^{1/3}$. Then the set $A = \{p, p^2 + p, 2p^2 + p, \ldots, (p-2)p^2 + p\}$ has iterated sumset with no powers.*

Again the set $A \subseteq [n]$ is of cardinality $\sim n^{1/3}$ and any its subset has its sum divisible by p but not by p^2.

Lastly, we give and infinite example of Dubickas and Stankevičius [23] that improve the examples given earlier in the chapter.

Example 6.7. *Let $p_1 < p_2 < \ldots$ be arbitrary sequence of primes. Put $p_0 = 1$ and*

$$A_k = \left\{ (jp_k^2 + p_k) \prod_{i=0}^{k-1} p_i^2 \mid j = 0, 1, \ldots, p_k - 2 \right\}, k \geqslant 1.$$

Take $A = \cup_{k=1}^{\infty} A_k$. Then S_A contains no powers of integers.

Again it is easy to see that each subset has sum divisible by some p_k but not by p_k^2. In order to get a slow growing set A one needs to select primes p_k carefully. In particular, if they satisfy $p_1 \cdots p_{k-1} < p_k < 1.4 p_1 \cdots p_{k-1}$ one gets a set A with $A[n] \gg n^{1/9}$.

Another way is to take any sequence of real numbers $\{g_n\}$ with $\lim_{n\to\infty} g_n = 0$ and primes satisfying $p_k \geqslant p_{k-1}$ and $(p_1 \cdots p_{k-1})^2 < g_{p_k}$. In this case one gets a set A locally nearly matching the lower bound, that is with infinitely many integers n_i satisfying $A[n_i] \geqslant n_i^{1/3} g_{n_i}^{-1/3}$. We direct the reader to their paper for more details.

7 Multiplicative functions additive on primes

Recall that a function $f : \mathbb{N} \to \mathbb{C}$ is called *multiplicative* if

$$f(ab) = f(a)f(b) \quad \text{for all} \quad a, b \in \mathbb{N} \quad \text{satisfying} \quad \gcd(a, b) = 1. \tag{7.1}$$

It is well known that a Cauchy functional equation $f(x+y) = f(x) + f(y)$ solved for functions defined on integers (or rationals) has two multiplicative solutions: $f(x) = x$ and a trivial one $f(x) = 0$. Spiro [78] has got a similar conclusion from a slightly weaker additivity condition. He proved that the only multiplicative function f which satisfies $f(p_0) \neq 0$ for at least one prime number p_0 and is additive on the set of primes, i.e. $f(p+q) = f(p) + f(q)$ for all primes p, q, is the identity function $f(n) = n$ for each $n \in \mathbb{N}$.

Fang [31] derived the same conclusion for multiplicative functions f which are additive on sums of three primes, namely, $f(p+q+r) = f(p) + f(q) + f(r)$ for all primes p, q, r. In this chapter we extend this result to multiplicative functions which are additive on sums of k primes, where $k \geqslant 2$ is a fixed integer.

Theorem 7.1. *Let $k \geqslant 2$ be a fixed integer. If a multiplicative function $f : \mathbb{N} \to \mathbb{C}$ satisfies*

$$f(p_1 + p_2 + \cdots + p_k) = f(p_1) + f(p_2) + \cdots + f(p_k) \tag{7.2}$$

for any primes $p_1, p_2, \ldots, p_k$ and $f(p_0) \neq 0$ for at least one prime p_0 then $f(n) = n$ for each $n \in \mathbb{N}$.

We shall only prove the theorem for $k \geqslant 3$, since the case $k = 2$ has been treated earlier in [78]. However, the case $k = 3$ (which was also treated earlier in [31]) is included here, since our proof is slightly different.

Note that selecting $a = 1$ and $b = p_0$ in (7.1) we obtain $f(p_0) = f(1)f(p_0)$. Since $f(p_0) \neq 0$, this implies

$$f(1) = 1. \tag{7.3}$$

Our aim is to show that under conditions of Theorem 7.1 the equality

$$f(p) = p \tag{7.4}$$

holds for each prime number p and then (via Lemma 7.7 below) to extend this equality to $f(n) = n$ for each $n \in \mathbb{N}$.

This chapter is organized as follows. In the next section we give some results on the representation of integers as sums of prime numbers. Then in Section 3 we prove two auxiliary results on functions $f : \mathbb{N} \to \mathbb{C}$ satisfying (7.2) for some fixed $k \geqslant 3$. These two (Lemmas 7.6 and 7.7) are the key results in our approach. The proof of Theorem 7.1 is then completed in Section 7.3, by proving (7.4).

7.1 Representation of integers as sums of primes

By a classical result of Vinogradov on the ternary Goldbach problem, every sufficiently large odd number can be expressed as the sum of three primes. Liu and Wang [57] showed that this is true for all odd integers greater than

$$n(V) := e^{3100}.$$

Under assumption of the generalized Riemann hypothesis every odd integer greater than or equal to 7 is the sum of three primes (see [21]). On the other hand, the binary Goldbach conjecture asserts that every even integer greater than or equal to 4 can be expressed as the sum of two primes. This famous conjecture is still open, although it has been checked up to $2 \cdot 10^{10}$ (see [36]), and then up to $4 \cdot 10^{14}$ (see [66]). Recently, Oliveira e Silva (see http://www.ieeta.pt/∼tos/goldbach.html) has checked it up to $4 \cdot 10^{18}$, so in this paper let us take

$$n(B) := 4 \cdot 10^{18}, \tag{7.5}$$

where $n(B)$ is the largest known integer until which the binary Goldbach conjecture has been verified.

An important ingredient in our proof of Theorem 7.1 is a recent result of Tao [80] who showed that

Theorem 7.2. *Every odd number greater than* 1 *can be expressed as the sum of at most five prime numbers.*

This improves an earlier result of Ramaré [65] who showed that every even number is the sum of at most six primes. In [46] Theorem 7.2 was proved under assumption of the Riemann hypothesis.

More precisely, we will use the following two implications of Theorem 7.2:

Lemma 7.3. *Every even number greater than or equal to* 12 *can be expressed as the sum of exactly six primes.*

Proof. The claim trivially holds for $n = 12$ and $n = 14$. Consider an even number $n \geqslant 16$. The number $n - 13 > 1$ is odd, so it can be written as the sum of at most 5 primes. Since 13 can be written as the sum of $1, 2, 3, 4$ or 5 primes (13, $2 + 11$, $3 + 3 + 7$, $2 + 2 + 2 + 7$, $2 + 2 + 2 + 2 + 5$), we conclude that $n = n - 13 + 13$ can be written as the sum of exactly six primes. □

Lemma 7.4. *For any $k \geqslant 7$ every number greater than or equal to $2k$ can be expressed as the sum of exactly k primes.*

Proof. Take any integer $n \geqslant 2k$. Consider the number $n - 2(k - 6)$ if n is even and $n - 3 - 2(k - 7)$ if n is odd. Both of them are even and greater than or equal to 12. By Lemma 7.3, they can be expressed as the sum of six primes. Then n itself can be written as the sum of exactly $6 + k - 6 = k$ primes, as claimed. □

Below, we will also use Lemma 3 from [31] (see also [78]):

Lemma 7.5. *Let $\nu_p(n)$ be the exponent of the prime number p in the prime factorization of n, and let*

$$H := \{n : \nu_p(n) \leqslant 1 \text{ if } p > 1000;\ \nu_p(n) \leqslant \lfloor (9 \log 10)/\log p \rfloor - 1 \text{ if } p < 1000\}\,. \tag{7.6}$$

Then for any integer $m > 10^{10}$ there exist at least four prime numbers $q_1, q_2, q_3, q_4 < m$ such that $m + q_i \in H$ $(i = 1, 2, 3, 4)$.

7.2 Two auxiliary lemmas

The next lemma uses the notation introduced in (7.5). (Note that in this lemma we are not using the multiplicativity of f.)

Lemma 7.6. *Fix $k \geqslant 3$. Let $f : \mathbb{N} \to \mathbb{C}$ be a any function satisfying* (7.2)*. Then*

$$f(p) = (p-2)f(3) - (p-3)f(2) \tag{7.7}$$

for all primes $p \leqslant n(B)$. Moreover, if $k \geqslant 6$ then (7.7) *holds for all primes p.*

Proof. Observe that if some prime numbers $p_1, p_2, p_3, q_1, q_2, q_3$ satisfy $p_1+p_2+p_3 = q_1 + q_2 + q_3$, then f satisfies

$$f(p_1) + f(p_2) + f(p_3) = f(q_1) + f(q_2) + f(q_3).$$

Indeed, for $k = 3$ this follows immediately from the condition (7.2), whereas for $k \geqslant 4$ we can add $k - 3$ equal summands (say 2) to both sides, then use the condition (7.2) and remove unnecessary terms $f(2)$ from both sides.

Equality (7.7) certainly holds for $p = 2$ and $p = 3$, since then it is an identity. By the above observation, equality $2+2+5 = 3+3+3$ implies $f(5) = 3f(3)-2f(2)$ and equality $2+2+7 = 3+3+5$ implies $f(7) = f(5)+2f(3)-2f(2) = 5f(3)-4f(2)$, so (7.7) holds for $p = 2, 3, 5, 7$.

Assume that $11 \leqslant q \leqslant n(B)$ is the smallest prime for which equality (7.7) is false. (The proof of the first part is finished if there is no such q.) Consider the number $q - 1$. It is even, greater than 4 and smaller than $n(B)$, so, by the definition of $n(B)$, it can be written as the sum of two primes $p_1 + p_2$. Clearly, $p_1, p_2 < q$. From $q+3+3 = 7+p_1+p_2$ and the validity of (7.7) for $p = 3, 7, p_1, p_2$ it follows that

$$\begin{aligned} f(q) &= f(p_1) + f(p_2) + f(7) - f(3) - f(3) \\ &= (p_1 - 2 + p_2 - 2 + 5 - 2)f(3) - (p_1 - 3 + p_2 - 3 + 4)f(2) \\ &= (q-2)f(3) - (q-3)f(2), \end{aligned}$$

a contradiction.

To prove the second part we assume that there exists a prime p for which (7.7) does not hold and take the smallest $p > n(B)$ with this property, so that

$$f(p) \neq (p-2)f(3) - (p-3)f(2).$$

Let us express the number $p+11$ as two different sums sum of 6 primes. Firstly, write it as $p+2+2+2+2+3$. This is also $13+q$, where $q := p-2$ is an odd number. As in the proof of Lemma 7.3, since q can be expressed as the sum of at most 5 primes, the number $13+q$ can be expressed as the sum of exactly six primes $p_1+\cdots+p_6$, all of them smaller than p, so this is a different representation of the same number as the sum of six primes. Then, since $k \geqslant 6$, from equality $p+2+2+2+2+3 = p_1+\cdots+p_6$ and equality (7.7) for $p = p_1, \ldots, p_6, 2, 3$ it follows that

$$\begin{aligned} f(p) &= f(p_1)+\cdots+f(p_6)-4f(2)-f(3) \\ &= (p_1+\cdots+p_6-13)f(3)-(p_1+\cdots+p_6-14)f(2) \\ &= (p+11-13)f(3)-(p+11-14)f(2) \\ &= (p-2)f(3)-(p-3)f(2), \end{aligned}$$

a contradiction. This completes the proof of the lemma. □

The next lemma shows that (7.4) can be indeed extended to $f(n) = n$ for every $n \in \mathbb{N}$.

Lemma 7.7. *Suppose that a multiplicative function f satisfies* (7.2) *for some fixed $k \geqslant 3$. If $f(p) = p$ for each prime number p then $f(n) = n$ for each $n \in \mathbb{N}$.*

Proof. By (7.1) and (7.3), it suffices to show that

$$f(p^\alpha) = f(p)^\alpha$$

for each prime number p and each integer $\alpha \geqslant 2$, since the case $\alpha = 1$ is covered by the condition of the lemma.

Suppose first that p is an odd prime. By Vinogradov's theorem, every sufficiently large odd integer $n \geqslant n(V) = e^{3100}$ is the sum of three primes. Take a prime number $q > n(V) + 2(k-3)$ and write the number $p^\alpha q - 2(k-3) > n(V)$ as the sum of three primes $p_1+p_2+p_3$. Then the odd integer $p^\alpha q$ is the sum of the following k primes $p_1, p_2, p_3, 2, \ldots, 2$. Applying (7.2) and the condition of the lemma, we find that

$$f(p^\alpha q) = f(p_1)+f(p_2)+f(p_3)+(k-3)f(2) = p_1+p_2+p_3+2(k-3) = p^\alpha q.$$

Hence $p^\alpha q = f(p^\alpha q) = f(p^\alpha)f(q) = f(p^\alpha)q$, by (7.1) and $f(q) = q$. Dividing both sides by q, we obtain $f(p^\alpha) = p^\alpha$. By the multiplicativity of f and $f(1) = 1$, this yields

$$f(n) = n \quad \text{for each odd} \quad n. \tag{7.8}$$

To complete the proof it remains to show that $f(2^\alpha) = 2^\alpha$ for each integer $\alpha \geqslant 2$. If k is even then $k \geqslant 4$ and, selecting as above some prime number $q > n(V)+2k-5$, we can express the odd number $2^\alpha q-3-2(k-4) > n(V)$ as the sum three primes $p_1+p_2+p_3$. Thus $2^\alpha q$ is the sum of k primes $p_1, p_2, p_3, 3, 2, \ldots, 2$ (where $k-4$ primes are equal to 2). We then arrive to the conclusion $f(2^\alpha) = 2^\alpha$ in the same manner as above.

In the case when $k \geqslant 3$ is odd the proof is different and uses Dirichlet's theorem on prime numbers in arithmetic progression. Note that the number $2^{\alpha-1} + 2 - k$ is odd whenever k is odd, so the integers 2^α and $2^{\alpha-1} + 2 - k$ are coprime. By Dirichlet's theorem, the arithmetic progression $2^\alpha s + 2^{\alpha-1} + 2 - k$, $s = k, k+1, \ldots,$ contains infinitely many primes. Take one of them, say, q. Clearly, $q := 2^\alpha s + 2^{\alpha-1} + 2 - k$ is odd. Selecting in (7.2) two primes equal to q and $k-2$ primes equal to 2, we find that

$$2^\alpha(2s+1) = 2q+2(k-2) = 2f(q)+(k-2)f(2) = f(2q+2(k-2)) = f(2^\alpha(2s+1)).$$

By (7.1) and (7.8), we see that the right hand side is equal to $f(2^\alpha)(2s+1)$. Dividing both sides by $2s+1$ we obtain $f(2^\alpha) = 2^\alpha$, as required. □

7.3 Proof of Theorem 7.1

In view of (7.3) and Lemma 7.7 in order to complete the proof of Theorem 7.1 it remains to prove (7.4). We first establish (7.4) for $k \geqslant 6$. In this case Lemma 7.6 asserts that $f(p) = (p-2)f(3) - (p-3)f(2)$ for all prime numbers p.

Take any prime number $p \geqslant k$. From Lemmas 7.3 and 7.4 we know that $2p \geqslant 2k$ can be expressed as the sum of exactly k primes $p_1, \ldots, p_k$. Using the multiplicativity of f, we find that

$$f(2)f(p) = f(p_1)+\cdots+f(p_k) = (p_1+\cdots+p_k-2k)f(3)-(p_1+\cdots+p_k-3k)f(2).$$

Hence

$$f(2)\left((p-2)f(3)-(p-3)f(2)\right)=(2p-2k)f(3)-(2p-3k)f(2)$$

and

$$p(f(2)-2)(f(3)-f(2))-(k-f(2))(3f(2)-2f(3))=0. \tag{7.9}$$

Since this equality holds for all primes $p \geqslant k$, dividing by p and letting $p \to \infty$ in (7.9) we derive that $(f(2)-2)(f(3)-f(2))=0$. Hence $f(2)=2$ or $f(2)=f(3)$. In the first case, substituting $f(2)=2$ into (7.9) and using $k>f(2)=2$, we find that $f(3)=3f(2)/2=3$. Hence

$$f(p)=(p-2)f(3)-(p-3)f(2)=3(p-2)-2(p-3)=p$$

for all primes $p \geqslant 2$.

We will show next that the second case, $f(2)=f(3)$, is impossible. Indeed, we then must have

$$f(p)=(p-2)f(3)-(p-3)f(2)=(p-2)f(2)-(p-3)f(2)=f(2)$$

for all prime numbers p. Taking any prime $p \geqslant k$ and applying (7.9) with $f(2)=f(3)$ we obtain $(k-f(2))f(2)=0$, i.e. $f(2)^2=kf(2)$. Now let us take two odd primes $r \neq q$ and express the even integer $2rq$ as the sum of k primes. This gives

$$f(2)^3=f(2)f(r)f(q)=f(2rq)=f(2)+\cdots+f(2)=kf(2),$$

because $f(p)=f(2)$ for all primes p. Hence $f(2)^3=kf(2)$. It follows that $f(2)^3=f(2)^2$, thus $f(2)=1$ or $f(2)=0$. In the first case, $f(2)=1$, we find that $k=f(2)=1$, a contradiction. In the second case, $f(2)=0$, we must have $f(p)=f(2)=0$ for all prime numbers p, which is not allowed by the condition of the theorem. This proves (7.4) for each $k \geqslant 6$.

We next prove (7.4) for $k \in \{3,4,5\}$. In this case we proceed as in [31]. (In principle, the method developed in [31] and [78] can be extended to some k greater than 5 as well by increasing the number of primes in Lemma 7.5 from four primes to more primes, but this eventually increases the constant 10^{10} of Lemma 7.5 beyond the constant $n(B)$, so one needs an alternative argument to cover all k.) Some parts of the proof for small k are the same, but we include them here for completeness.

Lemma 7.8. *Let $k = 3, 4$ or 5, and let f be a multiplicative function satisfying* (7.2)*. Then for all primes $p \leqslant n(B)$ we have either $f(p) = p$ or $f(p) = 0$.*

Proof. From Lemma 7.6 we know that such a function f satisfies (7.7) for all primes $p \leqslant n(B)$. Depending on the value of k, we consider the equalities

$$f(2) + f(2) + f(3) = f(7) \quad \text{for} \quad k = 3,$$

or

$$f(2) + f(3) + f(3) + f(3) = f(11) \quad \text{for} \quad k = 4,$$

or

$$f(2) + f(2) + f(2) + f(2) + f(3) = f(11) \quad \text{for} \quad k = 5.$$

After substituting $f(7) = 5f(3) - 4f(2)$ and $f(11) = 9f(3) - 8f(2)$ (see (7.7)) from each of these three equalities we find that $2f(3) = 3f(2)$. Hence

$$f(p) = (p-2)f(3) - (p-3)f(2) = 3(p-2)f(2)/2 - (p-3)f(2) = pf(2)/2 \quad (7.10)$$

for each prime $p \leqslant n(B)$.

Now, again depending on the value of k, consider the following equalities $f(2) + f(2) + f(2) = f(2)f(3) = 3f(2)^2/2$ for $k = 3$, or $5f(2) = f(2) + f(2) + f(3) + f(3) = f(2)f(5) = 5f(2)^2/2$ for $k = 4$, or $f(2) + f(2) + f(2) + f(2) + f(2) = f(2)f(5) = 5f(2)^2/2$ for $k = 5$. (These follow from the multiplicativity of f and (7.10) applied to $p = 3$ and $p = 5$.) Every one of those implies $f(2) = 2$ or $f(2) = 0$. Thus, by (7.10), either $f(p) = p$ for all primes p up to $n(B)$ or $f(p) = 0$ for all primes p up to $n(B)$. □

Finally, let $k = 3, 4$ or 5, and let f be a multiplicative function satisfying the conditions of Theorem 7.1. We claim that then $f(n) = n$ for all $n \in H$, where H is defined in (7.6). Of course, this assertion implies (7.4) for $k = 3, 4, 5$, because the set of prime numbers is a subset of the set H defined in (7.6).

From Lemma 7.8 we know that for all primes $p \leqslant n(B)$ we have either $f(p) = p$ or $f(p) = 0$. We will show that first case extends to $f(n) = n$ for all $n \in H$ and that the second extends to $f(n) = 0$ for all $n \in H$. Since all primes are in H, the latter function does not satisfy the conditions of Theorem 7.1. We will only show

how to extend the case $f(p) = p$, as the case $f(p) = 0$ can be handled exactly the same.

We first prove that $f(n) = n$ for all $n \leqslant n(B)$. By the definition of the constant $n(B)$, every even number n in the range $4 \leqslant n \leqslant n(B)$ can be written as the sum of two primes. Hence every number n in the range $6 \leqslant n \leqslant n(B)$ can be written as the sum of three primes (simply subtract 2 or 3 to get an even number greater than or equal to 4). Similarly, every n satisfying $8 \leqslant n \leqslant n(B)$ is the sum of four primes and every n in the range $10 \leqslant n \leqslant n(B)$ is the sum of five primes. From (7.2) and equality $f(p) = p$ for primes $p \leqslant n(B)$ it follows that $f(n) = n$ for $10 \leqslant n \leqslant n(B)$. For the remaining few numbers we can use the multiplicativity property of f. Indeed, take any integer a in the range $1 \leqslant a \leqslant 9$ and write $11f(a) = f(11a) = 11a$. This yields $f(a) = a$.

In order to prove that $f(n) = n$ holds for all $n \in H$ we use the induction on n. Let $n \in H$, $n > n(B)$, be the smallest integer for which the equality $f(n) = n$ has not yet been proven.

If n is not a prime power, then we can factor n as $n_1 n_2$, where $\gcd(n_1, n_2) = 1$, $n > n_1, n_2 \geqslant 2$. By the definition of H (see (7.6)), both n_1, n_2 belong to H. Hence, by induction and multiplicativity, $f(n) = f(n_1 n_2) = f(n_1)f(n_2) = n_1 n_2 = n$. On the other hand, if n is a prime power then, by the definition of the class H, it has to be a prime, say, $n = p$. We consider the cases $k = 3, 4, 5$ separately.

For $k = 3$ take $m := n + 2 = p + 2$. Using Lemma 7.5 select four primes p_1, p_2, p_3, p_4 smaller than $p + 2$ such that $p + 2 + p_i \in H$ for $i = 1, 2, 3, 4$. At least three of them are odd, and at least one of these three, say p_1, is smaller than $p - 2$. Then $p + 2 + p_1$ is in H and is even, but not a power of 2, by the definition of H given in (7.6). Thus $p + 2 + p_1$ can be factorized as $n_1 n_2$, where n_1, n_2 are coprime and smaller than p, since $p + 2 + p_1 < p + 2 + p - 2 = 2p$. Observe that $2, p_1, n_1, n_2 \in H$, by the definition of H. Therefore, by induction and multiplicativity, $f(p) + f(2) + f(p_1) = f(n_1)f(n_2)$ yields $f(p) = n_1 n_2 - 2 - p_1 = p$.

For $k = 4$ we take $m := p + 2 + 2$ and again using Lemma 7.5 select four primes p_1, p_2, p_3, p_4 smaller than $p + 4$ such that $p + 2 + 2 + p_i \in H$ for $i = 1, 2, 3, 4$. As above, at least three of those primes are odd. If one of these three is smaller than $p - 4$, we can proceed as in the case $k = 3$ (factorize and use induction as both factors are smaller than p). If not, then one of them, say p_1, has to be equal to

$p-2$ or to $p+2$. Then $p+2+2+p_1$ is divisible by 4, so we can factor it as $n_1 n_2$, where both n_1, n_2 are coprime and greater than 2, hence smaller than p. Now we can use the induction and multiplicativity as above to obtain $f(p) = p$.

Finally, for $k = 5$ take $m := p + 2 + 2 + 2$ and once again find four primes p_1, p_2, p_3, p_4 smaller than $p+6$ such that $p+2+2+2+p_i \in H$ for $i = 1, 2, 3, 4$. If one of the three odd primes is smaller than $p - 6$, continue as in the case $k = 3$. Else, one of them, say p_1, has to be equal to $p - 4, p$ or $p + 4$, since all three numbers $p - 6, p - 2, p + 2$ cannot be prime. Then $p + 2 + 2 + 2 + p_1$ is divisible by 4 and we continue as in the case $k = 4$ to obtain $f(p) = p$. This completes the induction.

8 Conclusions

From the results obtained in the previous chapters we derive the following conclusions:

- The size of the largest co-Sidon subset pair $A' \subseteq A, B' \subseteq B$ depends on the additive energy of the initial pair of sets $\mathsf{E}(A, B)$.
- Infinite sparse sets with bounded jumps can not have bounded doubling.
- The size of B_{2k} set in d-dimensional cube $[n]^d$ is asymptotically bounded by $(k!)^{\frac{1}{k}} k^{\frac{d}{2k}} n^{\frac{d}{2k}}$. The size of B_{2k-1} set in d-dimensional cube $[n]^d$ is asymptotically bounded by $(k!)^{\frac{2}{2k-1}} k^{\frac{d-1}{2k-1}} n^{\frac{d}{2k-1}}$.
- Divisibility properties of natural numbers are useful in constructing dense sets with square (or power) free iterated sumset
- Multivariate Cauchy functional equation with the additivity condition restricted to primes has essentially unique non-zero solution.

Bibliography

[1] Noga Alon, *Subset sums*, J. Number Theory **27** (1987), 196–205.

[2] Noga Alon and Paul Erdős, *An application of graph theory to additive number theory*, Eur. J. Comb. **6** (1985), 201–203.

[3] Noga Alon and Gregory A. Freiman, *On sums of subsets of a set of integers*, Combinatorica **8** (1988), no. 4, 297–306.

[4] Antal Balog and Endre Szemerédi, *A statistical theorem of set addition*, Combinatorica **14** (1994), no. 3, 263–268.

[5] Fabricio Benevides, Jonathan Hulgan, Nathan Lemons, Cory Palmer, Ago-Erik Riet, and Jeffrey P. Wheeler, *Additive properties of a pair of sequences*, Acta Arith. **139** (2009), no. 2, 185–197.

[6] Yuri Bilu, *Structure of sets with small sumset*, Structure Theory of Set Addition (Jean-Marc Deshouillers et al., ed.), vol. 258, Paris: Société Mathématique de France, Astérisque, 1999, pp. 77–108.

[7] Mei-Chu Chang, *A polynomial bound in Freiman's theorem*, Duke Math. J. **113** (2002), no. 3, 399–419.

[8] Kang-Kang Chen and Yong-Gao Chen, *On $f(p+q) = f(p)+f(q)$ for all odd primes p and q*, Publ. Math. Debrecen **76** (2010), 425–430.

[9] Sheng Chen, *On Sidon sequences of even orders*, Acta Arith. **64** (1993), no. 4, 325–330.

[10] ———, *On the size of finite Sidon sequences*, Proc. Am. Math. Soc. **121** (1994), no. 2, 353–356.

[11] Pham Van Chung, *Multiplicative functions satisfying the equation $f(m^2 + n^2) = f(m^2) + f(n^2)$*, Math. Slovaca **46** (1996), 165–171.

[12] Pham Van Chung and Bui Minh Phong, *Additive uniqueness sets for multiplicative functions*, Publ. Math. **55** (1999), no. 3-4, 237–243.

[13] Javier Cilleruelo, *Infinite sidon sequences*, arXiv preprint, arXiv:1209.0326.

[14] ———, *Solution of the problem 38*, Gaceta de la Real Sociedad Matematica Española **9** (2006), 455–460.

[15] ———, *Sidon sets in* $\mathbb{N}^d$, J. Comb. Theory, Ser. A **117** (2010), no. 7, 857–871.

[16] ———, *Combinatorial problems in finite fields and Sidon sets*, Combinatorica **32** (2012), no. 5, 497–511.

[17] Javier Cilleruelo, Imre Z. Ruzsa, and Carlos Vinuesa, *Generalized Sidon sets*, Adv. Math. **225** (2010), no. 5, 2786–2807.

[18] Javier Cilleruelo and Rafael Tesoro, *Dense infinite* b_h *sequences*, arXiv preprint, arXiv:1206.3087.

[19] Ernie Croot and Vsevolod F. Lev, *Problems presented at the workshop on recent trend in additive combinatorics, 2004*, www.aimath.org/WWN/additivecomb/additivecomb.pdf.

[20] Jean-Marie De Koninck, Imre Kátai, and Bui M. Phong, *A new characteristic of the identity function*, J. Number Theory **63** (1997), no. 2, 325–338.

[21] Jean-Marc Deshouillers, Gove Effinger, Herman J.J. te Riele, and Dmitrii Zinoviev, *A complete Vinogradov 3-primes theorem under the Riemann hypothesis*, Electron. Res. Announc. Am. Math. Soc. **3** (1997), 99–104.

[22] Artūras Dubickas, Tomasz Schoen, Manuel Silva, and Paulius Šarka, *Finding large co-sidon subsets in sets with a given additive energy*, Eur. J. Comb. **34** (2013), 1144–1157.

[23] Artūras Dubickas and Andrius Stankevičius, *Sumsets without powerful numbers*, Acta Arith. **130** (2007), no. 4, 381–387.

[24] Artūras Dubickas and Paulius Šarka, *On multiplicative functions which are additive on sums of primes*, to appear in Aequationes Math.

[25] ———, *Infinite sets of integers whose distinct elements do not sum to a power*, J. Integer Seq. **9** (2006), no. 4, 9p.

[26] ———, *Sumsets of sparse sets*, Period. Math. Hung. **64** (2012), no. 2, 169–179.

[27] Paul Erdős, *Some applications of Ramsey's theorem to additive number theory*, Eur. J. Comb. **1** (1980), 43–46.

[28] ———, *Extremal problems in number theory, combinatorics and geometry*, Proceedings of the International Congress of Mathematicians, Warszawa (Zbigniew Ciesielski and Czesław Olech, eds.), vol. 1, PWN-Polish Scientific Publishers, 1984.

[29] ———, *Some problems and results on combinatorial number theory*, Proceedings of the 1st China Conference in Combinatorics, 1989.

[30] Paul Erdős and Pál Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), no. 4, 212–215.

[31] Jin-Hui Fang, *A characterization of the identity function with equation* $f(p+q+r) = f(p) + f(q) + f(r)$, Combinatorica **31** (2011), no. 6, 697–701.

[32] Gregory A. Freiman, *Foundations of a structural theory of set addition (translations of mathematical monographs)*, Providence: AMS, 1973.

[33] Alfred Geroldinger and Imre Z. Ruzsa, *Combinatorial number theory and additive group theory*, Basel: Birkhäuser, 2009.

[34] Timothy W. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588.

[35] Sidney W. Graham, B_h *sequences*, Proceedings of Conference in Honor of Heini Halberstam (Bruce C. Berndt, Harold G. Diamond, and A. J. Hildebrand, eds.), Boston, MA: Birkhäuser, 1996.

[36] Andrew Granville, Jan Van de Lune, and Herman J.J. te Riele, *Checking the Goldbach conjecture on a vector computer*, Number theory and applications. Proceedings of the NATO Advanced Study Institute, Proc. NATO ASI,

Banff/Can. 1988, NATO ASI Ser., Ser. C 265, Kluwer Academic Publishers, 1989, pp. 423–433.

[37] Ben Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), no. 4, 365–390.

[38] Richard K. Guy, *Unsolved problems in number theory. 2nd ed.*, New York, NY: Springer-Verlag, 1994.

[39] Heini Halberstam and Klaus F. Roth, *Sequences. Vol. I*, Oxford university press, 1966.

[40] Herald Helfgott, *Major arcs for Goldbach's theorem*, arXiv preprint, arXiv:1305.2897.

[41] ———, *Minor arcs for Goldbach's theorem*, arXiv preprint, arXiv:1205.5252.

[42] Herald Helfgott and David J. Platt, *Numerical verification of the ternary Goldbach conjecture up to 8.875e30*, arXiv preprint, arXiv:1305.3062.

[43] David Hilbert, *Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n-ter Potenzen (Waringsches Problem)*, Mathematische Annalen **67** (1909), 281–300.

[44] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, New York, NY: Wiley, 2000.

[45] Xing-De Jia, *On finite Sidon sequences*, J. Number Theory **44** (1993), no. 1, 84–92.

[46] Leszek Kaniecki, *On Šnirelman's constant under the Riemann hypothesis*, Acta Arith. **72** (1995), no. 4, 361–374.

[47] Jeong H. Kim and Van H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.

[48] Yoshiharu Kohayakawa, Sang J. Lee, Vojtěch Rödl, and Wojciech Samotij, *The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers*, to appear in Random Struct. Algorithms.

[49] János Komlós, Miklós. Sulyok, and Endre Szemerédi, *Linear problems in combinatorial number theory*, Acta Math. Acad. Sci. Hung. **26** (1975), 113–121.

[50] Sergei Konyagin, http://atlas-conferences.com/cgi-bin/abstract/cbdg-67.

[51] Jeffrey C. Lagarias, Andrew M. Odlyzko, and James B. Shearer, *On the density of sequences of integers the sum of no two of which is a square. I: Arithmetic progressions*, J. Comb. Theory, Ser. A **33** (1982), 167–185.

[52] ———, *On the density of sequences of integers the sum of no two of which is a square. II: General sequences*, J. Comb. Theory, Ser. A **35** (1983), 123–139.

[53] Allison Lewko and Mark Lewko, *On the structure of sets of large doubling*, Eur. J. Comb. **32** (2011), no. 5, 688–708.

[54] Bernt Lindström, *A remark on B_4-sequences*, J. Comb. Theory **7** (1969), no. 3, 276–277.

[55] ———, *An inequality for B_2-sequences*, J. Comb. Theory **6** (1969), no. 2, 211–212.

[56] Edith Lipkin, *On representation of r-th powers by subset sums*, Acta Arith. **52** (1989), no. 4, 353–366.

[57] Ming-Chit Liu and Tianze Wang, *On the Vinogradov bound in the three primes Goldbach conjecture*, Acta Arith. **105** (2002), no. 2, 133–175.

[58] Florian Luca, *Infinite sets of positive integers whose sums are free of powers*, Rev. Colomb. Mat. **36** (2002), no. 2, 67–70.

[59] John C. M. Nash, *Freiman's theorem answers a question of Erdős*, J. Number Theory **27** (1987), 7–8.

[60] Melvyn B. Nathanson, *Additive number theory. Inverse problems and the geometry of sumsets*, New York, NY: Springer, 1996.

[61] ———, *Additive number theory. The classical bases*, New York, NY: Springer, 1996.

[62] Hoi H. Nguyen and Van H. Vu, *Squares in sumsets*, An irregular mind. Szemerédi is 70. Dedicated to Endre Szemerédi on the occasion of his seventieth birthday. (Imre Bárány and József Solymosi, eds.), Berlin: Springer, 2010.

[63] Bui M. Phong, *On sets characterizing the identity function*, Ann. Univ. Sci. Budap. Rolando Eőtvős, Sect. Comput. **24** (2004), 295–306.

[64] Laurence Rackham and Paulius Šarka, *B_h sequences in higher dimensions*, Electron. J. Comb. **17** (2010), no. 1, 15 p.

[65] Olivier Ramaré, *On Šnirel'man's constant*, Ann. Sc. Norm. Super. Pisa, Cl. Sci., IV. Ser. **22** (1995), no. 4, 645–706.

[66] Jörg Richstein, *Verifying the Goldbach conjecture up to* $4 \cdot 10^{14}$, Math. Comput. **70** (2001), no. 236, 1745–1749.

[67] J. Barkley Rosser and Lowell Schoenfeld, *Approximate formulas for some functions of prime numbers*, Ill. J. Math. **6** (1962), 64–94.

[68] Juanjo Rue, Paulius Šarka, and Ana Zumalacarregui, *On the error term of the logarithm of the lcm of a quadratic sequence*, to appear in J. Théor. Nombres Bordx.

[69] Diego Ruiz and Trujillo Carlos, *Construction of $B_h[g]$ sets in product of groups*, arXiv preprint, arXiv:1302.0071v1.

[70] Imre Z. Ruzsa, *Generalized arithmetical progressions and sumsets*, Acta Math. Hung. **65** (1994), no. 4, 379–388.

[71] ———, *An infinite Sidon sequence*, J. Number Theory **68** (1998), no. 1, 63–71.

[72] Tom Sanders, *On the Bogolyubov-Ruzsa lemma*, Anal. PDE **5** (2012), no. 3, 627–655.

[73] András. Sárközy, *Finite addition theorems. II*, J. Number Theory **48** (1994), no. 2, 197–218.

[74] Tomasz Schoen, *On sets of natural numbers whose sumset is free of squares*, J. Comb. Theory, Ser. A **88** (1999), no. 2, 385–388.

[75] ———, *Near optimal bounds in Freiman's theorem*, Duke Math. J. **158** (2011), no. 1, 1–12.

[76] Simon Sidon, *Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen*, Math. Ann. **106** (1932), 536–539.

[77] James Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Am. Math. Soc. **43** (1938), 377–385.

[78] Claudia A. Spiro, *Additive uniqueness sets of arithmetic functions*, J. Number Theory **42** (1992), no. 2, 232–246.

[79] Xing Sun and Andrew B. Nobel, *On the size and recovery of submatrices of ones in a random binary matrix*, J. Mach. Learn. Res. **9** (2008), 2431–2453.

[80] Terence Tao, *Every odd number greater than* 1 *is the sum of at most five primes*, Math. Comp., (submitted).

[81] Terence Tao and Van H. Vu, *Additive combinatorics*, Cambridge: Cambridge University Press, 2006.

[82] Robert C. Vaughan and Trevor D. Wooley, *Waring's problem: a survey*, Number theory for the millennium III. Proceedings of the millennial conference on number theory, Urbana-Champaign, IL, USA, May 21–26, 2000 (M. A. et al. Bennett, ed.), Natick, MA: A K Peters, 2002, pp. 301–340.