Zero-sum sets of prescribed size

Noga Alon and Moshe Dubiner
Department of Mathematics
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv, Israel

Abstract

Erdős, Ginzburg and Ziv proved that any sequence of $2n-1$ integers contains a subsequence of cardinality n the sum of whose elements is divisible by n. We present several proofs of this result, illustrating various combinatorial and algebraic tools that have numerous other applications in Combinatorial Number Theory. Our main new results deal with an analogous multi dimensional question. We show that any sequence of $6n-5$ elements of $Z_n \oplus Z_n$ contains an n-subset the sum of whose elements is the zero vector and consider briefly the higher dimensional case as well.

1 Introduction

The following theorem was proved in 1961 by Erdős, Ginzburg and Ziv.

Theorem 1.1 ([18]) *For any sequence $a_1, a_2, \ldots, a_{2n-1}$ of (not necessarily distinct) members of the cyclic group Z_n there is a set $I \subset \{1, 2, \ldots, 2n-1\}$ of cardinality $|I| = n$ so that $\sum_{i \in I} a_i = 0$ (in Z_n).*

This theorem has motivated the recent study of certain Ramsey type problems for graphs initiated by Bialostocki and Dierker in [8], [9] and studied by various researchers; see, e.g., [21], [32], [3]. The original proof of Theorem 1.1 is short and elementary. Somewhat surprisingly, this result can be proved in numerous distinct ways, which combine combinatorial and algebraic ideas. Here we present five proofs, which illustrate several powerful tools in Combinatorial Number Theory, some of which have other applications as well. We have been unable to modify any of these proofs and establish the following conjecture of Kemnitz [25], suggested, independently, by N. Zimmerman and Y. Peres, which deals with a two dimensional extension of the above theorem.

Conjecture 1.2 *For any sequence $a_1, a_2, \ldots, a_{4n-3}$ of (not necessarily distinct) members of the group $Z_n \oplus Z_n$ there is a set $I \subset \{1, 2, \ldots, 4n-3\}$ of cardinality $|I| = n$ so that $\sum_{i \in I} a_i = 0$ (in $Z_n \oplus Z_n$).*

The sequence consisting of $n-1$ copies of each of the four vectors $(0,0), (0,1), (1,0)$ and $(1,1)$ shows that the above conjecture, if true, is best possible. Similarly, the sequence consisting of $n-1$ copies of 0 and $n-1$ copies of 1 shows that Theorem 1.1 is tight. Our methods enable us to prove the following weaker version of the conjecture.

Theorem 1.3 *For any sequence $a_1, a_2, \ldots, a_{6n-5}$ of (not necessarily distinct) members of the group $Z_n \oplus Z_n$ there is a set $I \subset \{1, 2, \ldots, 6n-5\}$ of cardinality $|I| = n$ so that $\sum_{i \in I} a_i = 0$ (in $Z_n \oplus Z_n$).*

The above questions can be extended to higher dimensions. For two positive integers n and d, let $s(n,d)$ denote the smallest number s such that any sequence of s elements in $(Z_n)^d$ contains a 0-sum n-subsequence. By Theorem 1.1 $s(n,1) = 2n-1$, and if Conjecture 1.2 holds then $s(n,2) = 4n-3$. To the best of our knowledge, Harborth [24] was the first researcher who studied the function $s(n,d)$. His motivation was geometric; it is easy to see that $s(n,d)$ is the smallest integer s such that any set of s distinct lattice points in Z^d contains an n-subset whose centroid is also a lattice point. Several estimates for $s(n,d)$ appear in [24], [14], [25] and [26], and the best known general bounds for this function are given in the following inequality, whose simple proof appears in [24]:

$$(n-1)2^d + 1 \leq s(n,d) \leq (n-1)n^d + 1.$$

Here we are interested mainly in the case of small values of the dimension d and large n. In particular, we can show that

$$s(n,d) \leq c(d)n, \tag{1}$$

where $c(d)$ is a constant depending only on d.

The rest of this paper is organized as follows. In Section 2 we describe five proofs of Theorem 1.1. The proof of Theorem 1.3, as well as a proof of a stronger assertion for the case of large primes n are presented in Section 3. In this section we also include a sketch of a simpler proof of a slightly weaker version of the theorem. The final Section 4 contains some concluding remarks, together with a summary of the known estimates for the function $s(n,d)$ and a very brief outline of the methods used in our proof of the estimate (1). This proof combines combinatorial arguments with a Harmonic Analysis approach. Its detailed description will appear somewhere else.

2 Five proofs of the original theorem

As observed already in [18], it suffices to prove Theorem 1.1 for the case of prime n. To see this suppose the assertion of the theorem holds for prime n, and let us prove the general case by induction on the number of primes in the prime factorization of n. Put $n = pm$ where p is a prime, and let $a_1, \ldots, a_{2n-1}$ be the given sequence. By the result for the prime case, each subset of $2p-1$ members of the sequence contains a p-subset whose sum is 0 modulo p. Therefore, by repeatedly omitting from our sequence p-subsets of sum divisible by p, we can find $2m-1$ pairwise disjoint subsets $I_1, \ldots, I_{2m-1}$ of $\{1, \ldots, 2pm-1\}$, where $|I_i| = p$ for each i and the sum $\sum_{j \in I_i} a_j \equiv 0 (mod\ p)$ for each $1 \leq i \leq 2m-1$. (This is because as long as at most $2m-2$ such subsets have been chosen, the number of elements left is still at least $2pm - 1 - (2m-2)p = 2p-1$.) Define a sequence $a'_1, \ldots, a'_{2m-1}$, where $a'_i = \sum_{j \in I_i} a_j/p$. By the induction hypothesis this new sequence has a subset of m elements whose sum is divisible by m, and the union of the corresponding sets I_i supplies the desired n-subset whose sum is divisible by n. Therefore, the main step in the proof of Theorem 1.1 is the proof of the following proposition.

Proposition 2.1 *For a prime p and for any sequence $a_1, \ldots, a_{2p-1}$ of $2p-1$ elements in Z_p there is an $I \subset \{1, \ldots, 2p-1\}$ such that $|I| = p$ and $\sum_{i \in I} a_i = 0$ (in Z_p).*

In the rest of this section we present five proofs of this proposition.

2.1 The original proof

The original proof of Proposition 2.1 given in [18] is based on (a special case of) the Cauchy-Davenport Lemma. This Lemma, stated below, has many additional applications in Additive Number Theory.

Lemma 2.2 ([16]) *If p is a prime, and A, B are two nonempty subsets of Z_p, then $|A+B| \geq Min\{p, |A|+|B|-1\}$.*

This lemma can be proved quickly by induction on $|B|$. For $|B| = 1$ it is trivial. Assuming it holds for every A' and B' with $|B'| < |B|$, and given A and B, with $|A| < p$ and $|B| \geq 2$, suppose, first, that $A \cap B$ is a nonempty proper subset of B. In this case, one can apply the lemma to $A' = A \cup B$ and $B' = A \cap B$ and obtain the desired result, since $A' + B' \subset A + B$ and $|A'| + |B'| = |A| + |B|$. In case $A \cap B$ is not a nonempty, proper subset of B it is not too difficult to show that there is a $c \in Z_p$ so that $B \cap (A + c)$ is a nonempty proper subset of B and hence the result follows, as $|B + (A + c)| = |B + A|$. $\Box$

To prove Proposition 2.1, let us first renumber the elements $a_1, \ldots, a_{2p-1}$ so that $0 \leq a_1 \leq a_2 \leq \ldots \leq a_{2p-1}$. If $a_i = a_{i+p-1}$ for some $i \leq p-1$, then $a_i + a_{i+1} + \ldots + a_{i+p-1} = pa_i = 0$ (in Z_p) and the desired result follows. Otherwise, define $A_i = \{a_i, a_{i+p-1}\}$ for $1 \leq i \leq p-1$. By repeated application of the Cauchy-Davenport Lemma (Lemma 2.2), we conclude that $|A_1 + A_2 + \ldots + A_{p-1}| = p$, and hence every element of Z_p is a sum of precisely $p-1$ of the first $2p-2$ elements of our sequence. In particular, $-a_{2p-1}$ is such a sum, supplying the required p-subset whose sum is 0 in Z_p. □

2.2 The Chevalley Warning Theorem

The classical theorem of Chevalley and Warning, stated below, deals with the number of solutions of a system of polynomials with many variables over a finite field.

Theorem 2.3 (cf., [13], [33]) *For $j = 1, 2, \ldots, n$ let $P_j(x_1, \ldots, x_m)$ be a polynomial of degree r_j over a finite field F of characteristic p. If $\sum_{j=1}^{n} r_j < m$ then the number N of common zeros of $P_1, \ldots, P_n$ (in F^m) satisfies*

$$N \equiv 0 (mod\ p)\ .$$

In particular, if there is one common zero, then there is another one.

The proof is very short; clearly, if F has q elements then

$$N \equiv \sum_{x_1, \ldots, x_m \in F} \prod_{j=1}^{n} (1 - P_j(x_1, \ldots, x_m)^{q-1}) (mod\ p). \tag{2}$$

By expanding the right hand side we get a linear combination of monomials of the form $\prod_{i=1}^{m} x_i^{k_i}$ with $\sum_{i=1}^{m} k_i \leq (q-1) \sum_{j=1}^{n} r_j < (q-1)m$. Hence, in each such monomial there is an i with $k_i < q-1$. But then in $F = GF(q)$, $\sum_{x_i \in F} x_i^{k_i} = 0$, implying that the contribution of each monomial to the sum (2) is $0 (mod\ p)$ and completing the proof. □

As shown in [1], Proposition 2.1 (and hence Theorem 1.1) is an easy consequence of The Chevalley Warning Theorem. Given a sequence $a_1, \ldots, a_{2p-1}$ consider the following system of two polynomials in $2p-1$ variables x_i over the finite field Z_p:

$$\sum_{i=1}^{2p-1} a_i x_i^{p-1} = 0,$$

$$\sum_{i=1}^{2p-1} x_i^{p-1} = 0.$$

Since $2(p-1) < 2p-1$ and $x_1 = x_2 = \ldots = x_{2p-1} = 0$ is a solution, Theorem 2.3 implies the existence of a nontrivial solution $(y_1 \ldots y_{2p-1})$. Since by Fermat's little theorem, in Z_p $y^{p-1} = 1$ if

$y \neq 0$ and $0^{p-1} = 0$, the set $I = \{i : y_i \neq 0\}$ satisfies $\sum\{a_i : i \in I\} = 0$ and $|I| = p$, completing the proof. □

A similar, though slightly different derivation of Proposition 2.1 from the Chevalley Warning Theorem appears in [10].

2.3 A counting argument

Redmond and Ryavec [31] and, independently, Blokhuis [12], Gao [22] and Zimmerman ([34]) found the following short proof of Proposition 2.1. Let $a_1, \ldots, a_{2p-1}$ be the given sequence. Put $J = \{1, \ldots, 2p-1\}$ and consider the sum

$$S = \sum_{I \subset J, |I|=p} (\sum_{i \in I} a_i)^{p-1}$$

over the integers. It is obvious that S can be written as a sum of monomials of the form $c \prod_{i \in J} a_i^{k_i}$, where $\sum k_i = p-1$. In each such monomial, the number of positive numbers k_i is some j satisfying $1 \leq j \leq p-1$. Therefore, the number of distinct p-subsets I of J that contribute to the coefficient of this monomial in the sum S is $\binom{2p-1-j}{p-j}$ which is easily seen to be congruent to 0 modulo p. Since each such I contributes the same, this implies that the sum S is congruent to 0 modulo p. On the other hand, by Fermat's little theorem, if there is no subset $I \subset J$ of cardinality p so that $\sum_{i \in I} a_i \equiv 0 (mod\ p)$, then each of the $\binom{2p-1}{p}$ sets I contributes $1 (mod\ p)$ to the sum S, showing that in fact

$$S \equiv \binom{2p-1}{p} \equiv 1 (mod\ p).$$

This contradiction establishes the required assertion. □

2.4 Davenport's constant and group rings

For a finite abelian group G, define its *Davenport's constant* $s = s(G)$ to be the smallest positive integer s such that, for any sequence $g_1, g_2, \ldots, g_s$ of (not necessarily distinct) elements of G, there is an $\phi \neq I \subset \{1, \ldots, s\}$ such that $\sum\{g_i : i \in I\} = 0$. The problem of determining $s(G)$ was proposed by H. Davenport in 1966, and is related to the study of the maximal number of prime ideals in the decomposition of an irreducible integer in an algebraic number field whose class group is G. Olson [29] determined $s(G)$ for every p-group $G = Z_{p^{e_1}} \oplus \ldots \oplus Z_{p^{e_r}}$. Clearly

$$s(G) \geq 1 + \sum_{i=1}^{r} (p^{e_i} - 1) \ .$$

To see this let $x_1 \dots x_r$ be a basis for G, where x_i has order p^{e_i}, and consider the sequence of length $\sum_{i=1}^{r}(p^{e_i}-1)$ in which each x_i occurs $p^{e_i}-1$ times. No subsequence here has sum 0. Olson gave a beautiful proof of the following.

Theorem 2.4 $s(Z_{p^{e_1}} \oplus \ldots \oplus Z_{p^{e_r}}) = 1 + \sum_{i=1}^{r}(p^{e_i}-1)$.

The elegant proof, described below, is based on the fact that the ideal of nilpotent elements in the group-ring of a p-group over Z_p is nilpotent. Here are the details.

Proof. Let G be the finite abelian p-group with invariants $p^{e_1}, p^{e_2}, \ldots, p^{e_r}$, and let us use multiplicative notation for G. Let R be the group ring of G over Z_p. Suppose $k \geq 1 + \sum_{i=1}^{r}(p^{e_i}-1)$ and let $g_1, g_2, \ldots, g_k$ be a sequence of k members of G. We claim that in R

$$(1-g_1)\cdot(1-g_2)\cdot\ldots\cdot(1-g_k) = 0 \tag{3}$$

Indeed, let $x_1, x_2, \ldots, x_r$ be the standard basis for G, where the order of x_i is p^{e_i}. Since each g_j can be written as a product of the elements x_i, a repeated application of the identity $1-uv = (1-u)+u(1-v)$ enables us to express each expression of the form $1-g_j$ as a linear combination (with coefficients in R) of the elements $1-x_i$. Substituting into (3) and applying commutativity we conclude that the left hand side is a linear combination of elements of the form $\prod_{i=1}^{r}(1-x_i)^{k_i}$, where $\sum_{i=1}^{r} k_i = k > \sum_{i=1}^{r}(p^{e_i}-1)$. Hence, there is an i with $k_i \geq p^{e_i}$ and since in R, $(1-x_i)^{p^{e_i}} = 1 - x_i^{p^{e_i}} = 0$ this implies that (3) holds as claimed.

By interpreting (3) combinatorially we conclude that there is some nontrivial subsequence of $g_1, \ldots, g_k$ that has product 1, since otherwise, the coefficient of 1 in the above product will be nonzero. Hence $s(G) = 1 + \sum_{i=1}^{r}(p^{e_i}-1)$, as needed.□

There are many additional known results concerning the numbers $s(G)$. See [11], [17], [30], [27], [23] and their references. For our purposes here, a very special case of Theorem 2.4 suffices.

Proof of Proposition 2.1. By Theorem 2.4 any sequence of $2p-1$ elements of the group $Z_p \oplus Z_p$ has a nonempty subset whose sum is 0. The desired result follows by considering the sequence

$$(a_1, 1), (a_2, 1) \ldots, (a_{2p-1}, 1)$$

in this group. □

2.5 Permanents and vector sums

In this subsection we present another proof of Proposition 2.1, which will be useful in the the proof of Theorem 1.3 as well. The basic method here follows the argument in [7] and [5].

Lemma 2.5 *Let*

$$P = P(x_1, \ldots, x_m) = \sum_{U \subset \{1,\ldots,m\}} b_U \cdot \prod_{i \in U} x_i$$

be a multilinear polynomial over a commutative ring with identity. If $P(x_1, \ldots, x_m) = 0$ for each $(x_1, \ldots, x_m) \in \{0,1\}^m$ then $P \equiv 0$, that is, $b_U = 0$ for all $U \subset \{1, \ldots, m\}$.

This can be easily proved by induction on m, as shown in [4]. The following consequence of the above lemma is proved in [5] in a similar context.

Lemma 2.6 *Let $A = (a_{ij})$ be an m by m matrix over Z_p and suppose that the permanent $Per(A) \neq 0$ (in Z_p.) Then, for any $c_1, \ldots, c_m \in Z_p$ there are $\epsilon_1, \ldots, \epsilon_m \in \{0,1\}$ such that $\sum_{j=1}^m \epsilon_j a_{ij} \neq c_i$ for all $1 \leq i \leq m$.*

Proof. Suppose the lemma is false and there are no $\epsilon_1, \ldots, \epsilon_m$ as above. Consider the polynomial $P = P(x_1, \ldots, x_m) = \prod_{i=1}^m (\sum_{j=1}^m a_{ij} x_j - c_i)$. By assumption, $P(x_1, \ldots, x_m) = 0$ for all $(x_1, \ldots, x_m) \in \{0,1\}^m$. Let $\overline{P} = \overline{P}(x_1, \ldots, x_m)$ be the multilinear polynomial obtained from the standard representation of P as a sum of monomials by replacing each monomial of the form $a_U \prod_{i \in U} x_i^{\delta_i}$ with all $\delta_i > 0$, by $a_U \prod_{i \in U} x_i$. Clearly $\overline{P}(x_1, \ldots, x_m) = P(x_1, \ldots, x_m) = 0$ for all $(x_1, \ldots, x_m) \in \{0,1\}^m$. Therefore, by Lemma 2.5, $\overline{P} \equiv 0$. But this is impossible, since the coefficient of $\prod_{i=1}^m x_i$ in $\overline{P}$ (which equals the coefficient of this product in P) is $Per(A) \neq 0$. Therefore, the assertion of the lemma holds. □

Proof of Proposition 2.1. As in subsection 2.1, renumber the elements $a_1, \ldots, a_{2p-1}$ such that $0 \leq a_1 \leq \ldots \leq a_{2p-1}$. If there is an $i \leq p-1$ such that $a_i = a_{i+p-1}$ then $I = \{i, i+1, \ldots, i+p-1\}$ satisfies the assertion of the proposition. Otherwise, define $b_i = a_i - a_{i+p-1}$ $(\neq 0)$ for all $1 \leq i \leq p-1$, and let $c_1, \ldots, c_{p-1}$ be the set of all elements of Z_p besides the sum $-a_p - a_{p+1} - \ldots - a_{2p-1}$. Let $A = (a_{ij})$ be the $p-1$ by $p-1$ matrix defined by $a_{ij} = b_j$ for all $1 \leq i, j \leq p-1$. Clearly, $Per(A) = (p-1)! \cdot \prod_{j=1}^{p-1} b_j \neq 0$. Therefore, by Lemma 2.6, there are $\epsilon_1, \ldots, \epsilon_{p-1} \in \{0,1\}$ such that the sum $\sum_{j=1}^{p-1} \epsilon_j b_j$ differs from each c_i and is thus equal to $-a_p - a_{p+1} - \ldots - a_{2p-1}$. Hence, in Z_p,

$$a_{2p-1} + \sum_{i=1}^{p-1} [a_{i+p-1} + \epsilon_i (a_i - a_{i+p-1})] = 0,$$

and since each term $a_{i+p-1} + \epsilon_i(a_i - a_{i+p-1})$ is either a_{i+p-1} or a_i, this gives a p-subset of the sequence a_i the sum of whose elements is 0, as required. □

3 The two dimensional case

In this section we prove Theorem 1.3. We start by observing that just like in the one dimensional case it suffices to prove the theorem for prime n. Indeed, knowing the theorem holds for prime

n we prove the general case by induction on the number of primes in the prime factorization of n. Suppose $n = pm$, where p is a prime, and let $a_1, \ldots, a_{6pm-5}$ be a sequence of two dimensional integral vectors. By the result for the prime case any set of $6p-5$ members of the sequence contains a p-subset with sum divisible by p in each coordinate. By repeatedly deleting such subsets, and since $(6pm-5)-(6m-6)p = 6p-5$ we can find $6m-5$ pairwise disjoint subsets $I_1, \ldots, I_{6m-5}$ of $\{1, \ldots, 6n-5\}$, where $|I_i| = p$ for each i and where each of the two coordinates of each sum $\sum_{j \in I_i} a_j$ is divisible by p. We can now apply the induction hypothesis to the sequence $a'_i = \sum_{j \in I_i} a_j/p$ and complete the proof. Therefore, just like in the one dimensional case, the main step in the proof of Theorem 1.3 is the following.

Proposition 3.1 *For a prime p and for any sequence $a_1, \ldots a_{6p-5}$ of elements of $Z_p \oplus Z_p$ there is an $I \subset \{1, \ldots, 6p-5\}$ satisfying $|I| = p$ and $\sum_{i \in I} a_i = 0$ (in $Z_p \oplus Z_p$.)*

We start with the following simple consequence of the Chevalley Warning Theorem.

Lemma 3.2 *Let $a_1 = (a_{1,1}, a_{1,2}), a_2 = (a_{2,1}, a_{2,2}), \ldots, a_{3p} = (a_{3p,1}, a_{3p,2})$ be $3p$ elements of $Z_p \oplus Z_p$, where p is a prime, and suppose $\sum_{i=1}^{3p} a_i = (0,0)$. Then there is a p-subset $I \subset \{1, \ldots, 3p\}$ such that $\sum_{i \in I} a_i = (0,0)$.*

Proof. Consider the following system of 3 polynomial equations in the $3p-1$ variables x_i over the finite field Z_p:

$$\sum_{i=1}^{3p-1} a_{i,1} x_i^{p-1} = 0,$$

$$\sum_{i=1}^{3p-1} a_{i,2} x_i^{p-1} = 0,$$

$$\sum_{i=1}^{3p-1} x_i^{p-1} = 0.$$

Since $x_1 = x_2 = \ldots = x_{3p-1} = 0$ is a solution, there is another one, by Theorem 2.3, as the number of variables exceeds the sum of the degrees. Let $J \subset \{1, \ldots, 3p-1\}$ be the set of all indices of the nonzero entries of such a solution. Then $\sum_{i \in J} a_i = (0,0)$ in $Z_p \oplus Z_p$ and either $|J| = p$ or $|J| = 2p$. In the first case take $I = J$ and in the second define $I = \{1, \ldots, 3p\} \setminus J$. In both cases I satisfies the assertion of the lemma. □

Our basic approach here follows the one in [5]. For a two dimensional (column) vector $v = (d, e)$ in $Z_p \oplus Z_p$ let $v^* = v^*(p)$ denote the $2(p-1)$ dimensional (column) vector whose first $p-1$ coordinates are d and whose last $p-1$ coordinates are e. The next lemma is a simple corollary of Lemma 2.6.

Lemma 3.3 *Let $v_1, v_2, \ldots, v_{2p-2}$ be $2p-2$ vectors in $Z_p \oplus Z_p$ and let A be the $2p-2$ by $2p-2$ matrix whose columns are the vectors $v_1^*, v_2^*, \ldots, v_{2p-2}^*$. If in Z_p, $Per(A) \neq 0$, then for any vector $b = (d, e)$ in $Z_p \oplus Z_p$ there are $\epsilon_1, \ldots \epsilon_{2p-2} \in \{0, 1\}$ such that $b = \sum_{i=1}^{2p-2} \epsilon_i v_i$.*

Proof. Let $c = (c_1, \ldots, c_{2p-2})$ be a vector whose first $p-1$ coordinates are all the elements of Z_p different from d and whose last $p-1$ coordinates are all the elements of Z_p different from e. By Lemma 2.6 there are $\epsilon_1, \ldots, \epsilon_{2p-2} \in \{0, 1\}$ such that for every $1 \leq i \leq 2p-2$, the i^{th} coordinate of $\sum_{j=1}^{2p-2} \epsilon_j v_j^*$ differs from c_i. These inequalities, for $1 \leq i \leq p-1$, show that the first coordinate of $\sum_{i=1}^{2p-2} \epsilon_j v_j$ is d, as it differs from all the other elements of Z_p. Similarly, the inequalities for $p \leq i \leq 2p-2$ show that the second coordinate of that sum is e, as required. □

A *line* in $Z_p \oplus Z_p$ is the set of all vectors $\{x + ty : t \in Z_p\}$, where x and y are some fixed vectors in $Z_p \oplus Z_p$. It is easy to see that if a subset of $Z_p \oplus Z_p$ is not contained in a line then it contains three vectors u, v and w such that the set $u - w, v - w$ forms a basis of $Z_p \oplus Z_p$.

Lemma 3.4 *Let S be a sequence of $6p-7$ vectors in $Z_p \oplus Z_p$ and suppose that no line contains more than $2p-2$ members of S. Then there is a subsequence of $4p-4$ members of S, which we denote by $a_1, a_2, \ldots, a_{4p-4}$, so that if $b_i = a_{2i} - a_{2i-1}$ for $1 \leq i \leq 2p-2$, and A is the $2p-2$ by $2p-2$ matrix whose columns are the vectors b_i^*, then in Z_p, $Per(A) \neq 0$.*

Proof. Let e_1 and e_2 be the standard basis of $Z_p \oplus Z_p$ and let A_0 be the $2p-2$ by $2p-2$ matrix each of whose first $p-1$ columns is e_1^* and each of whose last $p-1$ columns is e_2^*. Trivially, $Per(A_0) = ((p-1)!)^2 \neq 0 (mod\ p)$. We next define the elements $a_1, a_2, \ldots, a_{4p-4}$ sequentially, so that after defining $a_1, \ldots, a_{2i}$ and $b_j = a_{2j} - a_{2j-1}$ for $1 \leq j \leq i$, the following holds. Let A_i be the matrix obtained from A_0 by replacing its first i columns by the columns $b_1^*, \ldots, b_i^*$, then in Z_p, $Per(A_i) \neq 0$. The desired result is just the existence of A_{2p-2}. Suppose $0 \leq i < 2p-2$ and suppose $a_1, \ldots, a_{2i}, b_1, \ldots b_i$ and A_i have already been defined as above such that $Per(A_i) \neq 0$. Our objective is to define a_{2i+1}, a_{2i+2}, b_{i+1} and A_{i+1} with the required properties. Put $S' = S \setminus \{a_1, \ldots, a_{2i}\}$. Clearly $|S'| \geq 6p - 7 - (2p-3)2 = 2p-1$. Therefore, S' is not contained in a line and it thus contains three vectors u, v and w so that $u - w, v - w$ is a basis. It follows that the column number $i+1$ of the matrix A_i is a linear combination of the two column vectors $(u-w)^*$ and $(v-w)^*$. By the multilinearity of the permanent this implies that $Per(A_i)$ is a linear combination of the permanents of the two matrices obtained from it by replacing its $i+1$ column by $(u-w)^*$ and by $(v-w)^*$. Hence at least one of these two permanents, say the first, is non-zero modulo p. We can now define $a_{2i+2} = u$ and $a_{2i+1} = w$, completing the proof. □

Proof of Proposition 3.1. Let S be a sequence of $6p-5$ elements of $Z_p \oplus Z_p$. Suppose, first, that there are $2p-1$ members of S on the line $\{x + ty : t \in Z_p\}$, where $x, y \in Z_p \oplus Z_p$. Let $x + t_i y$,

$(1 \leq i \leq 2p-1)$ be these elements. By Proposition 2.1 there is an $I \subset \{1, \ldots 2p-1\}$, $|I| = p$, so that $\sum_{i \in I} t_i \equiv 0 (mod\ p)$. Hence in $Z_p \oplus Z_p$, $\sum_{i \in I}(x + t_i y) = 0$, completing the proof in this case. Therefore, we may assume that there is no line containing more than $2p-2$ elements of S. By Lemma 3.4 we can renumber the elements of S by $a_1, \ldots, a_{6p-5}$ so that if $b_i = a_{2i} - a_{2i-1}$ for $1 \leq i \leq 2p-2$ then the permanent of the $2p-2$ by $2p-2$ matrix A whose columns are the vectors b_i^* is not zero modulo p. Therefore, by Lemma 3.3, every vector is a $0,1$-linear combination of the vectors b_i. In particular, there are $\epsilon_1, \ldots, \epsilon_{2p-2} \in \{0,1\}$ so that

$$\sum_{i=1}^{2p-2} \epsilon_i b_i = -a_1 - a_3 - \ldots - a_{4p-5} - a_{4p-3} - a_{4p-2} - a_{4p-1} - \ldots - a_{5p-2}.$$

By the definition of the vectors b_i this gives that in $Z_p \oplus Z_p$

$$a_{4p-3} + a_{4p-2} + a_{4p-1} + \ldots + a_{5p-2} + \sum_{i=1}^{2p-2} [a_{2i-1} + \epsilon_i (a_{2i} - a_{2i-1}) = 0.$$

Therefore, there is a $3p$-subset of S the sum of whose elements is 0, and by Lemma 3.2 it contains a zero-sum p-subset, completing the proof of Proposition 3.1 and establishing Theorem 1.3. $\Box$

Note that the above proof actually establishes a slightly stronger version of Proposition 3.1, as it shows that the $6p-5$ term in its statement can be reduced to $6p-7$. By modifying the above proof it is possible to improve the assertion of the Proposition further and show that for all sufficiently large primes p, any sequence of $5p-2$ elements of $Z_p \oplus Z_p$ contains a zero-sum p-subset. Since this still does not yield a proof of Conjecture 1.2 we only sketch the argument. The new idea is that by a more careful proof of Lemma 3.4 one can establish the following version of it.

Lemma 3.5 *Let S be a sequence of $5p-2$ vectors in $Z_p \oplus Z_p$, where p is a large prime, and suppose that no line contains more than $2p-2$ members of S. Then there is a subsequence of $4p-4$ members of S, which we denote by $a_1, a_2, \ldots, a_{4p-4}$, so that if $b_i = a_{2i} - a_{2i-1}$ for $1 \leq i \leq 2p-2$, and A is the $2p-2$ by $2p-2$ matrix whose columns are the vectors b_i^*, then in Z_p, $Per(A) \neq 0$.*

Proof (sketch). Let $\delta > 0$ be a fixed small constant. Split S into two random subsets S_1 and S_2 by choosing each element of S, randomly and independently, to be a member of S_1 with probability $(1+\delta)/2$ and a member of S_2 with probability $(1-\delta)/2$. By the standard estimates for Binomial distributions (see, e.g., [6], Appendix A), if p is sufficiently large as a function of δ then the following two conditions hold with high probability. (For the proof of the second condition the fact that the total number of lines is $O(p^2)$ has to be used.)
(i) $|S_1| \leq (\frac{1}{2} + \delta)(5p-2)$ (and hence $|S_2| \geq (\frac{1}{2} - \delta)(5p-2)$.)
(ii) Every line in $Z_p \oplus Z_p$ contains at most $p + 2\delta p$ members of S_1 and at most p members of S_2.

Fix a partition $S = S_1 \cup S_2$ satisfying (i) and (ii). We can now repeat the proof of Lemma 3.4, where we start choosing the triples of non-collinear vectors u, v and w among the members of S_1, as long as this is possible. Since no line contains more than $p + 2\delta p$ members of S_1 this can be done until at most that many elements of S_1 are left. Now we start choosing the required triples u, v, w in the next steps by always taking w from the remaining elements of S_1 and u and v from S_2. Observe that this can be done as long as there are elements left in S_1. To see this, note that by (i) and (ii) when j elements are left in S_1, then the number of elements in S_2 is still at least

$$(\frac{1}{2} - \delta)(5p - 2) - (p + 2\delta p - j) > p,$$

(provided δ is small and p is large enough). However, no line contains more than p elements among the remaining ones in S_2, and hence the required choices can be completed. Finally, only some of the elements of S_2 are left, and we can choose non-collinear triples u, v, w among those until only $p+2$ elements are left, since no line contains more than p elements of S_2. This completes the proof. □

The last lemma, together with the arguments in the proof of Proposition 3.1, clearly show that for large primes p we can replace the $6p - 5$ in the statement of the proposition by $5p - 2$.

We conclude this section with a sketch of a rather simple proof of a slightly weaker version of Proposition 3.1.

Lemma 3.6 *Let $B \subset Z_p \oplus Z_p$ be a set of cardinality $|B| = m < p^2/2$. Suppose u, v and w is a non-collinear triple of vectors in $Z_p \oplus Z_p$. Then, there is a subset C consisting of two of the elements u, v and w, so that $|B + C| \geq m + \lceil \sqrt{m} \rceil$.*

Proof. By applying an affine transformation to B and to $\{u, v, w\}$ we may assume that $\{u, v, w\}$ is simply the set $\{(0,0), (1,0), (0,1)\}$. We next show that either $C = \{(0,0), (1,0)\}$ or $C = \{(0,0), (0,1)\}$ will do. Let B_1 be the projection of B on the x-axis, and let B_2 be the projection on the y-axis. Since $|B| \leq |B_1| \cdot |B_2|$ either B_1 or B_2 contain at least $\sqrt{m}$ elements. Assume, without loss of generality, that $|B_1| \geq \sqrt{m}$. If B contains no full vertical line, then clearly, $C = \{(0,0), (0,1)\}$ will complete the proof. Otherwise, $|B_2| = p$ and again the conclusion follows unless B contains more than $p - \sqrt{m}$ full horizontal lines. But then $|B_1| = p$ as well, and we may assume that B contains more than $p - \sqrt{m}$ full vertical lines as well. Therefore, $|B| > 2p(p - \sqrt{m}) - (p - \sqrt{m})^2 = p^2 - m$, contradicting the assumption that $|B| = m < p^2/2$. □

Corollary 3.7 *If p is a prime, then any sequence S of at least $8\lceil p/\sqrt{2} \rceil + 2p - 11$ elements of $Z_p \oplus Z_p$ contains a zero-sum p-subset.*

Proof (sketch). Define $j = \lceil p/\sqrt{2} \rceil$. If $2p-1$ elements of S lie on a line, the result follows immediately, by Proposition 2.1. Otherwise, starting with a two element subset of S, we can repeatedly apply Lemma 3.6 to obtain $2j-2$ pairwise disjoint subsets $S_1, \ldots, S_{2j-2}$ of S, where $|S_i| = 2$ for all i and $|S_1 + S_2 + \ldots + S_{2j-2}| \geq j^2 > p^2/2$. In the same manner there are another $2j-2$ pairwise disjoint subsets $T_1, \ldots, T_{2j-2}$ of cardinality 2 each among the remaining elements of S so that $|T_1 + T_2 + \ldots + T_{2j-2}| \geq j^2 > p^2/2$. (Here we used the fact that during the whole process, before the last subset T_{2j-2} has been chosen there have still been at least $2p-1$ remaining elements of S, which are not all on a line. Again, the estimates can be somewhat improved for large p by applying the argument in the proof of Lemma 3.5.) Therefore, every element of $Z_p \oplus Z_p$ is in $S_1 + \ldots + S_{2j-2} + T_1 + \ldots + T_{2j-2}$. Since $4j-4 < 3p$ we can conclude, as in the proof of Proposition 3.1, that there are $3p$ elements of S whose sum is 0, and complete the proof by applying Lemma 3.2. $\Box$

4 Concluding remarks

The derivation of Theorem 1.3 from Proposition 3.1 can be easily extended to a proof of the following more general result, whose simple detailed proof is omitted.

Proposition 4.1 *Let G and H be two finite abelian groups, where $|G| = |H| = n$. Then any sequence of $6n-5$ elements in the group $G \oplus H$ contains an n-subset the sum of whose elements in $G \oplus H$ is 0.*

It is natural to try to extend Conjecture 1.2 and Theorem 1.3 to higher dimensions. Recall that $s(n,d)$ denotes the smallest number s such that any sequence of s elements in $(Z_n)^d$ contains a 0-sum n-subsequence. By the Erdős Ginzburg Ziv Theorem, $s(n,1) = 2n-1$ and if Conjecture 1.2 holds then $s(n,2) = 4n-3$. It is easy to see ([24]) that $s(n,d) \geq 2^d(n-1)+1$, as the sequence consisting of $n-1$ copies of each of the 2^n 0, 1-vectors of length d contains no 0-sum n-subsequence. Trivially $s(n,d) \leq (n-1)n^d + 1$, since any sequence of $(n-1)n^d + 1$ elements of $(Z_n)^d$ must contain the same vector n times. Thus, $s(2,d) = 2^d + 1$. An easy argument of [24] shows that $s(n_1 n_2, d) \leq s(n_1, d) + n_1(s(n_2, d) - 1)$ and since as shown by Kemnitz ([25]) $s(p,2) = 4p-3$ for $p = 2, 3, 5, 7$ this implies that $s(n,2) = 4n-3$ for all $n = 2^a 3^b 5^c 7^d$. Similarly, since $s(2,d) = 2^d + 1$ it follows that $s(2^a, d) = (2^a - 1)2^d + 1$.

Various researchers observed that $s(3,3) \geq 19$ $(> 8 \cdot 2 + 1)$. Examples appear in [24], [14], [25] and [26], where it is also shown that in fact $s(3,3) = 19$. In [25] it is shown that $s(3,4) = 41$; this has also been observed more recently by Doyle, Pemantle and Schwartz (private communication from Y. Peres). It is worth noting that by the main result in [15], [20], $s(3,d) = o(3^d)$, but it seems

very difficult to show that there exists some absolute constant $\delta > 0$ such that $s(3,d) \leq (3-\delta)^d$ for all sufficiently large d. The problem of determining $s(n,d)$ precisely for all n and d seems extremely difficult.

As mentioned in inequality (1) in the introduction we can prove that for every fixed d, $s(n,d)$ grows linearly with n. The proof of this fact combines the basic method of [19] (see also [28], [2]) with a Harmonic Analysis approach. Here is a very brief outline of the argument.

The proof is by induction on the dimension d, where Theorem 1.1 provides the beginning of the induction as it shows that $c(1) = 2$. It is convenient to prove the result for prime n, and this clearly suffices. Suppose we already know that $c(d-1)$ exists and try to prove that $c(d) \leq c'(d) \cdot c(d-1)$ for an appropriately defined $c'(d)$. Let S be a sequence of at least $c'(d)c(d-1)n$ elements of $(Z_n)^d$. If there are at least $c(d-1)n$ of them on a hyperplane, i.e., there is a vector v in $(Z_n)^d$ so that its scalar product (modulo n) with at least $c(d-1)n$ members of S is a constant, then the desired result follows, by the induction hypothesis. Thus we may assume that this is not the case. Define the *width* $w(T)$ of any sequence T of elements of $(Z_n)^d$ as follows. For each vector $v \in (Z_n)^d$ let $w_v(T)$ be the minimum length of a cyclic interval in Z_n that contains all the scalar products of members of T with v. The width $w(T)$ is simply the minimum of $w_v(T)$ over all the nonzero vectors $v \in (Z_n)^d$. Observe that since no hyperplane contains $c(d-1)n$ elements of S, the width of any subsequence T of, say, more than half the elements of S, exceeds $c'(d)/2$. We next show that this implies that there is an $l \leq p$ so that every element of $(Z_n)^d$ is a sum of a subsequence of cardinality l (and hence also a sum of a subsequence of cardinality p) of S. To do so we define, repeatedly, pairwise disjoint subsequences $S_1, S_2, \ldots$ of S, so that each S_i is of cardinality 2, and $|S_1 + \ldots + S_i|$ grows sufficiently fast. We stop once the size of this subset of $(Z_n)^d$ exceeds $n^d/2$, do it again with sets T_i and obtain the desired result as in the proof of Corollary 3.7. The main difficulty is to show that after $S_1, \ldots, S_{i-1}$ have been defined, we can still choose a set S_i among the remaining elements of S so that if S^* denotes the set of all members of $(Z_n)^d$ which belong to $S_1 + \ldots S_{i-1}$ then $|S^* + S_i|$ is sufficiently large. To this end we apply the simple (and clever) combinatorial method of [19]. This method implies the following. Let S^* and T be two subsets of an Abelian group, and let $T(k)$ denote the set of all elements of the group which can be written as a sum of at most k members of $T \cup (-T)$. Then, there is an element t of T so that

$$|(t + S^*) \setminus S^*| \geq |S^*| \frac{1}{k} (1 - \frac{|S^*|}{|T(k)|}). \tag{4}$$

Suppose, now, that $S_1, \ldots, S_{i-1}$ have already been defined and S^* is the set of all members of $(Z_n)^d$ that belong to $S_1 + \ldots + S_{i-1}$. Let T be the set of all the remaining members of S. Our objective is to define S_i so that $|S^* + S_i|$ will be sufficiently large. Observe that T contains (much)

more than half the members of S and hence has a large width. By an affine transformation we may assume that T contains the 0 vector and the standard basis of $(Z_n)^d$. The crucial point now is to show that since T has a large width, $T(k/2)$ is rather "dense" in the cube $(Z_n)^d$. This is done by defining appropriate "bump" functions and by analysing their Fourier transforms and that of the characteristic function of various subsets of $T(k)$. Once the asserted density is proved, the unit vectors in T can be used to show that $|T(k)|$ is sufficiently large. This can be applied to establish, using (4), the required lower bound for $|S_1 + \ldots + S_i|$ and complete the proof. The details are somewhat lengthy and will not appear here.

Finally we remark that the Chevalley Warning theorem supplies easy proofs for some (seemingly) minor variations of Conjecture 1.2 and its higher dimensional analogues. For example, it is easy to show that for any prime p and any integer d, any sequence of $(d+1)(p-1)+1$ terms in $(Z_p)^d$ contains a zero-sum subset whose cardinality is divisible by p. The problem of obtaining a zero-sum subset of cardinality p precisely seems much more difficult.

Acknowledgement We would like to thank Y. Peres for helpful discussions and comments.

References

[1] N. Alon, *Tools from higher algebra*, 1986. To appear in: "Handbook in Combinatorics", R. L. Graham, M. Grötschel and L. Lovász eds., North Holland.

[2] N. Alon, *Subset sums*, J. Number Theory 27 (1987), 196-205.

[3] N. Alon and Y. Caro, *On three zero-sum Ramsey-type problems*, J. Graph Theory, to appear.

[4] N. Alon, S. Friedland and G. Kalai, *Regular subgraphs of almost regular graphs*, J. Combinatorial Theory, Ser. B, 37 (1984), 79-91.

[5] N. Alon, N. Linial and R. Meshulam, *Additive bases of vector spaces over prime fields*, J. Combinatorial Theory, Ser. A 57 (1991), 203-210.

[6] N. Alon and J. H. Spencer, **The probabilistic method**, Wiley, 1991.

[7] N. Alon and M. Tarsi, *A nowhere-zero point in linear mappings*, Combinatorica 9 (1989), 393-395.

[8] A. Bialostocki and P. Dierker, *Zero sum Ramsey theorems*, Congressus Numerantium 70 (1990), 119-130.

[9] A. Bialostocki and P. Dierker, *On the Erdős Ginzburg Ziv theorem and the Ramsey numbers for stars and matchings*, Discrete Math. 110 (1992), 1-8.

[10] C. Bailey and R. B. Richter, *Sum zero (mod n), size n subsets of integers*, Amer. Math. Monthly 96 (1989), 240-242.

[11] R. C. Baker and W. Schmidt, *Diophantine problems in variables restricted to the values 0 and 1*, J. Number Theory 12 (1980), 460-486.

[12] A. Blokhuis, *Polynomials in finite geometries and combinatorics*, Proc. 14^{th} British Combinatorial Conference, London Mathematical Society Lecture Notes Series 187, edited by K. Walker, Cambridge University Press, 1993, 35-52.

[13] Z. I. Borevich and I. R. Shafarevich, **Number Theory**, Academic Press, New York, 1966.

[14] J. L. Brenner, Problem 6298, Amer. Math. Monthly 89 (1982), 279-280.

[15] T. C. Brown and J. C. Buhler, *A density version of a geometric Ramsey theorem*, J. Combinatorial Theory, Ser. A 32 (1982), 20-34.

[16] H. Davenport, *On the addition of residue classes*, J. London Math. Soc. 10 (1935), 30-32.

[17] P. van Emde Boas and D. Kruyswijk, *A combinatorial problem on finite abelian groups III*, Z. W. 1969-008 (Math. Centrum- Amsterdam).

[18] P. Erdős, A. Ginzburg and A. Ziv, *Theorem in the additive number theory*, Bull. Research Council Israel 10F (1961), 41-43.

[19] P. Erdős and H. Heilbronn, *On the addition of residue classes mod p*, Acta Arith. 9 (1964), 149-159.

[20] P. Frankl, R. L. Graham and V. Rödl, *On subsets of Abelian groups with no 3-term arithmetic progression*, J. Combinatorial Theory Ser. A 45 (1987), 157-161.

[21] Z. Füredi and D. J. Kleitman, *On zero-trees*, J. Graph Theory 16 (1992), 107-120.

[22] Weidong Gao, *Any* $2n-1$ *integers contain exactly* n *integers whose sum is a multiple of* n, J. of Northeast Normal University 4 (1985) (in Chinese).

[23] A. Geroldinger and R. Schneider, *On Davenport's constant*, J. Combinatorial Theory, Ser. A 61 (1992), 147-152.

[24] H. Harborth, *Ein Extremalproblem für Gitterpunkte*, J. Reine Angew. Math. 262/263 (1973), 356-360.

[25] A. Kemnitz, *On a lattice point problem*, Ars Combinatoria 16b (1983), 151-160.

[26] B. Leeb and C. Stahlke, *A problem on lattice points*, Crux Mathematicorum 13 (1987), 104-108.

[27] R. Meshulam, *An uncertainty inequality and zero subsums*, Disc. Math. 84 (1990), 197-200.

[28] J. E. Olson, *An addition theorem modulo p*, J. Combinatorial Theory 5 (1968), 45-52.

[29] J. E. Olson, *A combinatorial problem on finite abelian groups, I*, J. Number Theory 1 (1969), 8-10.

[30] J. E. Olson, *A combinatorial problem on finite abelian groups, II*, J. Number Theory 1 (1969), 195-199.

[31] T. Redmond and C. Ryavec, The Mathematical Intelligencer 2 (1980), 106.

[32] L. Schrijver and P. D. Seymour, *A simpler proof and a generalization of the zero-trees theorem*, J. Combinatorial Theory, Ser. A 58 (1991), 301-305.

[33] W. M. Schmidt, **Equations over finite fields, an elementary approach**, Springer Verlag Lecture Notes in Math., 1976.

[34] N. Zimmerman, private communication.