

Zero sums in restricted sequences

Niranjan Balachandran* and Eshita Mazumdar†

March 16, 2021

Abstract

Suppose $A \subset \mathbb{Z}_n \setminus \{0\}$. A sequence $\mathbf{x} = (x_1, \dots, x_m)$ of elements of \mathbb{Z}_n is called an *A-weighted Davenport Z-sequence* if there exists $\mathbf{a} := (a_1, \dots, a_m) \in (A \cup \{0\})^m \setminus \mathbf{0}_m$ such that $\sum_i a_i x_i = 0$, where $\mathbf{0}_m = (0, \dots, 0) \in \mathbb{Z}_n^m$. Similarly, the sequence \mathbf{x} is called an *A-weighted Erdős Z-sequence* if there exists $\mathbf{a} = (a_1, \dots, a_m) \in (A \cup \{0\})^m$ with $|\text{Supp}(\mathbf{a})| = n$, such that $\sum_i a_i x_i = 0$, where $\text{Supp}(\mathbf{a}) := \{i : a_i \neq 0\}$. A \mathbb{Z}_n -sequence \mathbf{x} is called *k-restricted* if no element of \mathbb{Z}_n appears more than *k* times in \mathbf{x} . In this paper, we study the problem of determining the least value of *m* for which a *k*-restricted \mathbb{Z}_n -sequence of length *m* is an *A*-weighted Davenport Z-sequence (resp. an *A*-weighted Erdős Z-sequence). We also consider the same problem for random \mathbb{Z}_n -sequences and some very natural choices of the set *A*.

Keywords: Davenport Constant, Erdős constant, Zero-Sum problems.

2010 AMS Classification Code: 11B50, 11B75, 11P70, 11K99.

1 Introduction

In this paper, k, ℓ, m, n shall always refer to positive integers. By $[n]$ we shall mean the set $\{1, \dots, n\}$, and for integers $a < b$, $[a, b]$ shall denote the set $\{a, a + 1, \dots, b\}$. By \mathbb{Z}_n we shall denote the cyclic group of order n .

Throughout this paper, we shall use the Landau asymptotic notation: For functions f, g , we write $f(n) = O(g(n))$ if there exists an absolute constant $C > 0$ and an integer n_0 such that for all $n \geq n_0$, $|f(n)| \leq C|g(n)|$. We write $f = \Omega(g)$ if $g = O(f)$, and we write $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$. We also write $f = o(g)$ if $\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$. For a real $a > 1$, we write $\log_a n$ to denote the logarithm of n to the base a .

*Dept.of Mathematics, IIT Bombay, Mumbai. email: niranj (at) math.iitb.ac.in. Supported by grant 12IR-CCSG016, IRCC, IIT Bombay

†Center for Combinatorics, Nankai University, Tianjin, China. email: eshitamazumdar@yahoo.com. Supported by NSFC with grant no. 11681217

For a finite abelian group $(G, +)$ a G -sequence of length m shall refer to a sequence $\mathbf{x} := (x_1, \dots, x_m)$ with $x_i \in G$ for all i . If some $x \in G$ appears r times in \mathbf{x} then we say that x has *multiplicity* r in \mathbf{x} . In particular, if x does not appear in \mathbf{x} then x has multiplicity zero in \mathbf{x} . For a sequence $\mathbf{x} = (x_1, \dots, x_m)$, and for a subset $I \subseteq [m]$ of the set of indices, we shall denote by \mathbf{x}_I the sum $\sum_{i \in I} x_i$. If $I = \emptyset$ then this corresponds to the empty sum. For sequences $\mathbf{x} = (x_1, \dots, x_m), \mathbf{y} = (y_1, \dots, y_m)$ of the same length, we shall denote by $\langle \mathbf{x}, \mathbf{y} \rangle$ the sum $\sum_{i \in [m]} x_i y_i$.

If $\mathbf{a} = (a_1, \dots, a_m)$ and $\mathbf{x} = (x_1, \dots, x_m)$ are \mathbb{Z}_n -sequences, then $\mathbf{a} \cdot \mathbf{x}$ shall denote the sequence $(a_1 x_1, \dots, a_m x_m)$, where the multiplication is from the ring structure of \mathbb{Z}_n . For $\mathbf{a} \in \mathbb{Z}_n^m$, we define $\text{Supp}(\mathbf{a}) := \{i : a_i \neq 0\}$. Finally, by $\mathbf{0}_m$ we shall mean the zero sequence $(0, \dots, 0) \in (\mathbb{Z}_n)^m$.

For $A \subseteq \mathbb{Z}_n \setminus \{0\}$, a sequence $\mathbf{x} = (x_1, \dots, x_m)$ of elements of \mathbb{Z}_n is called an A -weighted Davenport Z-sequence if there exists $\mathbf{a} = (a_1, \dots, a_m) \in (A \cup \{0\})^m \setminus \mathbf{0}_m$ such that $\langle \mathbf{a}, \mathbf{x} \rangle = 0$. In words, there is a choice of ‘coefficients’ from A , not all zero, such that the corresponding ‘linear combination’ equals zero. Similarly, the sequence \mathbf{x} is called an A -weighted Erdős Z-sequence if there exists $\mathbf{a} := (a_1, \dots, a_m) \in (A \cup \{0\})^m$ with $|\text{Supp}(\mathbf{a})| = n$, such that $\sum_i a_i x_i = 0$. When $A = \{a\}$ for some a co-prime to n , we shall refer to such a sequence simply as a Davenport Z-sequence (resp. an Erdős Z-sequence). When the set A is clear from the context, we shall drop any mention of the set A and refer simply to weighted Davenport Z-sequences (resp. weighted Erdős Z-sequences).

The notion of a weighted Davenport Z-sequence draws its motivation from a well-studied combinatorial invariant associated with a finite abelian group G , namely the *Davenport constant of G* (denoted $D(G)$), which is defined as the least positive integer m such that every G -sequence of length m admits a non-trivial finite subsequence whose sum is zero in G . A generalization of this notion in [1, 2]¹ introduces a weighted version of this combinatorial invariant as follows: For a given $A \subseteq \mathbb{Z}_n \setminus \{0\}$, the *weighted Davenport constant of \mathbb{Z}_n with respect to weight set A* (denoted by $D_A(\mathbb{Z}_n)$) is the least integer m such that for every \mathbb{Z}_n -sequence $\mathbf{x} = (x_1, \dots, x_m)$ of length m , there exists $\mathbf{a} \in (A \cup \{0\})^m \setminus \mathbf{0}_m$ satisfying $\langle \mathbf{a}, \mathbf{x} \rangle = 0$.

The following results are well known:

- If $a \in \mathbb{Z}_n^*$ and $A = \{a\}$ then any sequence of length n is an A -weighted Davenport Z-sequence. The bound n is sharp, as is witnessed by the sequence $(\underbrace{1, \dots, 1}_{n-1 \text{ times}})$. This is folklore and is a simple exercise on the application of the Pigeonhole Principle.
- (See [1]) For the set $A = \{1, -1\}$, any \mathbb{Z}_n -sequence of length $\lfloor \log_2 n \rfloor + 1$ is a weighted Davenport Z-sequence, and this result is again, best possible: The sequence $(1, 2, \dots, 2^{k-1})$, for $k = \lfloor \log_2 n \rfloor$, is not a weighted Davenport Z-sequence.
- (See [6], [8]) If $n = q_1 \cdots q_a$ is the product of a primes (not necessarily distinct) and $A = \mathbb{Z}_n^*$, the group of units of the ring \mathbb{Z}_n then any sequence of length $a + 1$ is an A -weighted Davenport Z-sequence, and again, this result is best possible: The sequence $(1, q_1, q_1 q_2, \dots, q_1 q_2 \cdots q_{a-1})$ is not a weighted Davenport Z-sequence.

¹While the generalization was for all finite abelian groups, we shall restrict our attention to the cyclic group \mathbb{Z}_n .

The notion of an Erdős Z-sequence draws its motivation from the following non-trivial theorem of Erdős-Ginzburg-Ziv [5]: Every \mathbb{Z}_n -sequence of length $2n - 1$ admits a subsequence of size n whose sum equals zero. Thus, in our language, every \mathbb{Z}_n -sequence of length $2n - 1$ is an Erdős Z-sequence, and again, this is best possible since the sequence $(\underbrace{0, \dots, 0}_{n-1 \text{ times}}, \underbrace{1, \dots, 1}_{n-1 \text{ times}})$ is not an Erdős Z-sequence. For

an arbitrary set $A \subseteq \mathbb{Z}_n \setminus \{0\}$, one can analogously define the parameter $E_A(\mathbb{Z}_n)$ as the least integer m such that for every \mathbb{Z}_n -sequence $\mathbf{x} = (x_1, \dots, x_m)$, there exists $\mathbf{a} = (a_1, \dots, a_m) \in (A \cup \{0\})^m$ with $|\text{Supp}(\mathbf{a})| = n$ such that $\langle \mathbf{a}, \mathbf{x} \rangle = 0$. It turns out [14] that $E_A(\mathbb{Z}_n) = D_A(\mathbb{Z}_n) + n - 1$. In particular, it follows that for $A = \{1, -1\}$, every sequence of length $n + \lfloor \log_2 n \rfloor$ is an A -weighted Erdős Z-sequence and this result is best possible.

One distinct feature of the aforementioned results pertaining to Erdős Z-sequences (and some others) is that the exact values of the invariants $D_A(\mathbb{Z}_n)$ are witnessed by highly structured sequences. For instance, for $A = \{a\}$ (for any $a \in \mathbb{Z}_n^*$), the maximal sequences \mathbf{x} that are not Davenport Z-sequences are necessarily of the form $\mathbf{x} = (\underbrace{x, \dots, x}_{n-1 \text{ times}})$ for some $x \in \mathbb{Z}_n^*$. However, if we

restrict our attention to \mathbb{Z}_n -sequences with a bound on the number of incidences of any particular element of $\mathbb{Z}_n \setminus \{0\}$, then it is conceivable that among this restricted class of sequences, the minimum value of m for which every restricted \mathbb{Z}_n -sequence (restricted in this sense, which we shall make more precise soon) of length m is an A -weighted Davenport (resp. Erdős) Z-sequence, might be considerably smaller. And this is the focal point of this paper: How large must a *restricted* sequence be if it is to be a Davenport (resp. Erdős) Z-sequence?

To make precise what we mean by the word restricted, we define the following set of \mathbb{Z}_n -sequences:

$$\mathcal{X}_k(n, m) := \{\mathbf{x} \in (\mathbb{Z}_n)^m : \text{every } x \in \mathbb{Z}_n \text{ has multiplicity at most } k \text{ in } \mathbf{x}\}.$$

We shall resort to some abuse of notation and denote $\mathcal{X}_k(n, m)$ by $\mathcal{X}_k(m)$ for simplicity. The preceding discussion leads us to the following natural problem:

Problem: Let $A \subset \mathbb{Z}_n \setminus \{0\}$ and suppose $1 \leq k \leq n - 1$ is a positive integer. Determine the least integer m such that every $\mathbf{x} \in \mathcal{X}_k(m)$ is an A -weighted Davenport (Erdős) Z-sequence.

This problem is part of a larger umbrella of problems that usually go by the name of Inverse Zero-Sum problems (see [7]). Indeed, some results in [7] are similar to some of the ones in this paper. However, the focus of the results is not the same as ours.

We set up some further notation. For $k \in \mathbb{N}$, by $\mathfrak{s}^{(k)}(\mathbb{Z}_n)$, we shall mean the least integer m such that every $\mathbf{x} \in \mathcal{X}_k(m)$ is an Erdős Z-sequence. The case $k = 1$ is usually awarded greater status and is referred to as the *Harborth constant* of \mathbb{Z}_n (see [9] for more results on the Harborth constant) and we shall not delve into that in this paper.

Our first result of this paper addresses the aforementioned problem posed above for $k \geq 2$:

Theorem 1. *Suppose $n - 1 \geq k \geq 2$ is an integer.*

1. $\mathfrak{s}^{(2)}(\mathbb{Z}_n) = n + 2$ for $n \geq 5$.

2. For any prime $p \geq 5$, $s^{(k)}(\mathbb{Z}_p) \leq p + k$. Furthermore, for each k , there exists an integer $p_0(k)$ such that $s^{(k)}(\mathbb{Z}_p) = p + k$ for all primes $p \geq p_0(k)$.
3. There exist constants $c, C > 0$ such that every sequence $\mathbf{x} \in \mathcal{X}_k(C\sqrt{nk})$ is a Davenport Z -sequence. Furthermore, there exist sequences $\mathbf{y} \in \mathcal{X}_k(c\sqrt{nk})$ that are not Davenport Z -sequences.

We now turn to the second objective of this paper, namely, the study of how these zero-sum related invariants behave if the \mathbb{Z}_n -sequence is picked *randomly*. More precisely, suppose $\mathfrak{X}_m = (X_1, \dots, X_m)$ is a \mathbb{Z}_n -sequence where each X_i is picked independently and uniformly at random from \mathbb{Z}_n . One of the motivations to study random \mathbb{Z}_n -sequences from the zero-sum problem perspective is that relatively short random sequences are very likely to be members of $\mathcal{X}_k(m)$. Indeed, for any fixed $k \geq 2$, the probability that a random sequence \mathfrak{X}_m contains some element of \mathbb{Z}_n with multiplicity at least $k + 1$ is at most $\binom{m}{k+1}(1/n)^k \leq \frac{m^{k+1}}{n^k} \rightarrow 0$ if $m = o(n^{k/(k+1)})$. Hence, if $m = o(n^{k/(k+1)})$ the sequence \mathfrak{X}_m is an element of $\mathcal{X}_k(m)$ with high probability (*whp for short*), i.e., with probability approaching 1 as $n \rightarrow \infty$. One also expects that with high probability, the random sequence \mathfrak{X}_m demonstrates no algebraic structure, so it is conceivable that random sequences of length significantly shorter than $D_A(\mathbb{Z}_n)$ are very likely to be weighted Davenport (resp. Erdős) Z -sequences. This leads us naturally to the following problem:

Problem: Let $A \subset \mathbb{Z}_n \setminus \{0\}$. Determine the least m such that a random \mathbb{Z}_n -sequence of length m is a weighted Davenport (Erdős) Z -sequence *whp*.

Our other results of this paper address this problem for weighted Erdős Z -sequences, and weighted Davenport Z -sequences respectively for some natural choices of the set A :

Theorem 2. Let $\mathfrak{X}_m = (X_1, \dots, X_m)$ be a random \mathbb{Z}_n -sequence. Then *whp* (as $n \rightarrow \infty$) the following statements hold:

1. \mathfrak{X}_{n+2} is an Erdős Z -sequence.
2. Let $A = \{a, b\}$. Suppose $A \cap \mathbb{Z}_n^* \neq \emptyset$ and suppose also that $a + b, a - b \in \mathbb{Z}_n^*$. Then \mathfrak{X}_n is an A -weighted Erdős Z -sequence.
3. Let $A = \{1, -1\}$. Then \mathfrak{X}_{n+1} is an A -weighted Erdős Z -sequence.

Theorem 3. Let $\mathfrak{X}_m = (X_1, \dots, X_m)$ be a random \mathbb{Z}_n -sequence and suppose $\omega(n)$ is a function that satisfies $\omega(n) \rightarrow \infty$ as $n \rightarrow \infty$.

1. The following hold *whp* (as $n \rightarrow \infty$) :

$$\begin{aligned} \mathfrak{X}_m \text{ is a Davenport } Z\text{-sequence if } m &\geq \log_2 n + \omega(n), \\ \mathfrak{X}_m \text{ is not a Davenport } Z\text{-sequence if } m &\leq \log_2 n - \omega(n). \end{aligned}$$

2. Suppose $A = \{-1, 1\}$. Then *whp* (as $n \rightarrow \infty$) the following hold:

$$\begin{aligned} \mathfrak{X}_m \text{ is an } A\text{-weighted Davenport } Z\text{-sequence if } m &\geq \log_3 n + \omega(n). \\ \mathfrak{X}_m \text{ is not an } A\text{-weighted Davenport } Z\text{-sequence if } m &\leq \log_3 n - \omega(n). \end{aligned}$$

3. Suppose $n = p_1 \cdots p_r$ where p_i are distinct odd primes and let $A = \mathbb{Z}_n^*$. Then if $m \geq \omega(n)$ then \mathfrak{X}_m is an A -weighted Davenport Z -sequence whp (as $n \rightarrow \infty$).

Note that in the last part of Theorem 3 (where the weight set is $A = \mathbb{Z}_n^*$) we do not have an analogous statement (as in the previous parts) with a lower bound for m for \mathfrak{X}_m to not be a weighted Davenport Z -sequence whp; indeed, as we shall see later, if p_1 is sufficiently large, a random sequence of length 3 is a \mathbb{Z}_n^* -weighted Davenport Z -sequence whp. This will follow as a consequence of our proof of Theorem 3.

An interesting aspect of our results is in the contrast between what we may dub the deterministic case versus the random case. For instance while $D(\mathbb{Z}_n) = n$, a random sequence of size approximately $\log_2 n$ is already a Davenport Z -sequence with high probability. One may suspect that with a random \mathbb{Z}_n -sequence, the length of the sequence in order that it is a weighted Davenport Z -sequence with high probability is significantly smaller than the corresponding length in the deterministic case. This is however not always true as evidenced by the second part of Theorem 3 corresponding to the case $A = \{-1, 1\}$ where the requisite sequence size drops down *merely by a constant factor*. But such instances may be more the exception than the rule. At the moment, we do not quite have a definitive answer to this question, but the overall motif seems unmistakable.

The rest of the paper is organized as follows. In the following three sections we prove Theorems 1, 2 and 3 respectively. The proof of the last part of Theorem 1 invokes a result of Szemerédi [12] that settled a problem posed by Erdős and Eggleston. While the paper [12] settles a different conjecture of Erdős and Heilbronn, the same proof can be suitably modified to also settle the Erdős-Eggleston conjecture, and Szemerédi notes this as such in [12]. We include a proof of this result in the appendix, for the sake of completeness since we were unable to find any published proof of the same. We make no claim to any originality towards *this* proof; only the presentation and perspective (which is implicit in Szemerédi's paper) is ours. We conclude the paper with some general remarks and a few open questions.

2 Proof of Theorem 1

We start with the following simple observation. For any finite abelian group G and $A, B \subset G$ satisfying $|A| + |B| > |G|$ we necessarily have $A + B = G$, where $A + B := \{a + b : a \in A, b \in B\}$. This follows easily since for any $x \in G$ we have $|A| + |x - B| > |G|$, so $A \cap (x - B) \neq \emptyset$, and that implies $x \in A + B$.

Proof. 1. In the proof of the first part, we shall deal with the case where n is even, or n is odd, separately. We start with the odd case. Let $\mathbf{x} = (a_1, a_1, \dots, a_k, a_k, a_{2k+1}, \dots, a_{2n+3})$ be a \mathbb{Z}_{2n+1} -sequence of length $2n + 3$. In words, the elements a_1, \dots, a_k appear twice (for some k) while the elements $a_{2k+1}, \dots, a_{2n+3}$ appear once. In particular, the a_i 's are pairwise distinct.

Set $\ell = 2(n - k) + 3$, and consider the sets $A = \{a_1, a_2, \dots, a_k, a_{2k+1}, \dots, a_{2k+(\ell-1)/2}\}$ and $B = \{a_1, a_2, \dots, a_k, a_{2k+(\ell+1)/2}, \dots, a_{2k+\ell}\}$. Clearly, $|A| = n + 1$ and $|B| = n + 2$. Let A' denote the set of all possible sums of n different elements of A and similarly, let B' denote the

set of all possible sums of $n + 1$ distinct elements of B . Clearly, $|A'| = n + 1$ and $|B'| = n + 2$, so $|A'| + |B'| = 2n + 3 > 2n + 1$, therefore by the observation at the beginning of this section, it follows that $A' + B' = \mathbb{Z}_{2n+1}$. Hence $0 \in A' + B'$, or equivalently, \mathbf{x} admits a non-trivial zero-sum subsequence of length $2n + 1$. This establishes $\mathfrak{s}^{(2)}(\mathbb{Z}_{2n+1}) \leq 2n + 3$.

To show that $\mathfrak{s}^{(2)}(\mathbb{Z}_{2n+1}) \geq 2n + 3$, consider the sequence

$$\mathbf{x} = (1, 2, \dots, n - 1, n, n, n + 1, n + 1, n + 2, n + 3, \dots, 2n).$$

It is easy to see that \mathbf{x} is a sequence of length $2n + 2$ satisfying $\sum_i x_i = 0$ but no element is equal to 0. Moreover, no element appears in \mathbf{x} more than twice, so this establishes that $\mathfrak{s}^{(2)}(\mathbb{Z}_{2n+1}) \geq 2n + 3$. Consequently, $\mathfrak{s}^{(2)}(\mathbb{Z}_{2n+1}) = 2n + 3$.

For the even case, again as before, let $\mathbf{x} = (a_1, a_1, \dots, a_k, a_k, a_{2k+1}, a_{2k+2}, \dots, a_{2n+2})$ be a \mathbb{Z}_{2n} -sequence of length $2n + 2$ where each a_i appears twice for $1 \leq i \leq k$ and the elements $a_{2k+1}, \dots, a_{2n+2}$ appear exactly once, and so again, the a_i 's are pairwise distinct. Let $\ell = 2(n - k) + 2$, and define subsets

$$A = \{a_1, a_2, \dots, a_k, a_{2k+1}, \dots, a_{2k+\ell/2}\}, B = \{a_1, a_2, \dots, a_k, a_{2k+(\ell/2)+1}, \dots, a_{2k+\ell}\}$$

of \mathbb{Z}_{2n} and let A' and B' be the sets of sums of n distinct elements of A and B respectively. Since $|A| = n + 1$ and $|B| = n + 1$, we have $|A'| = n + 1$ and $|B'| = n + 1$, so that $|A'| + |B'| = 2n + 2 > 2n$. By the observation made earlier and the same argument as in the previous case, it follows that $\mathfrak{s}^{(2)}(\mathbb{Z}_{2n}) \leq 2n + 2$ for $n \geq 3$.

To complete the proof of the theorem, consider the sequence $\mathbf{x} = (1, 2, 3, 1, 3, 5, 6, \dots, n - 1, n, n, n + 1, n + 2, \dots, 2n - 1)$ for $n \geq 5$ of length $2n + 1$. It is easy to check that again the sum of all the elements of \mathbf{x} equals 0 but since 0 itself is not in \mathbf{x} , it follows that $\mathfrak{s}^{(2)}(\mathbb{Z}_{2n}) \geq 2n + 2$ for $n \geq 5$. For $n = 4$ and $n = 3$ consider the sequence $\mathbf{x} = (1, 2, 3, 1, 3, 4, 5, 6, 7)$ and $\mathbf{x} = (1, 2, 1, 2, 3, 4, 5)$ respectively, and the same argument works with these sequences. This completes the proof.

2. We now turn to prove the second part of the theorem. We start with the proof of the upper bound which we shall prove by induction on k .

The case of $k = 2$ is just a special case of the first part of Theorem 1 that was proved above. Suppose now that $k \geq 3$, and suppose that the statement holds for values less than k .

Let \mathbf{a} be a sequence of size $p + k$, where each element appears at most k times. If no element in \mathbf{a} has multiplicity k , then since $p + k > p + k - 1$, we are through by induction, so we may assume that there is at least one element in \mathbf{a} that appears k times. For each $1 \leq i \leq k$, let ℓ_i denote the number of distinct elements of \mathbf{a} which appear precisely $k - i + 1$ times, and set $L_i := \sum_{j=1}^i \ell_j$. Write

$$\mathbf{a} = \left(\underbrace{a_1, \dots, a_1}_{k \text{ times}}, \dots, \underbrace{a_{L_1}, \dots, a_{L_1}}_{k \text{ times}}, \underbrace{a_{L_1+1}, \dots, a_{L_1+1}}_{k-1 \text{ times}}, \dots, \underbrace{a_{L_2}, \dots, a_{L_2}}_{k-1 \text{ times}}, \dots, \underbrace{a_{L_{k-1}+1}, \dots, a_{L_k}}_{1 \text{ time}} \right)$$

where the a_i are distinct for $1 \leq i \leq L_k$ and $k\ell_1 + (k - 1)\ell_2 + \dots + \ell_k = p + k$.

Consider the sets A_i consisting of those a_j which appear at least i times in \mathbf{a} . More precisely, let $A_i = \{a_1, a_2, \dots, a_{\ell_1 + \dots + \ell_{k-i+1}}\}$, for $1 \leq i \leq k$. Note that $|A_i| = \ell_1 + \dots + \ell_{k-i+1}$. As in the proof of the previous part, A'_i be the sum of all $(\ell_1 + \ell_2 + \dots + \ell_{k-i+1} - 1)$ -sum of elements

of A_i for each i , so that $|A'_i| = |A_i|$. By the Cauchy-Davenport Theorem (see [10], theorem 2.3, page 44, for instance) we have

$$\left| \sum_{i=1}^k A'_i \right| \geq \min(p, \sum_{i=1}^k |A'_i| - k + 1) = \min(p, p + k - k + 1) = p.$$

In particular we have $\sum_{i=1}^k A'_i = \mathbb{Z}_p$ and hence $0 \in \sum_{i=1}^k A'_i$, where A'_i is the sum of all $(\ell_1 + \ell_2 + \dots + \ell_{k-i+1} - 1)$ elements of the sequence. In particular, there is a subsequence of \mathbf{a} of length $\sum_{i=1}^k (\ell_1 + \ell_2 + \dots + \ell_{k-i+1} - 1) = p + k - k = p$ whose sum is zero, and that completes the induction.

To get the lower bound for $\mathfrak{s}^{(k)}(\mathbb{Z}_p)$, write $p = (k-1)\ell + r$ for some ℓ , and $0 < r < k-1$, so that $\ell = \frac{p-r}{k-1}$. Denote $p-1$ by -1 . We shall construct a sequence \mathbf{x} of length $p+k-1$ of the form

$$\mathbf{x} = \left(\underbrace{-1, \dots, -1}_{k \text{ times}}, \underbrace{0, \dots, 0}_{k-1 \text{ times}}, \dots, \underbrace{\ell-2, \dots, \ell-2}_{k-1 \text{ times}}, x_1, \dots, x_{r+k-2} \right)$$

where x_1, \dots, x_{r+k-2} are pairwise distinct such that \mathbf{x} has no zero-sum subsequence of length p .

Towards that end, we shall show that if p is sufficiently large (we will make this more precise soon) there exists a subset $\{x_1, \dots, x_{r+k-2}\} \subset \{0, 1, \dots, \ell-2\}$ in \mathbb{Z}_p such that

$$\sum_{i=1}^{r+k-2} x_i = -(k-1)(1+2+\dots+(\ell-2)) = -\frac{(p-r-(k-1))(p-r-2(k-1))}{2(k-1)} \quad (1)$$

in \mathbb{Z}_p .

First, to see why this suffices for our needs, observe that the sequence \mathbf{x} has no element appearing more than k times. If (1) holds then the sum of all the elements of \mathbf{x} equals $-k$. Denote the subsequence of \mathbf{x} omitting the -1 's by \mathbf{a} , so that \mathbf{a} is a \mathbb{Z}_p -sequence of length $p-1$. Then note that the sum of any $s \leq k-1$ elements of \mathbf{a} (when viewed as integers) is at most $(k-1)(\ell-2) < p-k$. Consequently, for any $J \subset [1, p-1]$ with $|J| = s \leq k-1$ we have $\mathbf{a}_J \neq -(s+1)$. The upshot of this observation is that if \mathbf{x} admits a zero-sum subsequence of length p , then its complementary subsequence, i.e., the subsequence consisting of the remaining elements in \mathbf{x} , is a subsequence of length $k-1$ which has sum $-k$. This complementary sequence consists of some subsequence \mathbf{b} of \mathbf{a} of length s (for some $1 \leq s \leq k-1$) along with some $k-1-s$ elements that are equal to -1 . But this implies that the sum of the elements of \mathbf{b} must equal $-(s+1)$ which is not possible by the observation made above. So, it suffices to find $x_1, \dots, x_{r+k-2} \subset \{0, \dots, \ell-2\}$ such that (1) holds.

Suppose ℓ is even, say $\ell = 2t$ so that $t = \frac{p-r}{2(k-1)}$. A straightforward calculation shows that (1) gives us

$$\begin{aligned} \sum_{i=1}^{r+k-2} x_i &= -\frac{(p-r-(k-1))(p-r-2(k-1))}{2(k-1)} \\ &= (r+k-1)(t-1) \end{aligned}$$

in \mathbb{Z}_p . Set $x_{r+k-2} = 2t - (k-1)$, $x_{r+k-3} = 2t - (k-2)$, $x_{r+k-4} = k-2-r$. If $r+k$ is odd then for $1 \leq i \leq \frac{r+k-5}{2}$ set

$$\begin{aligned} x_{2i-1} &= t - i, \\ x_{2i} &= t + i \end{aligned}$$

If $p \geq (k-1)(2k-1)$, then $t + \frac{r+k-5}{2} \leq \ell - 2$ and $0 \leq k-2-r < t - \frac{r+k-5}{2}$, so all these choices for x_i are pairwise distinct and lie in the set $\{0, \dots, \ell-2\}$. It is a straightforward

check to see that $\sum_{i=1}^{r+k-2} x_i = (r+k-1)(t-1)$.

If $r+k$ is even, then set $x_{r+k-5} = t$ for $1 \leq i \leq \frac{r+k-6}{2}$ set

$$\begin{aligned} x_{2i-1} &= t - i, \\ x_{2i} &= t + i. \end{aligned}$$

Again, it is straightforward to check that the x_i satisfy the requirements, and this settles this case.

If ℓ is odd, write $\ell = 2t + 1$ so that $t = \frac{p-(r+k-1)}{2(k-1)}$. In this case, (1) simplifies as

$$\begin{aligned} \sum_{i=1}^{r+k-2} x_i &= -\frac{(p-r-(k-1))(p-r-2(k-1))}{2(k-1)} \\ &= t(r+2(k-1)) \\ &= tr + p - (r+k-1) \\ &= (t-1)r - k + 1. \end{aligned}$$

in \mathbb{Z}_p . Again, to satisfy these requirements, set $a = \left\lfloor \frac{(t-2)r}{r+k-3} \right\rfloor$, and set $x_i = a - 1 - i$ for $1 \leq i \leq r+k-3$. Observe that if $p = \Omega(k^4)$ then $r+k \leq a \leq t-2 < \ell$, and

$$\sum_{i=1}^{r+k-3} x_i = (t-2)a - \frac{(r+k-3)(r+k-2)}{2} - s(r, k)$$

for some $0 \leq s(r, k) < r+k-3 \leq 2k-5$. Now set

$$x_{r+k-2} = \frac{(r+k-3)(r+k-2)}{2} + 2(r-1) - s(r, k)$$

so that $\sum_i x_i = (t-1)r - k + 1$. It is easy to see that for all integers k , $\frac{(k-2)(k-1)}{2} \geq 2k-5$, so

$$\frac{(r+k-3)(r+k-2)}{2} + 2(r-1) - s(r, k) > \frac{(k-2)(k-1)}{2} - (2k-5) \geq 0$$

which implies that $x_{r+k-2} > 0$. Also,

$$x_{r+k-2} + (r+k-2) \leq 2(2k-5) + (2k-5)(k-2) \leq \frac{t-2}{k-3} \leq a$$

if $p \geq \Omega(k^4)$. These inequalities establish that $x_i \in \{0, \dots, \ell-2\}$ for all $1 \leq r+k-2$, are pairwise distinct, and (1) holds; this completes the proof.

Remark: Our construction of the sequence \mathbf{x} seems somewhat ad hoc, and unlike some of the results to the Inverse Sum Problems (see [7]), these extremal sequences do not appear to be unique in any sense. Also, as we make no attempt to optimize for $p_0(k)$ (in the statement of part 2 of Theorem 1), it should be possible to find better examples than ours; as noted earlier, we may take $p_0(k) = \Omega(k^4)$ but there may be much sharper bounds for $p_0(k)$. Another side to this argument is that for any $k \leq O(p^{1/4})$ we have $\mathbf{s}^{(k)}(\mathbb{Z}_p) = p + k$.

3. Before we prove the third part of Theorem 1 we state a theorem due to Szemerédi, which settled a conjecture of Erdős and Eggleston. In order to state this, we need a definition. For a finite abelian group G and $A \subset G$, by $\mathcal{S}(A)$ we mean the set of all those elements that occur as a sum of elements of some non-trivial subset of A .

Theorem 4. *There exists an absolute constant $1/2 > \varepsilon_0 > 0$ such that² the following holds. If G is a finite abelian group and $A \subset G$ then either $0 \in \mathcal{S}(A)$ or $|\mathcal{S}(A)| \geq \varepsilon_0 |A|^2$.*

A conjecture of Erdős-Heilbronn, which was settled by Szemerédi [12] states that there exists an absolute constant $C \geq 1$ such that for any abelian group G of order n , and any subset $A \subset G$ with $|A| \geq C\sqrt{n}$, there exists some non-trivial subset of A the sum of whose elements equals zero. As we mentioned in the introduction, Szemerédi remarks [12] in his paper that the same methods actually can be extended to prove the aforementioned result as well. As mentioned in the Introduction, the proof of Theorem 4 appears in Appendix I.

We will also need another result due to Scherk, settling a problem proposed by L. Moser. The original version was for the case $k = 2$ but a simple induction (on k) establishes the following version as well.

Theorem 5. ([11], [13]) *Suppose $B_1, \dots, B_k \subset \mathbb{Z}_n$ such that $0 \in \cap_{i=1}^k B_i$, and suppose the equation $0 = b_1 + \dots + b_k$ with $b_i \in B_i$ has the unique solution $b_i = 0$ for all i , then*

$$\left| \sum_{i=1}^k B_i \right| \geq \min \left\{ \sum_{i=1}^k |B_i| - k + 1, n \right\}.$$

We are now in a position to prove the third part of Theorem 1. We start with the lower bound.

Consider the sequence $\mathbf{x} = (\underbrace{1, \dots, 1}_{k \text{ times}}, \underbrace{2, \dots, 2}_{k \text{ times}}, \dots, \underbrace{t, \dots, t}_{k \text{ times}})$, where $\frac{kt(t+1)}{2} < \frac{n}{2}$. Set $N = kt$ so

that $N = \Omega(\sqrt{nk})$. Then note that for any non-trivial $I \subset [N]$, $0 < \mathbf{x}_I < n/2$ by choice, so \mathbf{x} does not admit a zero-sum subsequence.

Let $C \geq 1$ be the constant from Szemerédi's theorem settling the Erdős-Heilbronn conjecture, and let ε_0 be the constant from Theorem 4. Let $C^* = 2C/\varepsilon_0$. We claim that if \mathbf{x} is a k -restricted \mathbb{Z}_n -sequence of length $m \geq \lceil C^* \sqrt{nk} \rceil + 1$, then \mathbf{x} is a Davenport Z -sequence, and we shall prove this by induction on k . The case $k = 1$ simply follows from Szemerédi's theorem, so suppose $k > 1$ and suppose the result holds for smaller values than k .

Let x_0 be a non-zero element of the sequence \mathbf{x} and let \mathbf{y} be the sequence obtained from \mathbf{x} by removing x_0 . Then \mathbf{y} is a sequence of length at least $\lceil C^* \sqrt{nk} \rceil \geq C^* \sqrt{nk}$. Write

$$\mathbf{y} = (\underbrace{x_1, \dots, x_1}_{\ell_1 \text{ times}}, \dots, \underbrace{x_r, \dots, x_r}_{\ell_r \text{ times}})$$

²The statement of Szemerédi's theorem in fact only states the existence of a constant $0 < \varepsilon_0 < 1$ but putting a smaller constant does not change the statement in any way, so we shall put a bound $\varepsilon_0 \leq 1/2$ here for convenience.

with $k \geq \ell_1 \geq \dots \geq \ell_r$ and where x_1, \dots, x_r are pairwise distinct non-zero elements of \mathbb{Z}_n . If $k > \ell_1$, then we are through by induction since $\lceil C^* \sqrt{nk} \rceil \geq \lceil C^* \sqrt{n(k-1)} \rceil + 1$ and no element of \mathbf{y} appears more than $k - 1$ times. So, we may assume that $\ell_1 = k$.

For $1 \leq i \leq k$ let R_i denote the set of those x_j that appear at least i times. Clearly $|R_1| \geq \dots \geq |R_k| \geq 1$. Let $r_i = |R_i|$. If $r_1 \geq C\sqrt{n}$ then again, by Szemerédi's Theorem $0 \in \mathcal{S}(R_1)$, and we are through, so we may assume that $r_1 < C\sqrt{n}$.

If $r_k \leq \frac{C^* \sqrt{n}}{\sqrt{k} + \sqrt{k-1}}$ then the subsequence of \mathbf{y} obtained by removing the elements of R_k is a $(k-1)$ -restricted subsequence of length at least $C^* \sqrt{nk} - r_k \geq C^* \sqrt{n(k-1)}$, so again, by induction, \mathbf{y} is a Davenport Z-sequence. So, again, we may assume that

$$r_k > \frac{C^* \sqrt{n}}{\sqrt{k} + \sqrt{k-1}} \geq (C/\varepsilon_0) \sqrt{\frac{n}{k}}.$$

Let $B_i = \mathcal{S}(R_i) \cup \{0\}$ for $i = 1, \dots, k$. By Theorem 4 either $0 \in \mathcal{S}(R_i)$ in which case we are through, or $|B_i| \geq \varepsilon_0 |R_i|^2$ for each i . If 0 can be written as $b_1 + \dots + b_k$ with $b_i \in B_i$ with at least one of the $b_i \neq 0$, then again, we are through, so we may assume that $0 = b_1 + \dots + b_k$ with $b_i \in B_i$ implies that $b_i = 0$ for each i . Then by the result of Scherk (Theorem 5),

$$\left| \sum_{i=1}^k B_i \right| \geq \min \left(n, \sum_{i=1}^k \varepsilon_0 |R_i|^2 - k + 1 \right) = n$$

as $|R_i| \geq (C/\varepsilon_0) \sqrt{n/k}$ for all i and $\frac{C^2}{\varepsilon_0} > 2$. Therefore, $\sum_{i=1}^k B_i = \mathbb{Z}_n$, which implies that $-x_0 = \sum_{i=1}^k b_i$ for some choices of non-zero $b_i \in B_i$. In particular, \mathbf{x} is a Davenport Z-sequence as required. This completes the induction and the proof. □

3 Proof of Theorem 2

Before we get into the proofs, we set up a bit of notation here which will be used repeatedly. By $\mathbb{I}(x)$ we mean the indicator function which equals 1 if $x = 0$ and is zero otherwise. For a subset $I \subset [m]$, we shall denote by \mathfrak{X}_I , the sum $\mathfrak{X}_I := \sum_{i \in I} X_i$. We shall sometimes write $\mathfrak{X}(I)$ instead of \mathfrak{X}_I for notational convenience.

Before we launch into the proofs, we state a simple lemma.

Lemma 6. (a) *Let X_1, \dots, X_m be uniformly distributed on \mathbb{Z}_n and suppose that for any $i \neq j$, X_i, X_j are independent. Then, for any $a_1, \dots, a_m \in \mathbb{Z}_n^*$ the random variable $\sum_{i=1}^m a_i X_i$ is uniformly distributed on \mathbb{Z}_n .*

(b) *Suppose $\mathfrak{X}_m = (X_1, \dots, X_m)$ where the X_i are independent and uniformly distributed over \mathbb{Z}_n . Let $I, J \subset [m]$ be non-empty subsets of indices, and let $Y_1 := \sum_{i \in I} a_i X_i$ and $Y_2 := \sum_{j \in J} b_j X_j$ for some $a_i, b_j \in \mathbb{Z}_n^*$. Let $i_0 \neq j_0$ be distinct indices in $I \cup J$, and let $x_i \in \mathbb{Z}_n$ for all $i \in I \cup J \setminus \{i_0, j_0\}$ be arbitrary. If the system of equations*

$$Y_1 = \alpha, Y_2 = \beta, X_i = x_i \text{ for all } i \neq i_0, j_0 \tag{2}$$

admits at most r solutions for the pair (X_{i_0}, X_{j_0}) for any α, β, x_i , then $\mathbb{P}(Y_1 = \alpha, Y_2 = \beta) \leq \frac{r}{n^2}$. In particular, if the system has a unique solution for the pair (X_{i_0}, X_{j_0}) , then Y_1 and Y_2 are independent.

Proof. (Of Lemma 6) Both the parts work by a conditioning argument. For the first part, pick an arbitrary \mathbb{Z}_n -sequence $\mathbf{x} = (x_2, \dots, x_m)$ and condition on $X_i = x_i$ for $i \geq 2$. Then

$$\mathbb{P}\left(\sum_i a_i X_i = \alpha \mid X_i = x_i \text{ for } i \geq 2\right) = \mathbb{P}\left(X_1 = \frac{1}{a_1} \left(\alpha - \sum_{i>1} a_i x_i\right)\right) = \frac{1}{n}$$

and this holds for all choices of the sequence \mathbf{x} . Hence the same holds for the unconditional probability as well.

For the second part, let \mathbf{x} be an arbitrary \mathbb{Z}_n -sequence indexed by the elements of $T := I \cup J \setminus \{i_0, j_0\}$ and condition on $X_\ell = x_\ell$ for all $\ell \in T$. Let the number of solutions for (X_{i_0}, X_{j_0}) to the system of equations (2) be $R(\alpha, \beta, \mathbf{x})$. Then

$$\mathbb{P}(Y_1 = \alpha, Y_2 = \beta \mid X_\ell = x_\ell \text{ for all } \ell \in T) = \frac{R(\alpha, \beta, \mathbf{x})}{n^2} \leq \frac{r}{n^2}$$

by the hypothesis. Since this holds for all choices of $\alpha, \beta, \mathbf{x}$, the same inequality holds unconditionally as well, and that completes the proof. \square

Proof. 1. Let $\mathfrak{X}_{n+2} = (X_1, \dots, X_{n+2})$ be a random \mathbb{Z}_n -sequence. Let $\mathcal{H} := \{I \subset [n+2] : |I| = n\}$ and $N := \sum_{I \in \mathcal{H}} \mathbb{I}(\mathfrak{X}_I)$. For any fixed $I \in \mathcal{H}$, $\mathbb{P}(\mathfrak{X}_I = 0) = 1/n$ by Lemma 6(a), so

$$\mathbb{E}(N) = \sum_{I \in \mathcal{H}} \mathbb{P}(\mathfrak{X}_I = 0) = \frac{1}{n} \binom{n+2}{n} = \frac{(n+2)(n+1)}{2n} = \Omega(n),$$

and,

$$\text{Var}(N) = \sum_{I \in \mathcal{H}} \text{Var}(\mathbb{I}(\mathfrak{X}_I)) + \sum_{\substack{I \neq J \\ I, J \in \mathcal{H}}} \text{Cov}(\mathbb{I}(\mathfrak{X}_I), \mathbb{I}(\mathfrak{X}_J)).$$

Pick $i \in I \setminus J$ and $j \in J \setminus I$, and consider the equations $\mathfrak{X}_I = \alpha, \mathfrak{X}_J = \beta$, and $X_\ell = x_\ell$ for all $\ell \neq i, j$. It is straightforward to check that this system of equations admits a unique solution for (X_i, X_j) , so by Lemma 6(b) it follows that $\{\mathfrak{X}_I\}_{I \in \mathcal{H}}$ are pairwise independent. Consequently, $\text{Cov}(\mathbb{I}(\mathfrak{X}_I), \mathbb{I}(\mathfrak{X}_J)) = 0$ for $I \neq J \in \mathcal{H}$. Also, $\text{Var}(\mathbb{I}(\mathfrak{X}_I)) = \frac{1}{n}(1 - \frac{1}{n})$, so

$$\text{Var}(N) = \sum_{I \in \mathcal{H}} \text{Var}(\mathbb{I}(\mathfrak{X}_I)) = \frac{1}{n} \left(1 - \frac{1}{n}\right) \frac{(n+2)(n+1)}{2} = O(n).$$

Therefore, by Chebyshev's inequality we have,

$$\mathbb{P}(N = 0) \leq \mathbb{P}(|N - \mathbb{E}(N)| \geq \mathbb{E}(N)) \leq \frac{\text{Var}(N)}{(\mathbb{E}(N))^2} = \frac{\frac{1}{2}(1 - \frac{1}{n})}{\frac{1}{4}(1 + \frac{2}{n})(n+1)} = O\left(\frac{1}{n}\right).$$

Hence $\mathbb{P}(N > 0) \rightarrow 1$ as $n \rightarrow \infty$.

2. Let $\mathfrak{X}_n = (X_1, \dots, X_n)$ be a random \mathbb{Z}_n -sequence and $A = \{a, b\}$. Without loss of generality, assume $\{a, a+b, a-b\} \subset \mathbb{Z}_n^*$. Let $\mathcal{H} := \{I \subset [n] : \emptyset \subsetneq I \subsetneq [n]\}$ and let

$$N := \sum_{I \in \mathcal{H}} \mathbb{I}_{a\mathfrak{X}(I)+b\mathfrak{X}(\bar{I})}$$

where $\bar{I} := [n] \setminus I$.

It is easy to see that

$$\mathbb{E}(N) = \sum_{I \in \mathcal{H}} \mathbb{P}\left(a\mathfrak{X}_I + b\mathfrak{X}_{\bar{I}} = 0\right) = \frac{2^n - 2}{n}$$

as each of the probabilities in the summation above equals $1/n$ by Lemma 6(a). Again,

$$\text{Var}(N) = \sum_{I \in \mathcal{H}} \text{Var}(\mathbb{I}_{a\mathfrak{X}(I)+b\mathfrak{X}(\bar{I})}) + \sum_{\substack{I \neq J \\ I, J \in \mathcal{H}}} \text{Cov}(\mathbb{I}_{a\mathfrak{X}(I)+b\mathfrak{X}(\bar{I})}, \mathbb{I}_{a\mathfrak{X}(J)+b\mathfrak{X}(\bar{J})})$$

and it follows that

$$\text{Var}\left(\mathbb{I}_{a\mathfrak{X}(I)+b\mathfrak{X}(\bar{I})}\right) = \frac{1}{n} \left(1 - \frac{1}{n}\right).$$

We claim that

$$\text{Cov}(\mathbb{I}_{a\mathfrak{X}(I)+b\mathfrak{X}(\bar{I})}, \mathbb{I}_{a\mathfrak{X}(J)+b\mathfrak{X}(\bar{J})}) = 0 \text{ for } I \neq J. \quad (3)$$

Hence

$$\text{Var}(N) = \sum_{I \in \mathcal{H}} \text{Var}(\mathbb{I}_{a\mathfrak{X}(I)+b\mathfrak{X}(\bar{I})}) < \frac{2^n}{n}$$

so again by Chebyshev's inequality,

$$\mathbb{P}(N = 0) \leq \mathbb{P}(|N - \mathbb{E}(N)| \geq \mathbb{E}(N)) \leq O(n2^{-n})$$

which gives us what we seek.

So to complete the proof, it remains to establish (3) which amounts to showing that

$$\mathbb{P}\left[\left(\mathfrak{X}(I) = -(b/a)\mathfrak{X}(\bar{I})\right) \text{ and } \left(\mathfrak{X}(J) = -(b/a)\mathfrak{X}(\bar{J})\right)\right] = \frac{1}{n^2}$$

whenever $I \neq J$.

Towards this end, suppose $I \neq J$ be subsets of $[m]$. We shall appropriately choose $i \neq j$ in $I \cup J$ and consider the equations $\mathfrak{X}_I = \alpha$, $\mathfrak{X}_J = \beta$, and $X_\ell = x_\ell$ for all $\ell \neq i, j$. We shall use Lemma 6(b) and show that this system of equations has a unique solution for (X_i, X_j) . We have the following cases:

- i. $I \setminus J \neq \emptyset$ and $J \setminus I \neq \emptyset$: Pick $i \in I \setminus J$, $j \in J \setminus I$. The system of equations reduces to two equations of the form $aX_i + bX_j = \xi_1$, $bX_i + aX_j = \xi_2$ for some $\xi_1, \xi_2 \in \mathbb{Z}_n$ and since by assumption $a^2 - b^2 \in \mathbb{Z}_n^*$, these equations admit a unique solution for (X_i, X_j) . Hence we are through by Lemma 6(b).

ii. If the previous case does not arise, then either $I \subset J$ or $J \subset I$. Suppose without loss of generality that $I \subset J$. Pick $i \in I, j \in J \setminus I$. Again, the system of equations reduces to two linear equations of the form $aX_i + bX_j = \xi_1, aX_i + aX_j = \xi_2$ for some $\xi_1, \xi_2 \in \mathbb{Z}_n$. Again, since $a, a - b \in \mathbb{Z}_n^*$, it is easy to see that the system of equations admits a unique solution for (X_i, X_j) and so by Lemma 6(b) we are through.

3. Let $\mathfrak{X}_{n+1} = (X_1, \dots, X_{n+1})$ be a random \mathbb{Z}_n -sequence, and define

$$\begin{aligned} \mathcal{H} &:= \{(\{I, J\} : I, J \subset [n+1], I, J \neq \emptyset, I \cap J = \emptyset, |I \cup J| = n)\}, \\ N &:= \sum_{\{I, J\} \in \mathcal{H}} \mathbb{I}(\mathfrak{X}_{I, J}), \end{aligned}$$

where $\mathbb{I}(\mathfrak{X}_{I, J}) := \mathbb{I}(\mathfrak{X}_I - \mathfrak{X}_J)$. Note that $|\mathcal{H}| = (n+1)(2^{n-1} - 1)$.

Again,

$$\mathbb{E}(N) = \sum_{\{I, J\} \in \mathcal{H}} \mathbb{P}(\mathfrak{X}_I = \mathfrak{X}_J) = \frac{(n+1)(2^{n-1} - 1)}{n}$$

for the same reason as in the preceding discussions. Again we shall bound $\text{Var}(N)$ but unlike the proof of parts 1 and 2 of Theorem 2, we do not always have pairwise independence of the random variables $\mathfrak{X}_{I, J}$ in this case. But it turns out that most pairs $\{I, J\} \neq \{I', J'\}$ are pairwise independent and that is sufficient for our purpose here.

Suppose $\{I, J\} \neq \{I', J'\}$ are pairs in \mathcal{H} such that $I \cup J = [n+1] \setminus \{a\}$ and $I' \cup J' = [n+1] \setminus \{b\}$ with $a \neq b$. We claim that the random variables $\mathfrak{X}_{I, J}$ and $\mathfrak{X}_{I', J'}$ are independent. Without loss of generality, suppose $a \in I'$ and $b \in I$. Let $T = [n+1] \setminus \{a, b\}$ and let \mathbf{x} be an arbitrary \mathbb{Z}_n -sequence indexed by the elements of T . The system of equations $\mathfrak{X}_I - \mathfrak{X}_J = \alpha, \mathfrak{X}_{I'} - \mathfrak{X}_{J'} = \beta, X_\ell = x_\ell$ for all $\ell \in T$ reduces to $X_a = \mathbf{x}_{J'} - \mathbf{x}_{I' \setminus \{a\}}$ and $X_b = \mathbf{x}_J - \mathbf{x}_{I \setminus \{b\}}$, thereby giving a unique solution for (X_a, X_b) ; consequently, by Lemma 6(b) $\mathfrak{X}_{I, J}, \mathfrak{X}_{I', J'}$ are independent.

The aforementioned argument in fact shows that if n is odd, then $\mathfrak{X}_{I, J}$ are *all* pairwise independent. Indeed, suppose $I \cup J = I' \cup J' = [n]$ (without loss of generality). Since at least one of I, J meets both I' and J' nontrivially, suppose I intersects both I', J' and pick $i \in I \cap I'$ and $j \in I \cap J'$. Then for $T = [n+1] \setminus \{i, j\}$ and any \mathbb{Z}_n -sequence \mathbf{x} indexed by the elements of T , the same system of equations as before reduces to $X_i + X_j = \xi_1, X_i - X_j = \xi_2$ for some $\xi_1, \xi_2 \in \mathbb{Z}_n$. But observe that if n is odd, this system of equations admits a unique solution for (X_i, X_j) and so again by Lemma 6(b), it follows that $\mathfrak{X}_{I, J}, \mathfrak{X}_{I', J'}$ are independent.

If n is even, then this might not admit any solution at all, in which case $\mathbb{I}(\mathfrak{X}_{I, J}), \mathbb{I}(\mathfrak{X}_{I', J'})$ are negatively correlated. If $\xi_1 + \xi_2$ is even, then the pair of linear equations above admit two possible solutions for X_i (say), and for each of these, a unique value for X_j . Consequently, in these cases, the system of equations considered above admits at most 2 solutions (X_i, X_j) for any choice of $\alpha, \beta, \mathbf{x}$. So again, by Lemma 6, it follows that

$$\text{Cov}(\mathbb{I}(\mathfrak{X}_{I, J}), \mathbb{I}(\mathfrak{X}_{I', J'})) = \mathbb{P}(\mathfrak{X}_{I, J} = 0 \text{ and } \mathfrak{X}_{I', J'} = 0) - \mathbb{P}(\mathfrak{X}_{I, J} = 0)\mathbb{P}(\mathfrak{X}_{I', J'} = 0) \leq \frac{2}{n^2} - \frac{1}{n^2} = \frac{1}{n^2}.$$

To complete the proof, we need a bound on the number of distinct pairs $\{I, J\}, \{I', J'\}$ of members of \mathcal{H} such that $I \cup J = I' \cup J'$. Since there are $n+1$ choices for the element not in

$I \cup J$, $2^n - 2$ choices for I and $2^n - 3$ further choices for I' , we have

$$\begin{aligned} \text{Var}(N) &= \sum_{\{I,J\} \in \mathcal{H}} \text{Var}(\mathbb{I}(\mathfrak{X}_{I,J})) + 2 \sum_{\substack{\{I,J\} \neq \{I',J'\} \\ \{I,J\}, \{I',J'\} \in \mathcal{H}}} \text{Cov}(\mathbb{I}(\mathfrak{X}_{I,J}), \mathbb{I}(\mathfrak{X}_{I',J'})) \\ &= \frac{1}{n} \left(1 - \frac{1}{n}\right) (n+1)(2^{n-1} - 1) + 2 \frac{(n+1)(2^n - 2)(2^n - 3)}{n^2}, \\ &\leq \frac{3 \cdot 2^{2n}}{n} \end{aligned}$$

so, by Chebyshev's inequality,

$$\mathbb{P}(N = 0) \leq \mathbb{P}(|N - \mathbb{E}(N)| \geq \mathbb{E}(N)) \leq \frac{\text{Var}(N)}{(\mathbb{E}(N))^2} = O\left(\frac{1}{n}\right),$$

so again, we have $\mathbb{P}(N > 0) \rightarrow 1$. This completes the proof of this part and that of the theorem as well. □

Remark: Since one needs a sequence of size at least n in order that it is an Erdős Z -sequence, the results of the previous theorem assert that for random \mathbb{Z}_n -sequences one does not need much more than the *absolute* minimum required size for it to be an Erdős Z -sequence *whp*. In particular, if n is prime, then for any A of size at least 3, a random \mathbb{Z}_n -sequence of length n is *whp* an A -weighted Erdős Z -sequence.

4 Proof of Theorem 3

Let $\mathfrak{X}_m = (X_1, \dots, X_m)$ be a random \mathbb{Z}_n -sequence. The proof of the first and second parts of this theorem again involve the application of Chebyshev's inequality. The third part needs an additional lemma which we shall state at the appropriate juncture.

Let $\omega(n)$ be an arbitrary function satisfying $\omega(n) \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. 1. We start with the lower bound. Fix $m \leq \log_2 n - \omega(n)$, and let $\mathcal{H}_m := \{I : I \subset [m], I \neq \emptyset\}$ and $N_m := \sum_{I \in \mathcal{H}_m} \mathbb{I}(\mathfrak{X}_I)$.

Then

$$\mathbb{E}(N_m) = \sum_{I \in \mathcal{H}_m} \mathbb{P}(\mathfrak{X}_I = 0) = \frac{2^m - 1}{n} \leq \frac{1}{2^{\omega(n)}}$$

and consequently it follows - by Markov's inequality - that $N_m = 0$ *whp*.

For the upper bound, let $m \geq \log_2 n + \omega(n)$; then

$$\text{Var}(N_m) = \sum_{I \in \mathcal{H}_m} \text{Var}(\mathbb{I}(\mathfrak{X}_I)) + \sum_{\substack{I \neq J \\ I, J \in \mathcal{H}_m}} \text{Cov}(\mathbb{I}(\mathfrak{X}_I), \mathbb{I}(\mathfrak{X}_J)).$$

By a very similar argument to that in the proof of part 1 of Theorem 2, it follows that \mathfrak{X}_I 's are pairwise independent, so that $\text{Cov}(\mathbb{I}(\mathfrak{X}_I), \mathbb{I}(\mathfrak{X}_J)) = 0$ for $I \neq J \in \mathcal{H}_m$. Hence,

$$\text{Var}(N_m) = \sum_{I \in \mathcal{H}_m} \text{Var}(\mathbb{I}(\mathfrak{X}_I)) = \frac{1}{n} \left(1 - \frac{1}{n}\right) (2^m - 1) \leq \frac{2^{(\log_2 n + \omega(n))}}{n},$$

so by Chebyshev's inequality

$$\mathbb{P}(N_m = 0) \leq \frac{\text{Var}(N_m)}{\mathbb{E}^2(N_m)} \leq \frac{1}{2^{\omega(n)}}.$$

This completes the proof of the first part.

2. Note that if $\mathbf{x} = (x_1, \dots, x_m)$ is a $\{1, -1\}$ -weighted Davenport Z-sequence, then there are disjoint subsets I, J with $I \cup J \neq \emptyset$ (with one of them possibly empty) such that $\mathbf{x}_I = \mathbf{x}_J$. In view of this, define

$$\begin{aligned} \mathcal{H}_m &:= \{\{I, J\} : I, J \subset [m], I \cup J \neq \emptyset, I \cap J = \emptyset\}, \\ N_m &:= \sum_{\{I, J\} \in \mathcal{H}_m} \mathbb{I}(\mathfrak{X}_{I, J}), \end{aligned}$$

where again $\mathbb{I}(\mathfrak{X}_{I, J}) := \mathbb{I}(\mathfrak{X}_I - \mathfrak{X}_J)$. Note that we allow for I (resp. J) to be the empty set in which case \mathfrak{X}_I (resp \mathfrak{X}_J) is zero.

First, observe that $|\mathcal{H}_m| = \frac{3^m - 1}{2}$. Indeed, the number of ordered pairs of sets $(I, I \cup J)$ where both are not empty is $3^n - 1$ and these ordered pairs count each member of \mathcal{H}_m twice. Hence if $m \leq \log_3 n - \omega(n)$,

$$\mathbb{E}(N_m) = \sum_{(I, J) \in \mathcal{H}_m} \mathbb{P}(\mathfrak{X}_I = \mathfrak{X}_J) = \frac{3^m - 1}{2n} < \frac{1}{3^{\omega(n)}}$$

which implies $\mathbb{P}(N_m > 0) \rightarrow 0$, and this establishes that *whp* \mathfrak{X}_m is not a $\{1, -1\}$ -weighted Davenport Z-sequence.

For the other part, suppose $m \geq \log_3 n + \omega(n)$. As before we bound

$$\text{Var}(N_m) = \sum_{\{I, J\} \in \mathcal{H}_m} \text{Var}(\mathbb{I}(\mathfrak{X}_{I, J})) + \sum_{\substack{\{I, J\}, \{I', J'\} \in \mathcal{H}_m \\ \{I, J\} \neq \{I', J'\}}} \text{Cov}(\mathbb{I}(\mathfrak{X}_{I, J}), \mathbb{I}(\mathfrak{X}_{I', J'}))$$

by using Lemma 6. The following claim is crucial:

Claim 4.1. *Cov*($\mathbb{I}(\mathfrak{X}_{I, J}), \mathbb{I}(\mathfrak{X}_{I', J'})$) = 0 unless, one of the following two cases arise:

- (a) $I = \emptyset, I', J' \neq \emptyset$ and $J = I' \cup J'$.
- (b) $I, J, I', J' \neq \emptyset, I \subset J' \subset I \cup J$ and $I' \subset J$.

In these cases,

$$\text{Cov}(\mathbb{I}(\mathfrak{X}_{I, J}), \mathbb{I}(\mathfrak{X}_{I', J'})) \leq \frac{1}{n^2}.$$

The proof of the claim while somewhat straightforward, is a study of several cases that arise, so we relegate it to Appendix II.

Call a pair $\{I, J\}, \{I', J'\} \in \mathcal{H}_m$ bad if one of the two cases above holds. Note that the number of bad pairs $\{I, J\}, \{I', J'\} \in \mathcal{H}_m$ of the first type is at most $4 \cdot 3^m$ and the number of bad pairs of the second type is at most $4 \cdot 5^m$. Consequently,

$$\mathbb{P}(N_m = 0) \leq \frac{\text{Var}(N_m)}{\mathbb{E}^2(N_m)} \leq \frac{1}{\mathbb{E}(N_m)} + O\left(\frac{5}{9}\right)^m = o(1)$$

and as before, that completes the proof.

3. We start with a lemma. For integers x_1, \dots, x_m we denote their gcd by $\gcd(x_1, \dots, x_m)$.

Lemma 7. *Suppose $n = \prod_{i=1}^r p_i$, where $p_1 < \dots < p_r$ are odd primes and let $m \geq 2$ be a positive integer. For any x_1, \dots, x_m there exist non-zero integers $a_1, \dots, a_m \in \mathbb{Z}$ such that $\gcd(a_i, n) = 1$ for all $1 \leq i \leq m$, and satisfying $\sum_{i=1}^m a_i x_i = \gcd(x_1, \dots, x_m)$.*

Proof. (Proof of Lemma 7) We proceed by induction on m . We first start with the case that $\gcd(x_1, \dots, x_m) = 1$ and then move to the general case later. Suppose $m = 2$, and we have relatively prime integers x_1, x_2 . We shall establish the lemma in this case by induction on r . Suppose $r = 1$. In particular, there are integers $a_1, a_2 \in \mathbb{Z}$ such that $a_1 x_1 + a_2 x_2 = 1$. We claim that there exists $\lambda \in \mathbb{Z}$ such that both $a_1^* = a_1 + \lambda x_2$ and $a_2^* = a_2 - \lambda x_1$ are relatively prime to $n = p_1$. This would complete the proof in this case since $a_1^* x_1 + a_2^* x_2 = 1$.

Suppose that p_1 divides x_1 . Then $\gcd(p_1, a_2) = \gcd(p_1, x_2) = 1$. In particular, $a_2, x_2 \neq 0$ in \mathbb{Z}_{p_1} , so $a_1 + \lambda x_2 \neq 0$ if and only if $\lambda \neq -(a_1/x_2)$ in \mathbb{Z}_{p_1} . Hence there is an available choice for λ that satisfies the claim. The same argument works if p_1 divides x_2 . So now suppose that $\gcd(p_1, x_1) = \gcd(p_1, x_2) = 1$. Then as before, a choice for λ is unsuitable if either $\lambda = (a_2/x_1)$ or $\lambda = -(a_1/x_2)$. Since p_1 is odd, there is a choice for $\lambda \in \mathbb{Z}_{p_1}$ such that both these choices are avoided, and so the claim holds.

Let $n' = n/p_r$ and suppose (by induction on r) that there are a_1, a_2 such that $\gcd(a_i, n') = 1$ for $i = 1, 2$ such that $a_1 x_1 + a_2 x_2 = 1$. Consider $a_1^* = a_1 + \lambda n' x_2, a_2^* = a_2 - \lambda n' x_1$. Since $\gcd(n', p_r) = 1$, arguing in the same manner as before, it follows that there is a choice for λ such that a_1^*, a_2^* are both non-zero in \mathbb{Z}_{p_r} . Furthermore, $\gcd(a_1^*, n') = \gcd(a_2^*, n') = 1$, so the proof of the lemma is complete for the case $m = 2$ when $\gcd(x_1, x_2) = 1$. In the general case, if $\gcd(x_1, x_2) = d$ then consider $x'_1 = x_1/d, x'_2 = x_2/d$ so that $\gcd(x'_1, x'_2) = 1$ and use the preceding argument with x'_1, x'_2 instead to get a_1, a_2 as desired.

Suppose now that the lemma holds for smaller values than m . Write $d = \gcd(x_1, \dots, x_m)$ and let $\gcd(x_1, \dots, x_{m-1}) = d'$ so that $\gcd(d', x_m) = d$. By the case $m = 2$, there exist a, b such that $\gcd(a, n) = \gcd(b, n) = 1$ and $ad' + bx_m = d$. Again by induction, there exist a'_i for

$1 \leq i \leq m-1$ such that $\gcd(a'_i, n) = 1$ and $\sum_{i=1}^{m-1} a'_i x_i = d'$. Then it follows that $a_i = aa'_i$ for

$1 \leq i \leq m-1$ and $a_m = b$ satisfy the requirements of the statement of the lemma, and this completes the induction. \square

Let $n = \prod_{i=1}^r p_i$ be a product of distinct odd primes and suppose $m \geq \omega(n)$. Let $\mathfrak{X}_{2m} := (X_1, \dots, X_{2m})$ be a random \mathbb{Z}_n -sequence. We make the following claim:

Claim 4.2. *whp* $\gcd(X_1, \dots, X_m)$ and $\gcd(X_{m+1}, \dots, X_{2m})$ are relatively prime to n .

Proof. (Proof of Claim 4.2) For any element $x \in \mathbb{Z}_n$, let $\pi(x) := \{i : p_i \text{ divides } x\}$. Suppose X is picked uniformly from \mathbb{Z}_n . For a fixed set $A \subset [r]$, we claim that

$$\mathbb{P}(\pi(X) = A) = \prod_{i \in A} \frac{1}{p_i} \prod_{i \notin A} \left(1 - \frac{1}{p_i}\right).$$

In other words, each element $i \in [r]$ is picked independently into $\pi(X)$ with probability $1/p_i$. This follows easily from the Chinese Remainder Theorem: $\mathbb{Z}_n \simeq \mathbb{Z}_{p_1} \times \dots \times \mathbb{Z}_{p_r}$ as rings. If $\pi_i : \mathbb{Z}_n \rightarrow \mathbb{Z}_{p_i}$ is the projection map then p_i divides x if and only if $\pi_i(x) = 0$ and the claim follows immediately.

Consequently, for each i , the probability that p_i divides X_j for all $j \in [m]$ is $1/p_i^m$ so that

$$\begin{aligned} \mathbb{P}(\text{There exists } i \text{ such that } p_i \text{ divides } X_j \text{ for all } j \in [m]) &\leq \sum_{i=1}^r \frac{1}{p_i^m} \\ &\leq \frac{1}{p_1^{m-2}} \sum_{p \text{ prime}} \frac{1}{p^2} \\ &\leq \frac{1}{p_1^{m-2}} \left(\sum_{n \geq 1} \frac{1}{n^2} \right) = o(1) \end{aligned}$$

since $m \geq \omega(n)$. Hence it follows that *whp* $\gcd(X_1, \dots, X_m)$ is relatively prime to n . By the same reasoning, $\gcd(X_{m+1}, \dots, X_{2m})$ is also relatively prime to n *whp*. \square

We are now in a position to complete the proof of the third part of Theorem 3. By lemma 7 there exist $a_i, b_i \in \mathbb{Z}_n^*$ such that *whp* $a_1 X_1 + \dots + a_m X_m = u_1$ and $b_1 X_{m+1} + \dots + b_m X_{2m} = u_2$, for $u_1, u_2 \in \mathbb{Z}_n^*$. Hence

$$\sum_{i=1}^m (-u_2 a_i) X_i + \sum_{i=1}^m (u_1 b_i) X_{m+i} = 0$$

holds *whp* and this establishes that *whp* \mathfrak{X}_{2m} is a \mathbb{Z}_n^* -weighted Davenport Z -sequence and the proof is complete. \square

Remark: As mentioned in the introduction, and as is evident from the proof, the bound $m \geq \omega(n)$ in the last part of Theorem 3 is not always necessary. Indeed, for $\varepsilon > 0$, if $p_1 > \varepsilon^{-1}$ then the same proof shows that for $m = 3$ the probability that a random sequence \mathfrak{X}_3 of length 3 is a Davenport Z -sequence is at least $1 - \varepsilon$.

5 Concluding remarks

- As we remarked, our proof of Theorem 1, part 2, works only for p prime, since we invoke the Cauchy-Davenport Theorem there. A more general result due to Kneser (see [10], chapter 4)

provides a lower bound for all general n , but it is not clear how to adapt our proof effectively to the general case, or even to the case $n = p^m$ for $m \geq 2$, and p prime. But we believe the following conjecture holds:

Conjecture 8. *For any integer n and any integer k ,*

$$s^{(k)}(\mathbb{Z}_n) = n + k.$$

- Our proofs of Theorems 2 and 3 also give a bound on the error probability of their corresponding statements. A more interesting version of the same problem is the following quantitative avatar: Suppose $\varepsilon > 0$, and suppose $A \subset \mathbb{Z}_n \setminus \{0\}$, and let $\mathfrak{X}_m := (X_1, \dots, X_m)$ be a random \mathbb{Z}_n -sequence. Determine the maximum and minimum $m := m_A(n, \varepsilon), M := M_A(n, \varepsilon)$ respectively such that

$$\begin{aligned} \mathbb{P}(\mathfrak{X}_m \text{ is not a weighted Davenport } Z\text{-sequence for } A) &\geq 1 - \varepsilon \\ \mathbb{P}(\mathfrak{X}_M \text{ is a weighted Davenport } Z\text{-sequence for } A) &\geq 1 - \varepsilon \end{aligned}$$

hold. For instance, our proof of the first part of Theorem 3 establishes that for the set $A = \{1\}$ we have the bounds $m_A(n, \varepsilon) \leq \log_2(\varepsilon n)$ and $M_A(n, \varepsilon) \geq \log_2(\frac{n}{\varepsilon})$. It would be interesting to see how much of an improvement is possible on these results and that would probably need more sophisticated probabilistic techniques.

- The proof of the last part of Theorem 3 establishes that sufficiently long sequences are already Davenport Z -sequences *whp*. We also remarked after the proof of the theorem that the same is not required; if the smallest prime factor on n is sufficiently large, then bounded length sequences are already Davenport Z -sequences with high probability. But whether $m = \omega(n)$ is *necessary in some cases* is not quite apparent.
- The authors had previously proposed an extremal problem relating to the weighted Davenport constant of a group in [3] which goes as follows. Given a finite abelian group G , and an integer $k \geq 2$, we define $f_G^{(D)}(k)$ to be the least integer ℓ such that there is some subset $A \subset [1, \exp(G) - 1]$ of size ℓ such that $D_A(G) \leq k$. The most interesting case is for the group $G = \mathbb{Z}_p$ where p is a prime and in this case, it turns out that [4] in this case

$$p^{1/k} < f_G^{(D)}(k) < 4^{k^2} p^{1/k}.$$

As one of the principal motifs of this paper is that random \mathbb{Z}_n -sequences of much smaller length than $D_A(G)$ are sufficient to get weighted Davenport Z -sequences, it is natural to pose the following problem: Suppose $\varepsilon > 0$ and let p be a prime, and let $k \geq 2$ be a fixed positive integer. Suppose $\mathfrak{X}_k = (X_1, \dots, X_k)$ is a random \mathbb{Z}_p -sequence. Let

$$\mathcal{A}_\varepsilon := \{A \subset [1, p - 1] : \mathbb{P}(\mathfrak{X}_k \text{ is an } A\text{-weighted Davenport } Z\text{-sequence}) \geq 1 - \varepsilon\}.$$

Determine

$$f_{\text{Rand}}^{(D)}(p, k, \varepsilon) := \min_{A \in \mathcal{A}_\varepsilon} |A|.$$

By the result in [4] it follows that $f_{\text{Rand}}^{(D)}(p, k, \varepsilon) < 4^{k^2} p^{1/k}$. A preliminary guess based on the results here suggests that it is likely that $f_{\text{Rand}}^{(D)}(p, k, \varepsilon) = o(p^{1/k})$.

- Finally, two new problems suggest themselves rather naturally as a consequence of the result in the second part of Theorem 3. Let $\varepsilon > 0$ and recall the notions of $M_A(n, \varepsilon)$ and \mathcal{A}_ε as defined above.

(a) Determine

$$\alpha(n, \varepsilon) := \max \left\{ \frac{M_A(n, \varepsilon)}{D_A(\mathbb{Z}_n)} : A \in \mathcal{A}_\varepsilon \right\}.$$

The second part of Theorem 3 tells us that $\alpha(\varepsilon, n) \geq \log_3 2 - o(1)$. It would be interesting to see if this can be improved.

(b) Suppose $\varepsilon > 0$ and let $K \geq 1$ be a fixed parameter, and let \mathfrak{X}_m be a random \mathbb{Z}_n -sequence. Determine

$$s(n, \varepsilon, K) := \max \left\{ |A| : A \in \mathcal{A}_\varepsilon, M_A(n, \varepsilon) \geq \frac{\alpha}{K} D_A(\mathbb{Z}_n) \right\}.$$

Again, the second part of Theorem 3 tells us that $s(n, \varepsilon, K) \geq 2$ for all $K \geq 1$ and all $\varepsilon > 0$. Can it get significantly bigger? That is not clear at the moment.

Acknowledgments

The authors are grateful to the anonymous referees for their careful reading and for pointing out discrepancies and errors that were there in the original manuscript and also for pointing out to some inaccuracies in the proofs which led us to rewrite significant portions of the paper, which has improved the quality of the paper overall.

References

- [1] S. D. Adhikari, Y. G. Chen, J. B. Friedlander, S. V. Konyagin, and F. Pappalardi, Contributions to zero-sum problems. *Discrete Math.* **306** (2006), no. 1, 1-10.
- [2] S. D. Adhikari, and Y. G. Chen, Davenport constant with weights and some related question II, *J. Combin. Theory Ser. A* **115** (2008), No. 1, 178-184.
- [3] N. Balachandran, E. Mazumdar, The Weighted Davenport constant of a group and a related extremal problem, *Elect. J. Combin.*, **26** (2019), Issue 4, P4, 51.
- [4] N. Balachandran, E. Mazumdar, The Weighted Davenport constant of a group and a related extremal problem - II, <https://arxiv.org/abs/1912.07509>.
- [5] P. Erdős, A. Ginzburg, and A. Ziv, Theorem in the additive number theory, *Bull. Res. Counc. Israel Sect. F Math. Phys.* **10F** (1961), no. 1, 41-43.
- [6] S. Griffiths, The Erdős- Ginzburg - Ziv Theorem with units, *Discrete Math.* **308** (2008), no. 23, 5473 - 5484.
- [7] W. Gao, A. Geroldinger, and W. Schmid, Inverse zero-sum problems, *Acta Arith.* 245-279.
- [8] F. Luca, A generalization of a classical zero-sum problem, *Discrete Math.* **307** (2007), 1672-1678.

- [9] L. E. Marchan, O. Ordaz, D. Ramos and W. A. Schmid, Some Exact Values of the Harborth Constant and Its Plus-Minus Weighted Analogue, *Archiv der Mathematik* **101** (2013), 501-512 .
- [10] M. B. Nathanson, *Additive Number Theory: Inverse Problems and the Geometry of Sumsets*, Graduate texts in Mathematics, Vol. 165, Springer-Verlag, New York, 1996.
- [11] L. Moser and P. Scherk, Solution to advanced problem 4466, *Amer. Math. Monthly*, **1** (1955), Vol. 62, 46-47.
- [12] E. Szemerédi, On a conjecture of Erdős and Heilbronn, *Acta Arith.*, **17** (1970), 227-229.
- [13] H.B. Yu, A Simple Proof of a theorem of Bollobás and Leader. *Proc. American Math. Society.* **131** (2003), 9, 2639- 2640.
- [14] P. Yuan and X. Zeng, Davenport constant with weights. *European J. Comb.* **31** (2010), 677-680.

6 Appendix I: Szemerédi’s proof of the Erdős-Eggleston Conjecture

In this section, we present a proof of the Erdős-Eggleston conjecture that was settled by Szemerédi. In fact, Szemerédi’s proof works for any abelian group. Our presentation of the proof alone is ours, as these ideas are all there in Szemerédi’s paper. We make no claim regarding the optimality of the constant that appears here, nor do we make any attempts to optimize. We shall also drop ceilings and floors to make the presentation simpler. Recall that for a set $A \subset G$, by $\mathcal{S}(A)$ we mean the set of all sums $\sum_{x \in X} x$ as X varies over all non-empty subsets of A .

Theorem 9. (Szemerédi) *Let G be a finite abelian group and suppose $A \subset G$ such that $0 \notin \mathcal{S}(A)$. Then $|\mathcal{S}(A)| \geq \frac{|A|^2}{10000}$.*

Proof. Suppose the statement of the theorem does not hold; in particular suppose $A \subset G$ satisfies $0 \notin \mathcal{S}(A)$ and $|\mathcal{S}(A)| < \frac{|A|^2}{10000}$. We may assume without loss of generality that $|A|$ is sufficiently large (how large $|A|$ could be assumed to be will be determined shortly by the inequalities that shall appear). Let us write $|A| = \ell$. For each $\ell/4 \leq k \leq 3\ell/4$ we define the bipartite graph G_k with vertex sets $\binom{A}{k}, \binom{A}{k+1}$ (For a set A , $\binom{A}{r}$ denotes the set of all r -subsets of A) as follows. $X \in \binom{A}{k}$ is adjacent to $Y \in \binom{A}{k+1}$ in G_k if and only if $X \subset Y$ and $|\mathcal{S}(Y) \setminus \mathcal{S}(X)| \leq \ell/100$. The upshot of this definition for the graphs G_k is this: If we consider the union of all the graphs $\mathcal{G} = \cup_k G_k$, and consider any chain $X_0 \subset \dots \subset X_{\ell/2}$ of sets with $|X_i| = \ell/4 + i$ (for $0 \leq i \leq \ell/2$), then there are at most $\ell/100$ ‘missing edges’ along the chain in the graph \mathcal{G} . This follows since if there are more than $\ell/100$ missing edges along some chain $(X_1, \dots, X_{\ell/2})$ then $|\cup_i \mathcal{S}(X_{i+1}) \setminus \mathcal{S}(X_i)| > (\ell/100)(\ell/100)$ and that contradicts the assumption that the statement is false.

Fix integers t, k and consider $D \in \binom{A}{k}$. Denote by $\deg_k(D)$ the degree of D in G_k . Suppose $\deg_k(D), \deg_{k-1}(D) \geq t$. Let D have neighbors B_1, \dots, B_t and A_1, \dots, A_t in G_k and G_{k-1}

respectively. Write $A_i = D \setminus \{a_i\}$, and $B_i = D \cup \{b_i\}$. Since

$$S := \left\{ \sum_{x \in D} x - a_i + b_j : i, j \in [t] \right\} \supset \mathcal{S}(B_j)$$

and $|\mathcal{S}(B_j) \setminus \mathcal{S}(D)| \leq \ell/100$, there are at least $t - \ell/100$ elements in $S \cap \mathcal{S}(D)$. Since this holds for each j , a simple averaging argument implies that there exists i such that $t - \ell/100$ elements of the set $U := \{\sum_{x \in D} x - a_i + b_j : 1 \leq j \leq t\}$ belong to $\mathcal{S}(D)$, and since $|\mathcal{S}(D) \setminus \mathcal{S}(A_i)| \leq \ell/100$, it follows again that at least $|U \cap \mathcal{S}(A_i)| \geq t - \ell/50$. In particular, if $t > \ell/50$, there is some $1 \leq j_0 \leq t$ and $D_1 \subset D \setminus \{a_i\}$ such that

$$\sum_{x \in D} x - a_i + b_{j_0} = \sum_{y \in D_1} y, \quad \text{which implies that } b_{j_0} + \sum_{x \in D \setminus (D_1 \cup \{a_i\})} x = 0$$

contradicting the hypothesis that $0 \notin \mathcal{S}(A)$.

So, to complete the proof, we need to show that there exists $k \in [\ell/4 + 1, 3\ell/4]$ and $D \in \binom{A}{k}$ such that D has degree more than $\ell/50$ in both G_k and G_{k-1} . Towards that end, let us denote by $d(G_k)$ the *density* of G_k , i.e., $d(G_k) = \frac{e(G_k)}{m_k}$ where $e(G_k)$ denotes the number of edges in G_k and m_k denotes the cardinality of the set $\{(X, Y) : X \in \binom{A}{k}, Y \in \binom{A}{k+1}, X \subset Y\}$.

We now claim the following: If there exists k in this range such that both $d(G_k), d(G_{k-1}) \geq 2/3$, then there is a vertex $D \in \binom{A}{k}$ such that $\deg_k(D), \deg_{k-1}(D) > \ell/50$. By the preceding observation, the proof of the theorem follows as a consequence. To prove the claim, suppose $\text{BAD}_k := \{D \in \binom{A}{k} : \deg_k(D) \leq \ell/48\}$. Then

$$\frac{2}{3} \binom{\ell}{k} (\ell - k) = \frac{2m_k}{3} \leq e(G_k) \leq |\text{BAD}_k|(\ell/48) + \left(\binom{\ell}{k} - |\text{BAD}_k| \right) (\ell - k)$$

which gives $|\text{BAD}_k| \leq \frac{18}{44} \binom{\ell}{k} < \frac{1}{2} \binom{\ell}{k}$. Similarly, we get $|\text{BAD}_{k-1}| < \frac{1}{2} \binom{\ell}{k}$, so there exists $D \in \binom{A}{k} \setminus (\text{BAD}_k \cup \text{BAD}_{k-1})$ and for this D , we have $\deg_k(D) \geq \ell/48, \deg_{k-1}(D) \geq \ell/48$, and that achieves our goal.

So finally, to establish that for some k we have $d(G_k), d(G_{k-1}) \geq 2/3$, we revert to our observation at the very beginning of the proof: Suppose $X_i \in \binom{A}{\ell/4+i}$ for $0 \leq i \leq \ell/2$ such that $X_i \subset X_{i+1}$ for each i , and let \mathcal{C} denote the chain of subsets $\mathcal{C} = (X_0, \dots, X_{\ell/2})$ with $X_i \subset X_{i+1}$ for each $0 \leq i \leq \ell/2 - 1$. Then among the pairs (X_i, X_{i+1}) at most $\ell/100$ are non-edges in \mathcal{G} . Equivalently, for every chain $\mathcal{C} = (X_0, \dots, X_{\ell/2})$, if we uniformly and randomly pick i with $0 \leq i \leq \ell/2 - 1$, then the probability that (X_i, X_{i+1}) is an edge of \mathcal{G} is at least 0.99.

Call k a *Bad level*, if $d(G_k) < 2/3$. If what we seek does not hold, then every alternate level is Bad. In particular, if $\mathcal{C} = (X_0, \dots, X_{\ell/2})$ is a chain chosen uniformly and randomly and $\text{Miss}(\mathcal{C})$ denotes the number of missing edges along the sets in \mathcal{C} , then

$$\ell/100 \geq \mathbb{E}(\text{Miss}(\mathcal{C})) \geq \sum_{k \text{ Bad}} \mathbb{P}((X_i, X_{i+1}) \text{ is not an edge}) \geq \frac{1}{2} \cdot \frac{\ell}{2}$$

(the first inequality follows by assumption on $\text{Miss}(\mathcal{C})$ for all chains \mathcal{C} , and the last inequality follows since there are at most $\ell/2$ levels, and at least half of those are Bad by assumption) and that is a contradiction. \square

7 Appendix II: Proof of Claim 4.1

Suppose $\{I, J\} \neq \{I', J'\}$ satisfy $I \cap J = I' \cap J' = \emptyset$. As in the previous case of Theorem 3 we compute $\mathbb{P}(\mathfrak{X}_I = \mathfrak{X}_J \text{ and } \mathfrak{X}_{I'} = \mathfrak{X}_{J'})$, and use Lemma 6. Towards that end, we shall consider the system of equations (2) for a specific choice for the special indices i_0, j_0 and show that except in the exceptional cases mentioned in Claim 4.1, the system of equations admit unique solutions for (X_{i_0}, X_{j_0}) which implies that $\mathbb{P}(\mathfrak{X}_I = \mathfrak{X}_J \text{ and } \mathfrak{X}_{I'} = \mathfrak{X}_{J'}) = 1/n^2$ which in turn implies that $\text{Cov}(\mathbb{I}(\mathfrak{X}_{I,J}), \mathbb{I}(\mathfrak{X}_{I',J'})) = 0$. Even in these exceptional cases too, we shall show that there are at most two solutions for the pair (X_{i_0}, X_{j_0}) (like in the case of the proof of Theorem 2, part 3) which implies that in the bad cases, $\text{Cov}(\mathbb{I}(\mathfrak{X}_{I,J}), \mathbb{I}(\mathfrak{X}_{I',J'})) \leq 2/n^2$ as desired.

Suppose first that $I = \emptyset$. If $I' = \emptyset$ too, then both $J, J' \neq \emptyset$. If both $J \setminus J'$ and $J' \setminus J$ are nonempty, then pick $i \in J \setminus J', j \in J' \setminus J$. Let $T = I \cup J \cup I' \cup J' \setminus \{i, j\}$ and let \mathbf{x} be an arbitrary \mathbb{Z}_n -sequence indexed by the elements of T . Consider the equations in (2); this reduces to two equations of the form $X_i = \xi, X_j = \xi'$, which admits a unique solution for (X_i, X_j) in \mathbb{Z}_n . If $J \subset J'$, then again, pick $i \in J$ and $j \in J' \setminus J$. This sets up equations of the form $X_i = \xi, X_i + X_j = \xi'$ which again admit a unique solution for (X_i, X_j) . In fact, the same argument also works if $I = I'$ as well.

Next suppose $I = \emptyset$ but $J, I', J' \neq \emptyset$. In this case, first suppose that $J \setminus (I' \cup J') \neq \emptyset$. In this case, pick $i \in J \setminus (I' \cup J')$ and $j \in J'$; this gives equations $X_i = \xi, X_j = \xi'$ if $j \in J' \setminus J$ and equations $X_j = \xi, X_i + X_j = \xi'$ if $j \in J \cap J'$. In either case, this admits a unique solution for the pair (X_i, X_j) . If $J \subset (I' \cup J')$, pick $i \in (I' \cup J') \setminus J$ and $j \in J$ to get equations of the form $X_j = \xi, X_i \pm X_j = \xi'$ (depending on where i and j lie) which again leads to a unique solution for (X_i, X_j) . This finally leads us to the case $J = I' \cup J'$. This is the first of the exceptional cases in Claim 4.1.

Henceforth we shall assume that all $I, J, I', J' \neq \emptyset$ and further that all the sets I, I', J, J' are pairwise distinct.

Pick $i \in I$ and $j \in I'$. We now consider the various possibilities for the membership of elements i, j in the sets J, J' respectively. We shall merely write down the nature of the equations it imposes upon X_i, X_j . To keep our notation succinct, we shall denote the profiles by 0-1 tuples (a, b, c, d) which shall denote the following: $a = \mathbb{I}_{i \in I'}, b = \mathbb{I}_{i \in J'}, c = \mathbb{I}_{j \in I}, d = \mathbb{I}_{j \in J}$. So for instance, $(a, b, c, d) = (0, 0, 0, 0)$ simply means $i \in \overline{I' \cup J'}$ and $j \in \overline{I \cup J}$, and so on. Observe that if (a, b) and (c, d) are interchanged, the result is merely the interchange of i and j , so the nature of the equations is the same, so we shall club those instances into the same case. We shall merely note down the equations this forces upon (X_i, X_j) (or (X_j, X_i)). Also note that since $I \cap J = I' \cap J'$ we cannot have $(a, b), (c, d) = (1, 1)$; indeed, if $(a, b) = (1, 1)$ then it implies $i \in I' \cap J'$.

1. $(a, b, c, d) = (0, 0, 0, 0)$: $X_i = \xi, X_j = \xi'$.
2. $(a, b, c, d) = (0, 0, 0, 1), (0, 1, 0, 0)$: $X_i - X_j = \xi, X_j = \xi'$.
3. $(a, b, c, d) = (0, 0, 1, 0), (1, 0, 0, 0)$: $X_i + X_j = \xi, X_j = \xi'$.
4. $(a, b, c, d) = (1, 0, 0, 1)$: If this is the profile for every choice of $i \in I, j \in I'$ then it forces $I \subset I' \subset J$. But since $I \cap J = \emptyset$, this forces $I = \emptyset$ which has already been considered before.

5. $(a, b, c, d) = (1, 0, 1, 0)$: In this case, this forces $I \subset I' \subset I$ which gives $I = I'$ and again, this case has already been dealt with earlier.
6. $(a, b, c, d) = (0, 1, 0, 1)$. In this final profile, the aforementioned argument then gives us that $I \subset J'$ and $I' \subset J$. If $J' \setminus (I \cup J) \neq \emptyset$, pick $i \in I, j \in J' \setminus (I \cup J)$ to get equations of the form $X_i = \xi, X_i + X_j = \xi'$ which admit unique solutions. Hence we may assume that $I \subset J' \subset I \cup J$. Here, pick $i \in I, j \in J' \setminus I$ and condition on all the other X_ℓ . This gives equations $X_i - X_j = \xi, X_i + X_j = \xi'$ and in \mathbb{Z}_n , this system of equations admits a unique solution for (X_i, X_j) if n is odd, or admits at most 2 solutions if n is even. This completes the case-by-case analysis and the proof.