Bounds for generalized Sidon sets

Xing Peng * Rafael Tesoro † Craig Timmons ‡

Abstract

Let $\Gamma$ be an abelian group and $g \geq h \geq 2$ be integers. A set $A \subseteq \Gamma$ is a $C_h[g]$-set if given any set $X \subseteq \Gamma$ with $|X| = k$, and any set $\{k_1, \ldots, k_g\} \subseteq \Gamma$, at least one of the translates $X + k_i$ is not contained in $A$. For any $g \geq h \geq 2$, we prove that if $A \subseteq \{1, 2, \ldots, n\}$ is a $C_h[g]$-set in $\mathbb{Z}$, then $|A| \leq (g - 1)^{1/2}n^{1/2} + O(n^{1/2-1/2h})$.

We show that for any integer $n \geq 1$, there is a $C_3[3]$-set $A \subseteq \{1, 2, \ldots, n\}$ with $|A| \geq (4^{-2/3} + o(1))n^{2/3}$. We also show that for any odd prime $p$, there is a $C_3[3]$-set $A \subseteq \mathbb{F}_p^4$ with $|A| \geq p^2 - p$, which is asymptotically best possible. Using the projective norm graphs from extremal graph theory, we show that for each integer $h \geq 3$, there is a $C_h[h+1]$-set $A \subseteq \{1, 2, \ldots, n\}$ with $|A| \geq (c_h + o(1))n^{1-1/h}$. A set $A$ is a weak $C_h[g]$-set if we add the condition that the translates $X + k_1, \ldots, X + k_g$ are all pairwise disjoint. We use the probabilistic method to construct weak $C_h[g]$-sets in $\{1, 2, \ldots, n\}$ for any $g \geq h \geq 2$. Lastly we obtain upper bounds on infinite $C_h[g]$-sequences. We prove that for any infinite $C_h[g]$-sequence $A \subseteq \mathbb{N}$, we have $A(n) = O(n^{1/h}(\log n)^{-1/2})$ for infinitely many $n$, where $A(n) = |A \cap \{1, 2, \ldots, n\}|$.

1 Introduction

Given an integer $n \geq 1$, write $[n]$ for $\{1, 2, \ldots, n\}$. Let $\Gamma$ be an abelian group and $g \geq h \geq 2$ be integers. A set $A \subseteq \Gamma$ is a $C_h[g]$-set if given any set $X \subseteq \Gamma$ with $|X| = k$, and any set $\{k_1, \ldots, k_g\} \subseteq \Gamma$, at least one of the translates $X + k_i := \{x + k_i : x \in X\}$ is not contained in $A$. These sets were introduced by Erdős and Harzheim in [8], and they are a natural generalization of the well-studied Sidon sets. A Sidon set is the same as a $C_2[2]$-set.

We will always assume that $g \geq h \geq 2$. The reason for this is that if $X = \{x_1, \ldots, x_k\}$ and $K = \{k_1, \ldots, k_g\}$, then $A$ contains each of the translates $X + k_1, \ldots, X + k_g$ if and only if $A$ contains each of the translates $K + x_1, \ldots, K + x_k$.

Our starting point is a connection between $C_h[g]$-sets and the famous Zarankiewicz problem from extremal combinatorics. Given integers $m, n, s, t$ with $m \geq s \geq 1$ and $n \geq t \geq 1$, let $z(m, n, s, t)$ be the largest integer $N$ such that there is an $m \times n$ 0-1 matrix $M$, that contains $N$ 1’s, and does not contain an $s \times t$ submatrix of all 1’s. Determining $z(m, n, s, t)$ is known as the problem of Zarankiewicz.

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA. (x2peng@ucsd.edu). Research is supported in part by ONR MURI N000140810747, and AFOSR AF/SUB 552082.
†Departamento de Matemáticas, Universidad Autónoma de Madrid. 28049 Madrid, Spain. (rafael.tesoro@estudiante.uam.es)
‡Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA. (c.timmons@ucsd.edu). Research is partially supported by NSF Grant DMS-1101489 through Jacques Verstraete.
Proposition 1. Let \( \Gamma \) be a finite abelian group of order \( n \). Let \( A \subset \Gamma \) and let \( g \geq h \geq 2 \) be integers. If \( A \) is a \( C_h[g] \)-set in \( \Gamma \), then

\[
n|A| \leq z(n, n, g, h). \tag{1}
\]

To see this, let \( A \subset \Gamma \) be a \( C_h[g] \)-set where \( \Gamma = \{b_1, \ldots, b_n\} \) is a finite abelian group of order \( n \). Define an \( n \times n \) 0-1 matrix \( M \) by putting a 1 in the \((i, j)\)-entry if \( b_i + b_j \in A \), and 0 otherwise. A \( g \times h \) submatrix of all 1’s consists of a set \( X = \{x_1, \ldots, x_h\} \) of \( h \) distinct elements of \( \Gamma \), and a sequence \( k_1, \ldots, k_g \) of \( g \) distinct elements of \( \Gamma \), such that \( x_i + k_j \in A \) for all \( 1 \leq i \leq h \), \( 1 \leq j \leq g \). There is no such submatrix since \( A \) is a \( C_h[g] \)-set. Furthermore, each row of \( M \) contains \(|A|\) 1’s so that \( n|A| \leq z(n, n, g, h) \).

Füredi [11] proved that

\[
z(m, n, s, t) \leq (s - t)^{1/2}mn^{1-1/t} + tm^{2-2/t} + tn \tag{2}
\]

for any integers \( m \geq s \geq t \geq 1 \) and \( n \geq t \). Therefore, if \( A \subset \Gamma \) is a \( C_h[g] \)-set and \( \Gamma \) is a finite abelian group of order \( n \), then

\[
|A| \leq (g - h + 1)^{1/h}n^{1-1/h} + hn^{1-2/h} + h. \tag{3}
\]

If \( A \subset [n] \) is a \( C_h[g] \)-set, then it is not difficult to show that \( A \) is a \( C_h[g] \)-set in \( \mathbb{Z}_{2n} \), thus by [3],

\[
|A| \leq (g - h + 1)^{1/h}2n^{1-1/h} + h(2n)^{1-2/h} + h.
\]

Our first result improves this upper bound.

Theorem 1. If \( A \subset [n] \) is a \( C_h[g] \)-set with \( g \geq h \geq 2 \), then

\[
|A| \leq (g - 1)^{1/h}n^{1-1/h} + O \left( n^{1/2-1/2h} \right). \tag{4}
\]

This theorem is a refinement of the estimate \(|A| = O(n^{1-1/h})\) proved by Erdős and Harzheim [8]. Recall that \( C_2[2] \)-sets are Sidon sets. Theorem [4] recovers the well-known upper bound for the size of Sidon sets in \([n]\) obtained by Erdős and Turán [9]. In general, \( C_2[g] \)-sets are those sets \( A \) such that each nonzero difference \( a - a' \) with \( a, a' \in A \) appears at most \( g - 1 \) times. Theorem [4] recovers Corollary 2.1 in [7].

If \( A \subset [n] \) is a Sidon set, then for any \( g \geq 2 \), \( A \) is a \( C_2[g] \)-set. There are Sidon sets \( A \subset [n] \) with \(|A| = (1 + o(1))n^{1/2} \) thus the exponent of [4] is correct when \( h = 2 \). Motivated by constructions in extremal graph theory, we can show that [4] is correct for other values of \( h \).

Theorem 2. Let \( p \) be an odd prime and \( \alpha \in \mathbb{F}_p \) be chosen to be a quadratic non-residue if \( p \equiv 1 \pmod{4} \), and a nonzero quadratic residue otherwise. The set

\[
A = \{(x_1, x_2, x_3) \in \mathbb{F}_p^3 : x_1^2 + x_2^2 + x_3^2 = \alpha\}
\]

is a \( C_3[3] \)-set in the group \( \mathbb{F}_p^3 \) with \(|A| \geq p^2 - p \).

Corollary 1. For any integer \( n \geq 1 \), there is a \( C_3[3] \)-set \( A \subset [n] \) with

\[
|A| \geq \left( 4^{-2/3} + o(1) \right) n^{2/3}.
\]

By [3], Theorem 2 is asymptotically best possible. It is an open problem to determine the maximum size of a \( C_3[3] \)-set in \([n]\).

Proposition [1] suggests that the methods used to construct \( K_{p,h} \)-free graphs may be used to construct \( C_h[g] \)-sets. Using the norm graphs of Kollár, Rónyai, and Szabó [12], we construct \( C_h[h! + 1] \)-sets \( A \subset [n] \) with \(|A| \geq c_h n^{1-1/h} \) for each \( h \geq 2 \).
Theorem 3. Let $h \geq 2$ be an integer. For any integer $n$, there is a $C_h[h! + 1]$-set $A \subset [n]$ with

$$|A| = (1 + o(1)) \left( \frac{n}{2^{n-1}} \right)^{1-1/h}.$$ 

Using the probabilistic method we can construct sets that are almost $C_h[g]$ for all $g \geq h \geq 2$. A set $A \subset \Gamma$ is a weak $C_h[g]$-set if given any set $X \subset \Gamma$ with $|X| = k$, and any set $\{k_1, \ldots, k_g\} \subset \Gamma$ such that $X + k_1, \ldots, X + k_g$ are all pairwise disjoint, at least one of the translates $X + k_i$ is not contained in $A$. Erdős and Harzheim used the probabilistic method to construct such sets. Here we do the same but obtain a better lower bound.

Theorem 4. For any integers $g \geq h \geq 2$, there exists a weak-$C_h[g]$-set $A \subset [n]$ such that

$$|A| \geq \frac{1}{8} n^{(1-\frac{1}{h})(1-\frac{1}{g})\left(1+\frac{1}{hg}-\epsilon\right)}.$$ 

It should be noted that for $h$ fixed, Theorem 4 gives $|A| \geq n^{1-1/h}$ for $g$ sufficiently large, being a lower bound close to the exponent given in Theorem 1.

Erdős and Harzheim also proved that for any infinite $C_h[g]$-sequence $A \subset \mathbb{N}$,

$$\lim \inf_{n \to \infty} \frac{A(n)}{n^{1-1/h}} = 0.$$ 

Here $A(n) = |A \cap \{1, 2, \ldots, n\}|$. We refine this result as follows.

Theorem 5. If $A$ is an infinite $C_h[g]$-sequence with $g \geq h \geq 2$, then

$$\lim \inf_{n \to \infty} \frac{A(n)(\log n)^{1/h}}{n^{1-1/h}} = O(1),$$

where the implicit constant depends only on $g$ and $h$.

Theorem 5 was proved by Erdős [10] when $h = g = 2$.

The rest of the paper is organized as follows. In Section 2 we prove Theorem 1. We shall prove Theorem 2, Corollary 1, and Theorem 3 in Section 3. Theorem 4 is proved in Section 4 and Theorem 5 is proved in Section 5. We conclude with some open problems.

2 Proof of Theorem 1

We will use an inequality due to Cilleruelo and Tenenbaum [5].

Theorem 6 (Overlapping Theorem [5]). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\{E_j\}_{j=1}^k$ denote a family of events. For $m \geq 1$, let

$$\sigma_m := \sum_{1 \leq j_1 < \cdots < j_m \leq k} \mathbb{P}(E_{j_1} \cap \cdots \cap E_{j_m}).$$

Then for any $m \geq 1$,

$$\sigma_m \geq \frac{\sigma_1}{m!} \left( \frac{(\sigma_1 - 1)\cdots(\sigma_1 - (m-1))}{m!} \right).$$

Proof of Theorem 1. Let $A \subset [n]$ be a $C_h[g]$-set and let $B$ be any subset of $[n]$ with size at least $h$. Let $\mathbb{Y}$ be a random variable with range the positive integers and law

$$\mathbb{P}(\mathbb{Y} = m) = \begin{cases} 1/|A + B| & \text{if } m \in A + B, \\ 0 & \text{otherwise.} \end{cases}$$
For every \( b \in B \) we define the event \( E_b = \{ \omega \in \Omega: Y(\omega) \in A + b \} \), that has probability \( P(E_b) = \sum_{a \in A} P(Y = a + b) = |A|/|A + B| \). We also write
\[
\sigma_m := \sum_{\{b_1, \ldots, b_m\} \in \binom{B}{m}} P(E_{b_1} \cap \cdots \cap E_{b_m}), \quad (m \geq 1).
\]
In particular
\[
\sigma_1 = \frac{|A||B|}{|A + B|}.
\]
Let \( b_1 > \cdots > b_h \) be \( h \) fixed elements of \( B \). We can write
\[
P(E_{b_1} \cap \cdots \cap E_{b_h}) = \sum_{\{a_1, \ldots, a_h\} \in \binom{A}{h}} P(Y = a_1 + b_1 = a_2 + b_2 = \cdots = a_h + b_h)
\]
\[
= \sum_{a_1 + \{0, b_1 - b_2, b_1 - b_3, \ldots, b_1 - b_h\} \in \binom{A}{h}} \frac{1}{|A + B|},
\]
the sum extending to all \( a_1 \in A \) such that \( a_1 + \{0, b_1 - b_2, b_1 - b_3, \ldots, b_1 - b_h\} \subset A \). These are congruent \( h \)-subsets of the \( C_h[g] \)-set \( A \), thus
\[
P(E_{b_1} \cap \cdots \cap E_{b_h}) \leq \frac{g - 1}{|A + B|}.
\]
Now we use Theorem 6 to obtain
\[
\left( \frac{|B|}{h} \right)^{\frac{g - 1}{|A + B|}} \leq \sigma_h \geq \frac{\sigma_1(\sigma_1 - 1) \cdots (\sigma_1 - h + 1)}{h!} \geq \frac{\sigma_1}{h!}(\sigma_1 - (h - 1))^{h - 1},
\]
and so
\[
\frac{|B|^h}{h!} \left( \frac{g - 1}{|A + B|} \right)^{\frac{h - 1}{h!}} \leq \left( \frac{|A||B|}{h!|A + B|} \right)^{\frac{h - 1}{h!(h - 1)}} (h - 1)^{h - 1},
\]
which implies
\[
|A|^{\frac{|B|}{h\ell}} \leq |A + B| \left( (g - 1)^{\frac{1}{h\ell}} + \frac{(h - 1)|A|^{1/(h - 1)}}{|B|} \right).
\]
If we choose \( B = [l] \), by the last inequality we have
\[
|A|^{\frac{|B|}{h\ell}} \leq (n + \ell) \left( (g - 1)^{\frac{1}{h\ell}} + \frac{(h - 1)|A|^{1/(h - 1)}}{\ell + 1} \right). \tag{5}
\]
We first take \( \ell = n \) and use \( |A| \leq n \) in the right side, getting
\[
|A|^{\frac{|B|}{h\ell}} = O(n) \implies |A|^{1/(h - 1)} = O(n^{1/h}).
\]
Inserting this in the second member of (5) we obtain
\[
|A|^{\frac{|B|}{h\ell}} \leq (g - 1)^{\frac{1}{h\ell}} + O(\ell) + O\left( \frac{n^{1 + 1/h}}{\ell + 1} \right) + O(n^{1/h}).
\]
To minimize this last upper bound we choose \( \ell \approx n^{1/2 + 1/2h} \). Then we can write
\[
|A|^{\frac{|B|}{h\ell}} \leq (g - 1)^{\frac{1}{h\ell}} + O\left( n^{1/2 + 1/2h} \right) = (g - 1)^{\frac{1}{h\ell}} + O\left( \left( n^{1/2h} - \frac{1}{2} \right) \right),
\]
which yields
\[
|A| \leq (g - 1)^{1/h}n^{1 - 1/h} \left( 1 + O\left( n^{1/2h - 1/2} \right) \right)^{1 - 1/h} = (g - 1)^{1/h}n^{1 - 1/h} + O\left( n^{1/2 - 1/2h} \right)\),
\]
as we claimed.
3 Proof of Theorem 2, Corollary 1, and Theorem 3

Proof of Theorem 2. Recall that we choose \( \alpha \in \mathbb{F}_p \) as a quadratic non-residue when \( p \equiv 1 \pmod{4} \) and a nonzero quadratic residue otherwise. Let \( G = (V, E) \) be a graph with vertex set \( V = \mathbb{F}_p^3 \). For \( x = (x_1, x_2, x_3) \) and \( y = (y_1, y_2, y_3) \) we have \( (x, y) \in E(G) \) if and only if

\[
\sum_{i=1}^{3} (x_i - y_i)^2 = \alpha.
\]

The graph \( G \) is \( K_{3,3} \)-free as shown by Brown [4]. Define

\[
S(\alpha) = \{(x_1, x_2, x_3) \in \mathbb{F}_p^3 : x_1^2 + x_2^2 + x_3^2 = \alpha\}.
\]

Let \( X = \{x, y, z\} \subset \mathbb{F}_p^3 \). Suppose \( X + a \subset S(\alpha) \) for some \( a \in \mathbb{F}_p^3 \). We first show \( -a \not\in \{x, y, z\} \). If \( x = -a \) then we get \( 0 = \alpha \) as \( x + a \in S(\alpha) \). This is a contradiction since we have chosen \( \alpha \) so that \( \alpha \neq 0 \). Therefore \( -a \neq x \) and similarly, \( -a \neq y \) and \( -a \neq z \). By definition, \((x, -a), (y, -a), \) and \((z, -a)\) are three edges in \( G \), which tell us that \( a \) is a common neighbor of \( x, y \), and \( z \). Assume that there are three translates \( X + a, X + b, X + c \) contained in \( S(\alpha) \) for distinct \( a, b, c \in \mathbb{F}_p^3 \). We have \( \{x, y, z\} \cap \{a, b, c\} = \emptyset \), and so \( L = \{x, y, z\} \) and \( R = \{a, b, c\} \) form a \( K_{3,3} \) in \( G \). However \( G \) is \( K_{3,3} \)-free, a contradiction. Thus there are at most two elements \( a, b \in \mathbb{F}_p^3 \) such that the translates \( X + a \) and \( X + b \) are contained in \( S(\alpha) \). This holds for every \( X \subset \mathbb{F}_p^3 \) with \( |X| = 3 \). We proved Theorem 2. \( \square \)

Next we prove Corollary 1 and Theorem 3. Both results rely on the following lemma.

Lemma 1. Let \( p \) be a prime and \( d \geq 1 \) be an integer. Define \( \phi : \mathbb{F}_p^d \to \mathbb{Z} \) by

\[
\phi((x_1, \ldots, x_d)) = x_1 + 2px_2 + (2p)^2x_3 + \cdots + (2p)^{d-1}x_d.
\]

where \( 0 \leq x_i \leq p - 1 \). The map \( \phi \) is 1-to-1 and furthermore, for any \( x, y, z, t \in \mathbb{F}_p^d \), we have \( x + y = z + t \) if and only if \( \phi(x) + \phi(y) = \phi(z) + \phi(t) \).

The proof of Lemma 1 is not difficult. In the language of additive combinatorics, the map \( \phi \) is a Freiman isomorphism of order 2 (see [15], Chapter 5, Section 3).

Proof of Corollary 1. Let \( n \) be a large integer. Choose an odd prime \( p \) with \( 4p^3 \leq n \) and \( p \) as large as possible. Let \( S \subset \mathbb{F}_p^3 \) be a \( C_3[3] \)-set in \( \mathbb{F}_p^3 \) with \( |S| \geq p^2 - p \) guaranteed by Theorem 2. Consider \( A = \phi(S) \) where \( \phi : \mathbb{F}_p^3 \to \mathbb{Z} \) is the map

\[
\phi((x_1, x_2, x_3)) = x_1 + 2px_2 + 4p^2x_3
\]

Here \( x_i \) is chosen so that \( 0 \leq x_i \leq p - 1 \). By Lemma 1, \( A \) is a \( C_3[3] \)-set. If \( a \in A \), then \( a \leq (p - 1)(1 + 2p + 4p^2) \leq 4p^3 \leq n \) so \( A \subset [n] \). Since \( \phi \) is 1-to-1, \( |A| \geq p^2 - p \). For large enough \( n \), there is always a prime between \( (n/4)^{1/3} - (n/4)^{2/3} \) and \( (n/4)^{1/3} \) for some \( \theta < 1 \). The results of [2] show that one can take \( \theta = 0.525 \). Therefore, \( |A| \geq (n/4)^{2/3} - O(n^{1/6}) = (1 + o(1))(n/4)^{2/3} \).

\( \square \)

Proof of Theorem 3. Let \( q \) be a prime power and \( h \geq 2 \) be an integer. Let \( N : \mathbb{F}_{q^h} \to \mathbb{F}_q \) be the norm map defined by

\[
N(x) = x^{1+q+q^2+\cdots+q^{h-1}}.
\]

Let \( A = \{ x \in \mathbb{F}_{q^h} : N(x) = 1 \} \). The norm map \( N \) is a group homomorphism that maps \( \mathbb{F}_{q^h}^* \) onto \( \mathbb{F}_q^* \). This implies \( q^{h-1} \mid |A| = q - 1 \) so \( |A| = \frac{q^{h-1}}{q-1} \). We now show that \( A \) is a \( C_h[h! + 1] \)-set in the group \( \mathbb{F}_{q^h}^* \).
Suppose \( X = \{x_1, \ldots, x_h\} \subset \mathbb{F}_{q^h} \). It follows from Theorem 3.3 of [12] that there are at most \( h! \) elements \( k \in \mathbb{F}_{q^h} \) such that

\[
N(k + x_i) = 1
\]

for all \( 1 \leq i \leq h \). Therefore, given any set \( \{k_1, \ldots, k_{h+1}\} \subset \mathbb{F}_{q^h} \), at least one of the translates \( X + k_i \) is not contained in \( A \).

Let \( \psi : \mathbb{F}_{q^h} \rightarrow \mathbb{Z}_q^h \) be a group isomorphism mapping the additive group \( \mathbb{F}_{q^h} \) onto the direct product \( \mathbb{Z}_q^h \). Let \( \phi : \mathbb{Z}_q^h \rightarrow \mathbb{Z} \) be the map

\[
\phi(x_1, \ldots, x_h) = x_1 + (2q)x_2 + \cdots + (2q)^{h-1}x_h
\]

where \( 0 \leq x_i \leq q - 1 \). By Lemma [1], \( A' := \phi(\psi(A)) \) is a \( C_h[h! + 1] \)-set. The set \( A' \) has \( \frac{q^h - 1}{q - 1} \) elements and is contained in the set \( [2^{h-1}q^h] \). By the same argument used to prove Corollary [1] we can choose a prime power \( q \) given a large enough integer \( n \) to obtain a \( C_h[h! + 1] \)-set in \( [n] \) with size \( (1 + o(1)) \left( \frac{n}{q^{h-1}} \right)^{1-1/h} \).

### 4 Proof of Theorem 4

The proof in this section uses the probabilistic method combined with the deletion technique. These ideas have appeared before in the literature, see for example [11 §3], [14], and [6].

We say that \( m \in S \) is \((h, g)\)-bad for \( S \) if there exist \( m_1 < \cdots < m_{g-1} \), with \( m_i < m \), and there exist \( \ell_1 < \ell_2 < \cdots < \ell_{h-1} \) such that the sums \( \{m_1, \cdots, m_{g-1}, m\} + \{0, \ell_1, \cdots, \ell_{h-1}\} \) are \( gh \) distinct elements of \( S \).

We define \( S_{bad} \) the set of \((h, g)\)-bad elements for \( S \). It is clear that for any set \( S \), the set

\[
S_{C_h[g]} = S \setminus S_{bad},
\]

is weak-\( C_h[g] \)-set with cardinality \( |S_{C_h[g]}| = |S| - |S_{bad}| \).

Define \( p \) as the number such that \( 2pn = n^{g+h-1}(2p)^{gh} \). It is straightforward to check that

\[
np = \frac{1}{2}n^{\left(1-\frac{1}{h}\right)} \left(1+\frac{1}{m^{\frac{1}{h-1}}}\right).
\]

We will prove that except for finitely many \( n \) there exist a set \( S \subset [n] \) such that

\[
|S| \geq \frac{np}{2} \quad \text{and} \quad |S_{bad}| \leq \frac{np}{4}.
\]

Note that for such a set we have

\[
|S_{C_h[g]}| = |S| - |S_{bad}| > \frac{np}{4} = \frac{1}{8}n^{\left(1-\frac{1}{h}\right)} \left(1+\frac{1}{m^{\frac{1}{h-1}}}\right),
\]

for all sufficiently large \( n \) and \( A = S_{C_h[g]} \) satisfies the conditions of Theorem 4.

Indeed we will prove that with probability at least \( 1/4 \), a random set \( S \) in \([n] \) satisfies [7] if each element in \([n] \) is independently chosen to be in \( S \) with probability \( p \).

Next we obtain estimates for the random variables \(|S|\) and \(|S_{bad}|\).

If \( m \) is \((h, g)\)-bad then the \( gh \) sums \( \{m_1, \cdots, m_{g-1}, m\} + \{0, \ell_1, \cdots, \ell_{h-1}\} \) are all distinct elements of \( S \) and so

\[
\mathbb{P}(\{m_1, \ldots, m_{g-1}, m\} + \{0, \ell_1, \cdots, \ell_{h-1}\} \subset S) = p^{gh}.
\]
Hence
\[
P(m \text{ is } (h, g)\text{-bad}) \leq \sum_{1 \leq m_1 < \cdots < m_g < m} p^{gh} \leq \binom{m}{h-1} \binom{n}{g-1} p^{gh} < n^{g+2-h} p^{gh}
\]
which implies
\[
E(|S_{bad}|) \leq \sum_{1 \leq m \leq n} P(m \text{ is } g\text{-bad}) \leq n^{g+h-1} p^{gh}.
\]
On the one hand by Markov’s inequality we have
\[
P\left(|S_{bad}| > \frac{np}{4}\right) = P\left(|S_{bad}| > \frac{n^{g+h-1}(2p)^{gh}}{8}\right)
\]
\[
= P\left(|S_{bad}| > 2^{gh-3} n^{g+h-1} p^{gh}\right) \leq P\left(|S_{bad}| > 2E(|S_{bad}|)\right) < 1/2.
\]
On the other hand, using that \(E(|S|) = np\) and \(\text{Var}(|S|) = np(1 - p)\) and applying Chebychev’s inequality we have
\[
P\left(|S| < \frac{np}{2}\right) = P\left(|S| < \frac{E(|S|)}{2}\right) < P\left(|S - E(|S|)| > \frac{E(|S|)}{2}\right)
\]
\[
< \frac{4\text{Var}(|S|)}{(E(|S|))^2} = \frac{4np(1 - p)}{(pn)^2} < \frac{4}{pn} < \frac{1}{4},
\]
except for finitely many \(n\). By (8) and (9) we have
\[
P(|S| \geq np/2 \text{ and } |S_{bad}| \leq np/4) \geq 1 - (1/2 + 1/4) \geq 1/4,
\]
as we wanted.

5 Proof of Theorem 5

In order to simplify notation, when \(f(n) = O(g(n))\) we write \(f(n) \ll g(n)\) or \(g(n) \gg f(n)\) through this section.

Proof of Theorem 5. Let \(A\) be an infinite \(C_h[g]\)-sequence. For a positive integer \(N\), let \([N^2]\) denote all the positive integers less or equal to \(N^2\). We divide \([N^2]\) into equally sized intervals
\[
I_\nu := [(\nu - 1)N, \nu N], \quad \nu = 1, \cdots, N.
\]
Let \(C\) denote the collection of all \(h\)-subsets of \([N^2]\) that are included in one of the intervals \(I_\nu\):
\[
C := \left\{ C \in \left(\frac{[N^2]}{h}\right): C \subset I_\nu \text{ for some } \nu \right\}.
\]
We say that the sets in the collection \(C\) are “small” as their diameter is at most \(N\). We classify the elements of \(C\) so that each class groups all the sets that are pairwise congruent. Each class \(\alpha\) contains a set \(C_\alpha\) that contains 1, and the remaining \(h - 1\) elements of \(C_\alpha\) can be chosen in \(\binom{N-1}{h-1}\) different ways; each of the choices determines a class different from the others. Then the number of classes is
\[
\binom{N-1}{h-1}.
\]
Let $A_\nu$ denote the size of $A \cap I_\nu$, we have $A_\nu = A(\nu N) - A((\nu - 1)N)$, where $A(x) := |\{a \in A : a \leq x\}|$ is the counting function of the sequence.

One the one hand as $A$ is a $C_h[g]$-sequence then in every class of $\mathcal{C}$ there are at most $g - 1$ subsets of $A$. Hence we have the following upper bound for the total number of “small” subsets of $A$ that belong to $\mathcal{C}$

$$
\sum_{\nu=1}^{N} \frac{A_\nu}{h} \leq \binom{N-1}{h-1}(g-1) \ll N^{h-1} \quad (N \to \infty),
$$

Now we prove by induction in $h$ that

$$
\sum_{\nu=1}^{N} A_\nu^h \ll N^{h-1} \quad (N \to \infty). \tag{10}
$$

For $h = 2$ we know by Theorem 4 that $A(N^2) \ll N$, so

$$
\sum_{\nu=1}^{N} A_\nu^2 = 2 \sum_{\nu=1}^{N} \left(\frac{A_\nu}{2}\right) + \sum_{\nu=1}^{N} A_\nu \ll N + A(N^2) \ll N.
$$

If (10) holds for all exponents up to $h - 1$, then

$$
\sum_{\nu=1}^{N} A_\nu^h = h! \sum_{\nu=1}^{N} \left(\frac{A_\nu}{h}\right) + O \left(\sum_{\nu=1}^{N} A_\nu^{h-1}\right) \ll N^{h-1} + N^{h-2}, \quad (N \to \infty),
$$

thus it also holds for $h$. Using (10) and Hölder inequality we can write

$$
\sum_{\nu=1}^{N} A_\nu \left(\frac{1}{\nu}\right)^{1-1/h} \leq \left(\sum_{\nu=1}^{N} A_\nu^h\right)^{1/h} \left(\sum_{\nu=1}^{N} \frac{1}{\nu}\right)^{1-1/h} \ll (N \log N)^{1-1/h}, \quad (N \to \infty). \tag{11}
$$

On the other hand as $\sum_{\nu \leq t} A_\nu = A(tN)$ and summing by parts

$$
\sum_{\nu=1}^{N} A_\nu \left(\frac{1}{\nu}\right)^{1-1/h} = \frac{A(N^2)}{N^{1-1/h}} + \int_{1}^{N} \frac{A(tN)}{t^{2-1/h}} dt.
$$

In this sum the first summand is bounded by Theorem 4 as follows

$$
\frac{A(N^2)}{N^{1-1/h}} \ll N^{2(1-1/h)}/N^{(1-1/h)} = N^{(1-1/h)}, \quad (N \to \infty) \tag{12}
$$

and as consequence we shall prove next that the second summand is the main term in the sum. Let us write

$$
\tau(m) := \inf_{n \geq m} \frac{A(n)(\log n)^{1/h}}{n^{1-1/h}}.
$$

For $N \geq m$ and $t \geq 1$ we have

$$
A(tN) = \frac{A(tN)(\log(tN))^{1/h}(tN)^{1-1/h}}{(tN)^{1-1/h}(\log(tN))^{1/h}} \geq \tau(m) \frac{t^{1-1/h}N^{1-1/h}}{(\log N)^{1/h}}.
$$
Thus for $N \geq m$ we have
\[ \int_1^N \frac{A(tN)}{t^{2-1/h}} \, dt \gg \frac{\tau(m)N^{1-1/h}}{(\log N)^{1/h}} \int_1^N \frac{1}{t} \, dt \gg \tau(m)(N \log N)^{1-1/h}, \]
and so by (12)
\[ \sum_{\nu=1}^N A_{\nu} \left( \frac{1}{\nu} \right)^{1-1/h} \gg \int_1^N \frac{A(tN)}{t^{2-1/h}} \, dt \gg \tau(m)(N \log N)^{1-1/h}. \]
Inserting (11) we have $\lim_{m \to \infty} \tau(m) \ll 1$, that is what we wanted to prove.

6 Open problems

In this final section we mention several open problems.

**Problem 1:** Determine the maximum size of a $C_3[3]$-set contained in $[n]$.

Our results show that if $A \subset [n]$ is a $C_3[3]$-set of maximum size then
\[ (4^{-2/3} + o(1))n^{2/3} \leq |A| \leq (2^{1/3} + o(1))n^{2/3}. \]
It seems likely that both of these bounds can be improved. Perhaps the correct answer is $(1 + o(1))n^{2/3}$.

**Problem 2:** Remove the condition weak in Theorem 4.

A much harder problem is the following.

**Problem 3:** Construct $C_h[g]$-sets in $[n]$ with the order $n^{1-\frac{1}{h}}$ for each $g \geq h \geq 3$.

For $g \geq h = 3$ and $g \geq h! + 1$, we constructed $C_h[g]$-sets in $[n]$ whose sizes matches the order given by Theorem 4. We believe for any other $g$ and $h$ the upper bound by Theorem 4 gives the correct exponent. We note that solving Problem 3 would imply $z(n, n, g, h) \geq C(g, h)n^{1-1/h}$ for some constant $C$ (see Proposition 1).

**Problem 4:** Construct an infinite $C_h[g]$-sequence $A \in \mathbb{N}$ which has counting function $A(n) \gg n^{(1-\frac{1}{g})(1-\frac{1}{h})(1 + \frac{1}{m}) + o(1)}$ for all $n$.

We have found technical difficulties to deal with Problems 2 and 4, which were suggested to the second author by Javier Cilleruelo.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method*, John Wiley & Sons Inc., Third edition (2008).

[2] R. C. Baker, G. Harman, and J. Pintz, *The difference between consecutive primes, II*, Proc. London Math. Soc., 83-(3) (2001), 532-562.

[3] R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv., 37 (1962), 141-147.

[4] W. G. Brown, *On graphs that do not contain a Thomsen graph*, Canad. Math. Bull., 9 (1966), 281-289.

[5] J. Cilleruelo and G. Tenenbaum, *An Overlapping theorem with applications*, Publicacions matemàtiques, Extra (2007) 107-118.
[6] J. Cilleruelo, *Probabilistic constructions of $B_2[g]$ sequences*, Acta Mathematica Sinica, English Series **26-7** (2010) 1309-1314.

[7] J. Cilleruelo, *Sidon sets in $\mathbb{N}^d$*, Journal of Combinatorial Theory, Series A, **117** (2010) 857-871.

[8] P. Erdős and E. Harzheim, *Congruent subsets of infinite sets of natural numbers*, J. Reine Angew. Math., **367** (1986) 207-214.

[9] P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, Journal of the London Mathematical Society, **16** (1941) 212-215.

[10] P. Erdős, *Einige Bemerkungen zur Arbeit von A. Stöhr: “Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe”* J. Reine Angew. Math., **197** (1957) 216-219.

[11] Z. Füredi, *An upper bound on Zarankiewicz’ Problem*, Combinatorics, Probability and Computing, **5** (1996), 29-33.

[12] J. Kollár, L. Rónyai, and T. Szabó, *Norm-graphs and bipartite Turán numbers*, Combinatorica **16-3** (1996) 399-406.

[13] R. Lidl and H. Niederreiter, *Finite Fields*, Encyclopedia of Mathematics and its Applications 20, Cambridge University Press 2nd ed., 1997.

[14] J. Spencer and P. Tetali, *Sidon Sets with Small Gaps*, Discrete Probability and Algorithms,

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