DIMACS Technical Report 94-2
January 1994

Ballot numbers, alternating products, and the Erdős-Heilbronn conjecture

by

Melvyn B. Nathanson [1]
Department of Mathematics
Lehman College (CUNY)
Bronx, New York 10468
e-mail: nathansn@dimacs.rutgers.edu

[1]This research was supported in part by grants from the PSC-CUNY Research Award Program. The paper was written while I was a visitor at the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) at Rutgers University, and I thank DIMACS for its hospitality.

DIMACS is a cooperative project of Rutgers University, Princeton University, AT&T Bell Laboratories and Bellcore.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

1 Introduction

Let A be a subset of an abelian group. Let hA denote the set of all sums of h elements of A with repetitions allowed, and let $h\hat{}A$ denote the set of all sums of h distinct elements of A, that is, all sums of the form $a_1 + \cdots + a_h$, where $a_1, \ldots, a_h \in A$ and $a_i \neq a_j$ for $i \neq j$.

Let A be a set of k congruence classes modulo a prime p. The Cauchy-Davenport theorem states that

$$|2A| \geq \min(p, 2k-1),$$

and, by induction,

$$|hA| \geq \min(p, hk - h + 1)$$

for every $h \geq 2$. Erdős and Heilbronn conjectured 30 years ago that

$$|2\hat{}A| \geq \min(p, 2k-3).$$

They did not include this conjecture in their paper on addition of residue classes [9], but Erdős has frequently mentioned this problem in lectures and papers (for example, Erdős-Graham [8, p. 95]). Dias da Silva and Hamidoune recently prove this conjecture. They used linear algebra and the representation theory of the symmetric group to show that

$$|h\hat{}A| \geq \min(p, hk - h^2 + 1)$$

for every $h \geq 2$.

The purpose of this paper is to give a complete and elementary exposition of this proof. Instead of representation theory, we will use the combinatorics of the h-dimensional ballot numbers.

2 Multi-dimensional ballot numbers

The standard basis for $\mathbf{R}^h$ is the set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$, where

$$\begin{aligned} \mathbf{e}_1 &= (1,0,0,0,\ldots,0) \\ \mathbf{e}_2 &= (0,1,0,0,\ldots,0) \\ &\vdots \\ \mathbf{e}_h &= (0,0,0,\ldots,0,1). \end{aligned}$$

The lattice $\mathbf{Z}^h$ is the subgroup of $\mathbf{R}^h$ generated by the set $\{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$, so $\mathbf{Z}^h$ is the set of vectors in $\mathbf{R}^h$ with integral coordinates. Let

$$\mathbf{a} = (a_0, a_1, \ldots, a_{h-1}) \in \mathbf{Z}^h$$

and

$$\mathbf{b} = (b_0, b_1, \ldots, b_{h-1}) \in \mathbf{Z}^h.$$

A *path* in $\mathbf{Z}^h$ is a finite sequence of lattice points

$$\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$$

such that

$$\mathbf{v}_j - \mathbf{v}_{j-1} \in \{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$$

for $j = 1, \ldots, m$. Let $\mathbf{v}_{j-1}, \mathbf{v}_j$ be successive points on a path. We call this a *step* in the direction $\mathbf{e}_i$ if

$$\mathbf{v}_j = \mathbf{v}_{j-1} + \mathbf{e}_i.$$

The vector $\mathbf{a}$ is called *nonnegative* if $a_i \geq 0$ for $i = 0, 1, \ldots, h-1$. We write

$$\mathbf{a} \leq \mathbf{b}$$

if $\mathbf{b} - \mathbf{a}$ is a nonnegative vector.

Let $P(\mathbf{a}, \mathbf{b})$ denote the number of paths from $\mathbf{a}$ to $\mathbf{b}$. The path function $P(\mathbf{a}, \mathbf{b})$ is translation invariant in the sense that

$$P(\mathbf{a} + \mathbf{c}, \mathbf{b} + \mathbf{c}) = P(\mathbf{a}, \mathbf{b})$$

for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{Z}^h$. In particular,

$$P(\mathbf{a}, \mathbf{b}) = P(\mathbf{0}, \mathbf{b} - \mathbf{a}).$$

The path function satisfies the boundary conditions

$$P(\mathbf{a}, \mathbf{a}) = 1,$$

and

$$P(\mathbf{a}, \mathbf{b}) > 0 \text{ if and only if } \mathbf{a} \leq \mathbf{b},$$

If $\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$ is a path, then

$$\mathbf{v}_{m-1} = \mathbf{b} - \mathbf{e}_i$$

for some $i = 1, \ldots, h$, and there is a unique path from $\mathbf{b} - \mathbf{e}_i$ to $\mathbf{b}$. It follows that the path counting function $P(\mathbf{a}, \mathbf{b})$ also satisfies the difference equation

$$P(\mathbf{a}, \mathbf{b}) = \sum_{i=0}^{h-1} P(\mathbf{a}, \mathbf{b} - \mathbf{e}_i).$$

Let $\mathbf{a} \leq \mathbf{b}$. For $i = 0, 1, \ldots, k-1$, every path from $\mathbf{a}$ to $\mathbf{b}$ contains exactly $b_i - a_i$ steps in the direction $\mathbf{e}_{i+1}$. Let

$$m = \sum_{i=0}^{h-1} (b_i - a_i).$$

Every path from $\mathbf{a}$ to $\mathbf{b}$ has exactly m steps, and the number of different paths is the multinomial coefficient

$$P(\mathbf{a}, \mathbf{b}) = \frac{\left(\sum_{i=0}^{h-1} (b_i - a_i)\right)!}{\prod_{i=0}^{h-1} (b_i - a_i)!} = \frac{m!}{\prod_{i=0}^{h-1} (b_i - a_i)!}. \tag{1}$$

Let $h \geq 2$. There are h candidates in an election. The candidates will be labelled by the integers $0, 1, \ldots, h-1$. Suppose that m_0 votes have already been cast, and that candidate i has received a_i votes. Then

$$m_0 = a_0 + a_1 + \cdots + a_{h-1}.$$

We shall call

$$\mathbf{v}_0 = \mathbf{a} = (a_0, a_1, \ldots, a_{h-1})$$

the initial *ballot vector*. There are m remaining voters, each of whom has one vote, and these votes will be cast sequentially. Let $v_{i,k}$ denote the number of votes that candidate i has received after k additional votes have been cast. We represent the distribution of votes at step k by the ballot vector

$$\mathbf{v}_k = (v_{0,k}, v_{1,k}, \ldots, v_{h-1,k}).$$

Then

$$v_{0,k} + v_{1,k} + \cdots + v_{h-1,k} = k + m_0$$

for $k = 0, 1, \ldots, m$. Let

$$\mathbf{v}_m = \mathbf{b} = (b_0, b_1, \ldots, b_{h-1})$$

be the final ballot vector. It follows immediately from the definition of the ballot vectors that

$$\mathbf{v}_k - \mathbf{v}_{k-1} \in \{\mathbf{e}_1, \ldots, \mathbf{e}_h\}$$

for $k = 1, \ldots, m$, and so

$$\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$$

is a path in $\mathbf{Z}^h$ from $\mathbf{a}$ to $\mathbf{b}$. Therefore, the number of distinct sequences of m votes that can lead from the initial ballot vector $\mathbf{a}$ to the final ballot vector $\mathbf{b}$ is the multinomial coefficient

$$\frac{\left(\sum_{i=0}^{h-1} (b_i - a_i)\right)!}{\prod_{i=0}^{h-1} (b_i - a_i)!} = \frac{m!}{\prod_{i=0}^{h-1} (b_i - a_i)!}.$$

Let $\mathbf{v} = (v_1, \ldots, v_h)$ and $\mathbf{w} = (w_1, \ldots, w_h)$ be vectors in $\mathbf{R}^h$. The vector $\mathbf{v}$ will be called *increasing* if

$$v_1 \leq v_2 \leq \cdots \leq v_h,$$

and *strictly increasing* if

$$v_1 < v_2 < \cdots < v_h.$$

Now suppose that the initial ballot vector is

$$\mathbf{a} = (0, 0, 0, \ldots, 0)$$

and that the final ballot vector $\mathbf{b} = (b_0, b_1, \ldots, b_{h-1})$ is nonnegative and increasing. Let

$$m = b_0 + b_1 + \ldots + b_{h-1}.$$

Let $B(b_0, b_1, \ldots, b_{h-1})$ denote the number of ways that m votes can be cast so that all of the k-th ballot vectors are nonnegative and increasing. This is the classical *h-dimensional ballot number*. Observe that

$$B(0, 0, \ldots, 0) = 1$$

and

$$B(b_0, b_1, \ldots, b_{h-1}) > 0$$

if and only if $(b_0, b_1, \ldots, b_{h-1})$ is a nonnegative, increasing vector. These boundary conditions and the difference equation

$$B(b_0, b_1, \ldots, b_{h-1}) = \sum_{i=0}^{h-1} B(b_0, \ldots, b_{i-1}, b_i - 1, b_{i+1}, \ldots, b_{h-1})$$

completely determine the function $B(b_0, b_1, \ldots, b_{h-1})$.

There is an equivalent combinatorial problem. Suppose that the initial ballot vector is

$$\mathbf{a}^* = (0, 1, 2, \ldots, h-1),$$

and that the final ballot vector

$$\mathbf{b} = (b_0, b_1, \ldots, b_{h-1})$$

is nonnegative and strictly increasing. Let

$$m = \sum_{i=0}^{h-1} (b_i - i) = \sum_{i=0}^{h-1} b_i - \binom{h}{2}.$$

Let $\hat{B}(b_0, b_1, \ldots, b_{h-1})$ denote the number of ways that m votes can be cast so that all of the ballot vectors $\mathbf{v}_k$ are nonnegative and strictly increasing. We shall call this the *strict h-dimensional ballot number.*.

A path $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m$ in $\mathbf{Z}^h$ will be called *strictly increasing* if every lattice point $\mathbf{v}_k$ on the path is strictly increasing. Then $\hat{B}(b_0, b_1, \ldots, b_{h-1})$ is the number of strictly increasing paths from $\mathbf{a}^*$ to $\mathbf{b} = (b_0, \ldots, b_{h-1})$.

The strict h-dimensional ballot numbers satisfy the boundary conditions

$$\hat{B}(0, 1, \ldots, h-1) = 1$$

and

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) > 0$$

if and only if $(b_0, b_1, \ldots, b_{h-1})$ is a nonnegative, strictly increasing vector. These boundary conditions and the difference equation

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) = \sum_{i=0}^{h-1} \hat{B}(b_0, \ldots, b_{i-1}, b_i - 1, b_{i+1}, \ldots, b_{h-1})$$

completely determine $\hat{B}(b_0, b_1, \ldots, b_{h-1})$.

There is a simple relationship between the h-dimensional ballot numbers $B(b_0, b_1, \ldots, b_{h-1})$ and $\hat{B}(b_0, b_1, \ldots, b_{h-1})$. The lattice point

$$\mathbf{v} = (v_0, v_1, \ldots, v_{h-1})$$

is nonnegative and strictly increasing if and only if the lattice point

$$\mathbf{v}' = \mathbf{v} - (0, 1, 2, \ldots, h-1) = \mathbf{v} - \mathbf{a}^*$$

is nonnegative and increasing. It follows that

$$\mathbf{a}^* = \mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m = \mathbf{b}$$

is a path of strictly increasing vectors from $\mathbf{a}^*$ to $\mathbf{b}$ if and only if

$$\mathbf{0}, \mathbf{v}_1 - \mathbf{a}^*, \mathbf{v}_2 - \mathbf{a}^*, \ldots, \mathbf{b} - \mathbf{a}^*$$

is a path of increasing vectors from $\mathbf{0}$ to $\mathbf{b} - \mathbf{a}^*$. Thus,

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) = B(b_0, b_1 - 1, b_2 - 2, \ldots, b_{h-1} - (h-1)).$$

For $1 \le i < j \le h$, let $H_{i,j}$ be the hyperplane in $\mathbf{R}^h$ consisting of all vectors $(x_1, \ldots, x_n)$ such that $x_i = x_j$. There are $\binom{h}{2}$ such hyperplanes. A path

$$\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m = \mathbf{b}$$

will be called *intersecting* if there exists at least one vector $\mathbf{v}_k$ on the path such that $\mathbf{v}_k \in H_{i,j}$ for some hyperplane $H_{i,j}$.

The symmetric group S_h acts on $\mathbf{R}^h$ as follows: For $\sigma \in S_h$ and $\mathbf{v} = (v_0, v_1, \ldots, v_{h-1}) \in \mathbf{R}^h$, let

$$\sigma\mathbf{v} = (v_{\sigma(0)}, v_{\sigma(1)}, \ldots, v_{\sigma(h-1)}).$$

A path is intersecting if and only if there is a transposition $\tau = (i, j) \in S_h$ such that $\tau\mathbf{v}_k = \mathbf{v}_k$ for some lattice point $\mathbf{v}_k$ on the path.

Let $I(\mathbf{a}, \mathbf{b})$ denote the number of intersecting paths from $\mathbf{a}$ to $\mathbf{b}$. Let $J(\mathbf{a}, \mathbf{b})$ denote the number of paths from $\mathbf{a}$ to $\mathbf{b}$ that do not intersect any of the hyperplanes $H_{i,j}$. Then

$$P(\mathbf{a}, \mathbf{b}) = I(\mathbf{a}, \mathbf{b}) + J(\mathbf{a}, \mathbf{b}).$$

Lemma 1 *Let* $\mathbf{a}$ *be a lattice point in* $\mathbf{Z}^h$*, and let* $\mathbf{b} = (b_0, \ldots, b_{h-1})$ *be a strictly increasing lattice point in* $\mathbf{Z}^h$*. A path from* $\mathbf{a}$ *to* $\mathbf{b}$ *is strictly increasing if and only if it intersects none of the hyperplanes* $H_{i,j}$*, and so*

$$\hat{B}(b_0, \ldots, b_{h-1}) = J(\mathbf{a}^*, \mathbf{b}).$$

Proof. Let $\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$ be a path, and let

$$\mathbf{v}_k = (v_{0,k}, v_{1,k}, \ldots, v_{h-1,k})$$

for $k = 0, 1, \ldots, m$. If the path is strictly increasing, then every vector on the path is strictly increasing, and so the path does not intersect any of the hyperplanes $H_{i,j}$. Conversely, if the path is not strictly increasing, then there exists a greatest integer k such that the lattice point $\mathbf{v}_{k-1}$ is not strictly increasing. Then $1 \leq k \leq m$, and

$$v_{j,k-1} \leq v_{j-1,k-1}$$

for some $j = 2, \ldots, n$. Since the vector $\mathbf{v}_k$ is strictly increasing, we have

$$v_{j-1,k} \leq v_{j,k} - 1.$$

Since $\mathbf{v}_{k-1}$ and $\mathbf{v}_k$ are successive vectors in a path, we have

$$v_{j-1,k-1} \leq v_{j-1,k}$$

and

$$v_{j,k} - 1 \leq v_{j,k-1}.$$

Combining these inequalities, we obtain

$$v_{j,k-1} \leq v_{j-1,k-1} \leq v_{j-1,k} \leq v_{j,k} - 1 \leq v_{j,k-1}.$$

This implies that

$$v_{j,k-1} = v_{j-1,k-1}$$

and so the vector $\mathbf{v}_{k-1}$ lies on the hyperplane $H_{j-1,j}$. Therefore, if $\mathbf{b}$ is a strictly increasing vector, then a path from $\mathbf{a}$ to $\mathbf{b}$ is strictly increasing if and only if it is non-intersecting, and so $J(\mathbf{a}, \mathbf{b})$ is equal to the number of strictly increasing paths from $\mathbf{a}$ to $\mathbf{b}$. It follows that $J(\mathbf{a}^*, \mathbf{b})$ is equal to the strict ballot number $\hat{B}(b_0, \ldots, b_{h-1})$. This completes the proof.

Lemma 2 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be strictly increasing vectors. Then*

$$P(\sigma\mathbf{a}, \mathbf{b}) = I(\sigma\mathbf{a}, \mathbf{b})$$

for every $\sigma \in S_h, \sigma \neq id$.

Proof. If $\mathbf{a}$ is strictly increasing and if $\sigma \in S_h, \sigma \neq \mathrm{id}$, then $\sigma\mathbf{a}$ is not strictly increasing, and so every path from $\sigma\mathbf{a}$ to $\mathbf{b}$ must intersect at least one of the hyperplanes $H_{i,j}$. This completes the proof.

Lemma 3 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be strictly increasing lattice points. Then*

$$\sum_{\sigma \in S_h} \varepsilon(\sigma) I(\sigma\mathbf{a}, \mathbf{b}) = 0.$$

Proof. Since $\mathbf{a}$ is strictly increasing, it follows that there are $h!$ distinct lattice points of the form $\sigma\mathbf{a}$, where $\sigma \in S_h$, and none of these lattice points lies on a hyperplane $H_{i,j}$. Let Ω be the set of all intersecting paths that start at any one of the $n!$ lattice points $\sigma\mathbf{a}$ and end at $\mathbf{b}$. We shall construct an involution from the set Ω to itself.

Let $\sigma \in S_h$, and let

$$\sigma\mathbf{a} = \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m = \mathbf{b}$$

be a path that intersects at least one of the hyperplanes. Let k be the least integer such that $\mathbf{v}_k \in H_{i,j}$ for some $i < j$. Then $k \geq 1$ since $\mathbf{a}$ is strictly increasing, and the hyperplane $H_{i,j}$ is uniquely determined since $\mathbf{v}_k$ lies on a path. Consider the transposition $\tau = (i,j) \in S_h$. Then

$$\tau\mathbf{v}_k = \mathbf{v}_k \in H_{i,j},$$
$$\tau\sigma\mathbf{a} \neq \sigma\mathbf{a},$$

and

$$\tau\sigma\mathbf{a} = \tau\mathbf{v}_0, \tau\mathbf{v}_1, \ldots, \tau\mathbf{v}_k = \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_m = \mathbf{b}$$

is an intersecting path in Ω from $\tau\sigma\mathbf{a}$ to $\mathbf{b}$. Moreover, k is the least integer such that a lattice point on this new path lies in one of the hyperplanes, and $H_{i,j}$ is still the unique hyperplane containing $\mathbf{v}_k$. Since τ^2 is the identity permutation for every transposition τ, it follows that if we apply the same mapping to this path from $\tau\sigma\mathbf{a}$ to $\mathbf{b}$, we recover the original path from $\sigma\mathbf{a}$ to $\mathbf{b}$. Thus, this mapping is an involution on the set Ω of intersecting paths from the $h!$ lattice points $\sigma\mathbf{a}$ to $\mathbf{b}$. Moreover, if σ is an even (resp. odd) permutation, then an intersecting path from $\sigma\mathbf{a}$ is sent to an intersecting path from $\tau\sigma\mathbf{a}$, where τ is a transposition and so $\tau\sigma$ is an odd (resp. even) permutation. It follows that the number of intersecting paths that start at even permutations of $\mathbf{a}$ is equal to the number of intersecting paths that start at odd permutations of $\mathbf{a}$. This means that

$$\sum_{\substack{\sigma \in S_h \\ \varepsilon(\sigma)=1}} I(\sigma\mathbf{a}, \mathbf{b}) = \sum_{\substack{\sigma \in S_h \\ \varepsilon(\sigma)=-1}} I(\sigma\mathbf{a}, \mathbf{b}).$$

This statement is equivalent to the Lemma.

Let $[x]_k$ denote the polynomial $x(x-1)\cdots(x-k+1)$. If $b_i, \sigma(i)$ are nonnegative integers, then

$$\begin{aligned} [b_i]_{\sigma(i)} &= b_i(b_i-1)(b_i-2)\ldots(b_i-\sigma(i)+1) \\ &= \begin{cases} b_i!/(b_i-\sigma(i))! & \text{if } \sigma(i) \leq b_i \\ 0 & \text{if } \sigma(i) > b_i. \end{cases} \end{aligned}$$

Theorem 1 *Let $h \geq 2$, and let $b_0, b_1, \ldots, b_{h-1}$ be integers such that*

$$0 \leq b_0 < b_1 < \cdots < b_{h-1}.$$

Then

$$\hat{B}(b_0, b_1, \ldots, b_{h-1}) = \frac{(b_0 + b_1 + \cdots + b_{h-1} - \binom{h}{2})!}{b_0! b_1! \cdots b_{h-1}!} \prod_{0 \leq i < j \leq h-1} (b_j - b_i).$$

Proof. Let $\mathbf{a}^* = (0, 1, 2, \ldots, h-1)$, and let $\mathbf{b} = (b_0, b_1, \ldots, b_{h-1}) \in \mathbf{Z}^h$. Applying the previous lemmas, we obtain

$$\begin{aligned}
&\hat{B}(b_0, b_1, \ldots, b_{h-1}) = \\
&= J(\mathbf{a}^*, \mathbf{b}) \\
&= P(\mathbf{a}^*, \mathbf{b}) - I(\mathbf{a}^*, \mathbf{b}) \\
&= P(\mathbf{a}^*, \mathbf{b}) + \sum_{\substack{\sigma \in S_h \\ \sigma \neq id}} \varepsilon(\sigma) I(\sigma \mathbf{a}^*, \mathbf{b}) \\
&= P(\mathbf{a}^*, \mathbf{b}) + \sum_{\substack{\sigma \in S_h \\ \sigma \neq id}} \varepsilon(\sigma) P(\sigma \mathbf{a}^*, \mathbf{b}) \\
&= \sum_{\sigma \in S_h} \varepsilon(\sigma) P(\sigma \mathbf{a}^*, \mathbf{b}) \\
&= \sum_{\substack{\sigma \in S_h \\ \sigma a^* \leq b}} \varepsilon(\sigma) \frac{(b_0 + \cdots + b_{h-1} - \binom{h}{2})!}{\prod_{i=0}^{h-1} (b_i - \sigma(i))!} \\
&= \frac{(b_0 + \cdots + b_{h-1} - \binom{h}{2})!}{b_0! b_1! \cdots b_{h-1}!} \sum_{\substack{\sigma \in S_h \\ \sigma a^* \leq b}} \varepsilon(\sigma) [b_0]_{\sigma(0)} [b_1]_{\sigma(1)} \cdots [b_{h-1}]_{\sigma(h-1)} \\
&= \frac{(b_0 + \cdots + b_{h-1} - \binom{h}{2})!}{b_0! b_1! \cdots b_{h-1}!} \sum_{\sigma \in S_h} \varepsilon(\sigma) [b_0]_{\sigma(0)} [b_1]_{\sigma(1)} \cdots [b_{h-1}]_{\sigma(h-1)} \\
&= \frac{(b_0 + \cdots + b_{h-1} - \binom{h}{2})!}{b_0! b_1! \cdots b_{h-1}!} \begin{vmatrix} 1 & [b_0]_1 & [b_0]_2 & \cdots & [b_0]_{h-1} \\ 1 & [b_1]_1 & [b_1]_2 & \cdots & [b_1]_{h-1} \\ \vdots & & & & \\ 1 & [b_{h-1}]_1 & [b_{h-1}]_2 & \cdots & [b_{h-1}]_{h-1} \end{vmatrix} \\
&= \frac{(b_0 + \cdots + b_{h-1} - \binom{h}{2})!}{b_0! b_1! \cdots b_{h-1}!} \prod_{0 \leq i < j \leq h-1} (b_j - b_i).
\end{aligned}$$

This completes the proof.

We state the following corollary with the notation that is used later in the proof of the Erdős-Heilbronn conjecture.

Corollary 1 *Let $h \geq 2$, let p be a prime number, and let $i_0, i_1, \ldots, i_{h-1}$ be integers such that*

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} < p$$

and

$$i_0 + i_1 + \cdots i_{h-1} < \binom{h}{2} + p.$$

Then

$$\hat{B}(i_0, i_1, \ldots, i_{h-1}) \not\equiv 0 \pmod{p}.$$

Proof. This follows immediately from the Theorem.

3 A review of linear algebra

Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear operator. Let $I : V \to V$ be the identity operator. For every nonnegative integer i, we define $T^i : V \to V$ by

$$\begin{aligned} T^0(\mathbf{v}) &= I(\mathbf{v}) = \mathbf{v}, \\ T^i(\mathbf{v}) &= T\left(T^{i-1}(\mathbf{v})\right) \end{aligned}$$

for all $\mathbf{v} \in V$. To every polynomial

$$p(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0 \in F[x]$$

we associate the linear operator $p(T) : V \to V$ defined by

$$p(T) = c_n T^n + c_{n-1} T^{n-1} + \cdots + c_1 T + c_0 I.$$

The set of all polynomials $p(x)$ such that $p(T) = 0$ forms a nonzero, proper ideal J in the polynomial ring $F[x]$. Since every ideal in $F[x]$ is principal, there exists a unique monic polynomial $p_T(x) = p_{T,V}(x) \in J$ such that $p_T(x)$ divides every other polynomial in J. This polynomial is called the *minimal polynomial* of T over the vector space V.

A subspace W of V is called *invariant* with respect to T if $T(W) \subseteq W$, that is, if $T(\mathbf{w}) \in W$ for all $\mathbf{w} \in W$. Then T restricted to the subspace W is a linear operator on W with minimal polynomial $p_{T,W}(x)$. Since $p_{T,V}(T)(\mathbf{w}) = \mathbf{0}$ for all $\mathbf{w} \in W$, it follows that $p_{T,W}(x)$ divides $p_{T,V}(x)$, and so

$$\deg(p_{T,W}) \leq \deg(p_{T,V}), \tag{2}$$

where $\deg(p)$ denotes the degree of the polynomial p.

For $\mathbf{v} \in V$, the *cyclic subspace* with respect to T generated by $\mathbf{v}$ is the smallest subspace of V that contains $\mathbf{v}$ and is invariant under the operator T. We denote this subspace by

$\mathcal{C}_T(\mathbf{v})$. Let $\mathbf{v}_i = T^i(\mathbf{v})$ for $i = 0, 1, 2, \ldots$. Then $\mathcal{C}_T(\mathbf{v})$ is the subspace generated by the vectors

$$\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), T^3(\mathbf{v}), \ldots\} = \{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots\}$$

and

$$\dim(\mathcal{C}_T(\mathbf{v})) = l,$$

where l is the smallest integer such the vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_l$ are linearly dependent. This means that there exist scalars $c_0, c_1, \ldots, c_{l-1}$ in the field F such that

$$\mathbf{v}_l + c_{l-1}\mathbf{v}_{l-1} + \cdots + c_1\mathbf{v}_1 + c_0\mathbf{v}_0 = \mathbf{0}.$$

Let

$$p(x) = x^l + c_{l-1}x^{l-1} + \cdots + c_1 x + c_0.$$

Then

$$\begin{aligned} p(T)(\mathbf{v}_0) &= T^l(\mathbf{v}_0) + c_{l-1}T^{l-1}(\mathbf{v}_0) + \cdots + c_1 T(\mathbf{v}_0) + c_0 I(\mathbf{v}_0) \\ &= \mathbf{v}_l + c_{l-1}\mathbf{v}_{l-1} + \cdots + c_1\mathbf{v}_1 + c_0\mathbf{v}_0 \\ &= \mathbf{0}, \end{aligned}$$

and so

$$p(T)(\mathbf{v}_i) = p(T)(T^i(\mathbf{v}_0)) = T^i\left(p(T)(\mathbf{v}_0)\right) = T^i(\mathbf{0}) = \mathbf{0}$$

for $i = 0, 1, 2, \ldots$. Therefore, $p(T) = 0$ on the cyclic subspace $\mathcal{C}_T(\mathbf{v}) = \mathcal{C}$, and so $p(x)$ is divisible by the minimal polynomial $p_{T,\mathcal{C}}(x)$, and

$$m = \deg\left(p_{T,\mathcal{C}}\right) \leq \deg(p) = l.$$

On the other hand, since

$$p_{T,\mathcal{C}}(\mathbf{v}) = \mathbf{0},$$

it follows that the vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly dependent, and so

$$l \leq m.$$

This implies that $l = m$ and so, by inequality (2),

$$\dim(\mathcal{C}_T(\mathbf{v})) = \deg\left(p_{T,\mathcal{C}}\right) \leq \deg\left(p_{T,V}\right)$$

for all $\mathbf{v} \in V$.

If $T(\mathbf{f}) = a\mathbf{f}$ for some $a \in F$ and some nonzero vector $\mathbf{f} \in V$, then a is called an *eigenvalue* of T and $\mathbf{f}$ is called an *eigenvector* of T with eigenvalue a. The *spectrum* of T, denoted $\sigma(T)$, is the set of all eigenvalues of T. If V has a basis consisting entirely of eigenvectors of T, then T is called a *diagonal operator*.

The following inequality plays a central role in the proof of the Erdős-Heilbronn conjecture.

Lemma 4 *Let T be a diagonal linear operator on a finite-dimensional vector space V, and let $\sigma(T)$ be the spectrum of T. Then*

$$\dim\left(\mathcal{C}_T(\mathbf{v})\right) \leq |\sigma(T)| \tag{3}$$

for every $\mathbf{v} \in V$.

Proof. Let $a \in \sigma(T)$, and let $\mathbf{f}$ be an eigenvector with eigenvalue a. Let W be the one-dimensional subspace generated by $\mathbf{f}$. Then W is invariant with respect to T, and $p_{T,W}(x) = x - a$. It follows that $x - a$ divides $p_{T,V}(x)$, and so

$$\prod_{a \in \sigma(T)} (x - a)$$

divides $p_{T,V}(x)$. Let $\dim(V) = k$. If T is a diagonal linear operator, then V has a basis $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ of eigenvectors, and

$$\prod_{a \in \sigma(T)} (T - a)(\mathbf{f}_i) = \mathbf{0}$$

for $i = 0, 1, \ldots, k-1$. It follows that $\prod_{a \in \sigma(T)}(T - a)(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$, and so

$$p_{T,V}(x) = \prod_{a \in \sigma(T)} (x - a).$$

In particular, the degree of $p_{T,V}(x)$ is equal to the number of distinct eigenvalues of T. It follows that, if T is a diagonal operator on a finite-dimensional vector space V, then

$$\dim\left(\mathcal{C}_T(\mathbf{v})\right) \leq \deg\left(p_{T,V}\right) = |\sigma(T)|$$

for every $\mathbf{v} \in V$. This completes the proof.

Lemma 5 *Let $T : V \to V$ be a linear operator on the vector space V, and let $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ be eigenvectors of T with distinct eigenvalues. Let*

$$\mathbf{v}_0 = \mathbf{f}_0 + \mathbf{f}_1 + \cdots + \mathbf{f}_{k-1},$$

and let $\mathcal{C}_T(\mathbf{v}_0)$ be the cyclic subspace generated by $\mathbf{v}_0$. Then

$$\dim\left(\mathcal{C}_T(\mathbf{v}_0)\right) = k$$

and

$$\{\mathbf{v}_0, T(\mathbf{v}_0), T^2(\mathbf{v}_0), \ldots, T^{k-1}(\mathbf{v}_0)\}$$

is a basis for $\mathcal{C}_T(\mathbf{v}_0)$. If $\dim(V) = k$, then $\mathcal{C}_T(\mathbf{v}_0) = V$.

Proof. We first show that the vectors $\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}$ are linearly independent. If they are linearly dependent, then there is a minimal subset of the vectors $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ that is linearly dependent, say, $\mathbf{f}_0, \ldots, \mathbf{f}_{l-1}$. Moreover, $l \geq 2$ since $\mathbf{f}_i \neq \mathbf{0}$ for $i = 0, 1, \ldots, k-1$. There exist nonzero scalars $c_0, c_1, \ldots, c_{l-1}$ such that $\sum_{i=0}^{l-1} c_i \mathbf{f}_i = \mathbf{0}$. Let $a_i \in \sigma(T)$ be the eigenvalue corresponding to the eigenvector $\mathbf{f}_i$. Then

$$T\left(\sum_{i=0}^{l-1} c_i \mathbf{f}_i\right) = \sum_{i=0}^{l-1} c_i T(\mathbf{f}_i) = \sum_{i=0}^{l-1} c_i a_i \mathbf{f}_i = \mathbf{0}.$$

Since

$$\sum_{i=0}^{l-1} c_i a_{l-1} \mathbf{f}_i = a_{l-1} \sum_{i=0}^{l-1} c_i \mathbf{f}_i = \mathbf{0},$$

it follows that

$$\sum_{i=0}^{l-1} c_i (a_i - a_{l-1}) \mathbf{f}_i = \sum_{i=0}^{l-2} c_i (a_i - a_{l-1}) \mathbf{f}_i = \mathbf{0},$$

which contradicts the minimality of l, since $c_i(a_i - a_{l-1}) \neq 0$ for $i < l-1$. Thus, the vectors $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ are linearly independent, and span a k-dimensional subspace W of V. Moreover, W is an invariant subspace, since it has a basis of eigenvectors of T. Since

$$\mathbf{v}_0 = \mathbf{f}_1 + \cdots + \mathbf{f}_k \in W,$$

it follows that

$$\mathcal{C}_T(\mathbf{v}_0) \subseteq W$$

and so

$$\dim\left(\mathcal{C}_T(\mathbf{v}_0)\right) \leq \dim(W) = k.$$

The vector $T^i(\mathbf{v}_0) \in \mathcal{C}_T(\mathbf{v}_0)$ for every nonnegative integer i. Since

$$T^i(\mathbf{v}_0) = a_0^i \mathbf{f}_0 + a_1^i \mathbf{f}_1 + \cdots + a_{k-1}^i \mathbf{f}_{k-1},$$

the matrix of the set of vectors $\{\mathbf{v}_0, T(\mathbf{v}_0), T^2(\mathbf{v}_0), \ldots, T^{k-1}(\mathbf{v}_0)\}$ with respect to the basis $\{\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}\}$ is

$$\begin{pmatrix} 1 & a_0 & a_0^2 & \cdots & a_0^{k-1} \\ 1 & a_1 & a_1^2 & \cdots & a_1^{k-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{k-1} \\ \vdots & & & & \\ 1 & a_{k-1} & a_{k-1}^2 & \cdots & a_{k-1}^{k-1} \end{pmatrix},$$

and its determinant is the Vandermonde determinant

$$\prod_{0 \leq i < j \leq h} (a_j - a_i) \neq 0.$$

It follows that $\{\mathbf{v}_0, T(\mathbf{v}_0), T^2(\mathbf{v}_0), \ldots, T^{k-1}(\mathbf{v}_0)\}$ is a set of linearly independent vectors, and so

$$\dim\left(\mathcal{C}_T(\mathbf{v}_0)\right) \geq k = \dim(W).$$

Therefore, $\dim\left(\mathcal{C}_T(\mathbf{v}_0)\right) = k$. If $\dim(V) = k$, then $\mathcal{C}_T(\mathbf{v}_0) = V$. This completes the proof.

4 Alternating products

Let $\wedge^h V$ denote the h-th *alternating product* of the vector space V. Then $\wedge^h V$ is a vector space whose elements are linear combinations of expressions of the form

$$\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1},$$

where $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{h-1} \in V$. These *wedge products* have the property that

$$\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} = \mathbf{0}$$

if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$, and

$$\mathbf{v}_{\sigma(0)} \wedge \mathbf{v}_{\sigma(1)} \wedge \cdots \wedge \mathbf{v}_{\sigma(h-1)} = \varepsilon(\sigma)\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1}$$

for all $\sigma \in S_h$.

If $\{\mathbf{e}_0, \ldots, \mathbf{e}_{k-1}\}$ is a basis for V, then a basis for $\wedge^h V$ is the set of all wedge products of the form

$$\mathbf{e}_{i_0} \wedge \mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_{h-1}},$$

where

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq k-1,$$

and

$$\dim(\wedge^h V) = \binom{k}{h}.$$

Every linear operator $T : V \to V$ induces a linear operator

$$DT : \bigwedge^h V \to \bigwedge^h V$$

that acts on wedge products according to the rule

$$DT(\mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}) = \tag{4}$$

$$\sum_{j=0}^{h-1} \mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{j-1}} \wedge \ T(\mathbf{v}_{i_j}) \wedge \mathbf{v}_{i_{j+1}} \cdots \wedge \mathbf{v}_{i_{h-1}}. \tag{5}$$

The operator DT is called the *derivative* of T.

Lemma 6 *Let T be a diagonal linear operator on V, and let $\sigma(T)$ be the spectrum of T. For $h \geq 2$, let $DT : \bigwedge^h V \to \bigwedge^h V$ be the derivative of T. If T has distinct eigenvalues, that is, if $|\sigma(T)| = \dim(V)$, then*

$$\sigma(DT) = h^{\wedge}\sigma(T)$$

and

$$|h^{\wedge}\sigma(T)| \geq \dim\left(\mathcal{C}_{DT}(\mathbf{w})\right) \tag{6}$$

for every $\mathbf{w} \in \wedge^h V$.

Proof. Let $\sigma(T) = \{a_0, a_1, \ldots, a_{k-1}\}$, and let $\{\mathbf{f}_0, \mathbf{f}_1, \cdots, \mathbf{f}_{k-1}\}$ be a basis of eigenvectors of V such that $T(\mathbf{f}_i) = a_i\mathbf{f}_i$ for $i = 0, 1, \ldots, k-1$. Then (4) implies that

$$DT(\mathbf{f}_{i_0} \wedge \cdots \wedge \mathbf{f}_{i_{h-1}}) = (a_{i_0} + \cdots + a_{i_{h-1}})(\mathbf{f}_{i_0} \wedge \cdots \wedge \mathbf{f}_{i_{h-1}}).$$

It follows that DT is a diagonal linear operator on $\wedge^h V$, and its spectrum $\sigma(DT)$ consists of all sums of h distinct eigenvalues of T, that is,

$$\sigma(DT) = h^{\wedge}\sigma(T).$$

Applying inequality (3) to the vector space $\wedge^h V$ and the operator DT, we obtain

$$|h^{\wedge}\sigma(T)| = |\sigma(DT)| \geq \dim\left(\mathcal{C}_{DT}(\mathbf{w})\right)$$

for every $\mathbf{w} \in \wedge^h V$. This completes the proof.

Theorem 2 *Let T be a linear operator on the finite-dimensional vector space V. Let $h \geq 2$, and let $DT : \wedge^h V \to \wedge^h V$ be the derivative of T. For $\mathbf{v}_0 \in V$, define*

$$\mathbf{v}_i = T^i(\mathbf{v}_0) \in V$$

for $i \geq 1$, and let

$$\mathbf{w} = \mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} \in \wedge^h V.$$

Then for every $r \geq 0$

$$\begin{aligned}(DT)^r(\mathbf{w}) &= (DT)^r(\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1}) \\ &= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1})\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}},\end{aligned}$$

where the sum is over all integer lattice points $(i_0, i_1, \ldots, i_{h-1}) \in \mathbf{Z}^h$ such that

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq r + h - 1$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r,$$

and where $\hat{B}(i_0, i_1, \ldots, i_{h-1})$ is the strict h-dimensional ballot number corresponding to the lattice point $(i_0, i_1, \ldots, i_{h-1})$.

Proof. The proof will be by induction on r. Let $r = 0$. Since $\hat{B}(0, 1, 2, \ldots, h-1) = 1$, we have

$$\begin{aligned}(DT)^0(\mathbf{w}) &= \mathbf{w} \\ &= \mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1} \\ &= \hat{B}(0, 1, 2, \ldots, h-1)\mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{h-1}.\end{aligned}$$

Suppose the result holds for some integer $r \geq 0$. Then

$$\begin{aligned}
(DT)^{r+1}(\mathbf{w}) &= DT\left((DT)^r(\mathbf{w})\right) \\
&= DT\left(\sum \hat{B}(i_0, i_1, \ldots, i_{h-1})\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}\right) \\
&= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1}) DT\left(\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}\right) \\
&= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1}) \sum_{j=0}^{h-1} \left(\mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{j-1}} \wedge T(\mathbf{v}_{i_j}) \wedge \mathbf{v}_{i_{j+1}} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}\right) \\
&= \sum \hat{B}(i_0, i_1, \ldots, i_{h-1}) \sum_{j=0}^{h-1} \left(\mathbf{v}_{i_0} \wedge \cdots \wedge \mathbf{v}_{i_{j-1}} \wedge \mathbf{v}_{i_j+1} \wedge \mathbf{v}_{i_{j+1}} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}\right) \\
&= \sum C(i_0, i_1, \ldots, i_{h-1})\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}},
\end{aligned}$$

where the last sum is over all integer lattice points $(i_0, i_1, \ldots, i_{h-1}) \in \mathbf{Z}^n$ such that

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq r + h$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r + 1,$$

and the integer $C(i_0, i_1, \ldots, i_{h-1})$ satisfies the difference equation

$$C(i_0, i_1, \ldots, i_{h-1}) = \sum_{j=0}^{h-1} \hat{B}(i_0, \ldots, i_{j-1}, i_j - 1, i_{j+1}, \ldots, i_{h-1}).$$

This difference equation determines the strict h-dimensional ballot numbers, and so

$$C(i_0, i_1, \ldots, i_{h-1}) = \hat{B}(i_0, i_1, \ldots, i_{h-1}).$$

Therefore, the result holds in the case $r + 1$. This completes the induction.

5 Erdős-Heilbronn, concluded

Theorem 3 *Let p be a prime number, and let $A \subseteq \mathbf{Z}/p\mathbf{Z}$, where $|A| = k$. Let $2 \leq h \leq k$. Then*

$$|h^{\wedge}A| \geq \min(p, hk - h^2 + 1).$$

Proof. Let $A = \{a_0, a_1, \ldots, a_{k-1}\}$. Let V be a vector space of dimension k over the field $\mathbf{Z}/p\mathbf{Z}$, and let $\{\mathbf{f}_0, \mathbf{f}_1, \ldots, \mathbf{f}_{k-1}\}$ be a basis for V. We define the diagonal linear operator $T : V \to V$ by

$$T(\mathbf{f}_i) = a_i\mathbf{f}_i$$

for $i = 0, 1, \ldots, k-1$. The spectrum of T is

$$\sigma(T) = A.$$

Let

$$\mathbf{v}_0 = \mathbf{f}_0 + \mathbf{f}_1 + \cdots + \mathbf{f}_{k-1},$$

and define

$$\mathbf{v}_{i+1} = T(\mathbf{v}_i) = T^i(\mathbf{v}_0)$$

for $i \geq 0$. By Lemma 5, the cyclic subspace generated by $\mathbf{v}_0$ is V, and the set of vectors $\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$ is a basis for V. The alternating product $\bigwedge^h V$ is a vector space with a basis consisting of the $\binom{k}{h}$ wedge products of the form

$$\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}},$$

where

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq k-1.$$

Let

$$\mathbf{w} = \mathbf{v}_0 \wedge \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_{k-1} \in \bigwedge^h V.$$

By inequality (6),

$$|h^{\wedge}A| = |\sigma(DT)| \geq \dim \mathcal{C}_{DT}(\mathbf{w}).$$

Therefore, it suffices to prove that

$$\dim \mathcal{C}_{DT}(\mathbf{w}) \geq \min(p, hk - h^2 + 1).$$

This is equivalent to proving that the vectors

$$\mathbf{w}, (DT)(\mathbf{w}), (DT)^2(\mathbf{w}), \ldots, (DT)^n(\mathbf{w})$$

are linearly independent in the alternating product $\bigwedge^h V$, where

$$n = \min(p, hk - h^2 + 1) - 1 = \min(p - 1, hk - h^2).$$

Let $0 \leq r \leq n$. By Theorem 2, the vector $(DT)^r(\mathbf{w})$ is a linear combination of the vectors

$$\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$$

such that

$$0 \leq i_0 < i_1 < \cdots < i_{h-1} \leq r + h - 1$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r. \tag{7}$$

Let I be the interval of integers $[0, k-1]$. Since

$$h^{\wedge}I = \left[\binom{h}{2}, hk - \binom{h+1}{2}\right] = \binom{h}{2} + [0, hk - h^2],$$

it follows that there is at least one basis vector $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$ in the expansion of $(DT)^r(\mathbf{w})$ such that

$$0 \le i_0 < i_1 < \cdots < i_{h-1} \le k-1 < k \le p$$

and

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r \le \binom{h}{2} + n <\le \binom{h}{2} + p.$$

By Theorem 2, the coefficient of this basis vector is the strict h-dimensional ballot number $\hat{B}(i_0, i_1, \ldots, i_{h-1})$, and

$$\hat{B}(i_0, i_1, \ldots, i_{h-1}) \not\equiv 0 \pmod{p}$$

by Corollary 1.

Suppose that the vector $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$ satisfies $i_l \ge k$ for some l. Since every vector $\mathbf{v}_l \in V$ with $l \ge k$ is a linear combination of $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$, it follows that $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$ is a linear combination of basis vectors of the form $\mathbf{v}_{j_0} \wedge \mathbf{v}_{j_1} \wedge \cdots \wedge \mathbf{v}_{j_{h-1}}$, where

$$0 \le j_0 < j_1 < \cdots < j_{h-1} \le k-1$$

and

$$j_0 + j_1 + \cdots + j_{h-1} < \binom{h}{2} + r.$$

It follows that $(DT)^r(\mathbf{w})$ is a linear combination of basis vectors $\mathbf{v}_{i_0} \wedge \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_{h-1}}$ such that either

$$i_0 + i_1 + \cdots + i_{h-1} < \binom{h}{2} + r$$

or

$$i_0 + i_1 + \cdots + i_{h-1} = \binom{h}{2} + r,$$

and in the latter case the basis vector appears with a coefficient that is nonzero modulo p. This implies that the vectors $\mathbf{w}, (DT)(\mathbf{w}), \ldots, (DT)^n(\mathbf{w})$ are linearly independent in the cyclic subspace $\mathcal{C}_{DT}(\mathbf{w})$, and the proof of the Erdős-Heilbronn conjecture is complete.

6 Remarks

This proof only requires that A be a subset of a field, and does not require that the field be $\mathbf{Z}/p\mathbf{Z}$. Let F be an arbitrary field. Let p be the characteristic of F if the characteristic is positive, and let $p = \infty$ if the characteristic is zero. Then we have, in fact, proved that if $A \subseteq F$ and $|A| = k \le p$, then $|h^{\wedge}A| \ge \min(p, hk - h^2 + 1)$ for all $h \ge 1$.

The Cauchy-Davenport theorem was proved by Cauchy [2] in 1813. Davenport [4] rediscovered the result in 1935. I. Chowla [3] immediately extended the Cauchy-Davenport theorem to composite moduli. Other generalizations have been obtained by Pillai [13], Shatrovskii [19], Pollard [14, 15], Brakemaier [1], and Hamidoune [11]. Davenport [5] discovered in 1947 that Cauchy had proved the Cauchy-Davenport theorem first.

The Erdős-Heilbronn conjecture originated in the 1960's. Partial results on the Erdős-Heilbronn conjecture were obtained by Rickert [17], Mansfield [12], Rödseth [18], Pyber [16], and Freiman, Low, and Pitman [10]. Dias da Silva and Hamidoune [7] proved the conjecture by using results from representation theory and linear algebra. This algebraic technique had previously been applied to additive number theory by Dias da Silva and Hamidoune [6] and Spigler [20].

The derivation of the formula for the strict ballot number $\hat{B}(b_0, \ldots, b_{n-1})$ follows a paper of Zeilberger [21].

Using combinatorial ideas from the proof of Dias da Silva and Hamidoune, Nathanson and Ruzsa found a different proof of the Erdős-Heilbronn conjecture that uses only the simplest properties of polynomials.

References

[1] W. Brakemeier. Eine Anzahlformel von Zahlen modulo n. *Monat. Math.*, 85:277–282, 1978.

[2] A. L. Cauchy. Recherches sur les nombres. *J. École polytech.*, 9:99–116, 1813.

[3] I. Chowla. A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring's problem. *Proc. Indian Acad. Sci., Section A*, 1:242–243, 1935.

[4] H. Davenport. On the addition of residue classes. *J. London Math. Soc.*, 10:30–32, 1935.

[5] H. Davenport. A historical note. *J. London Math. Soc.*, 22:100–101, 1947.

[6] J. A. Dias da Silva and Y. O. Hamidoune. A note on the minimal polynomial of the Kronecker sum of two linear operators. *Linear Algebra and its Applications*, 141:283–287, 1990.

[7] J. A. Dias da Silva and Y. O. Hamidoune. Cyclic spaces for Grassmann derivatives and additive theory. *Bull. London Math. Soc.*, 26:to appear, 1994.

[8] P. Erdős and R. L. Graham. *Old and New Problems and Results in Combinatorial Number Theory*. L'Enseignement Mathématique, Geneva, 1980.

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

[10] G. A. Freiman, L. Low, and J. Pitman. On the Erdős-Heilbronn conjecture. preprint, 1993.

[11] Y. O. Hamidoune. A generalization of an addition theorem of Shatrowsky. *Europ. J. Combin.*, 13:249–255, 1992.

[12] R. Mansfield. How many slopes in a polygon? *Israel J. Math.*, 39:265–272, 1981.

[13] S. S. Pillai. Generalization of a theorem of Davenport on the addition of residue classes. *Proc. Indian Acad. Sci. (Series A)*, 6:179–180, 1938.

[14] J. M. Pollard. A generalization of a theorem of Cauchy and Davenport. *J. London Math. Soc.*, 8:460–462, 1974.

[15] J. M. Pollard. Additive properties of residue classes. *J. London Math. Soc.*, 11:147–152, 1975.

[16] L. Pyber. On the Erdős-Heilbronn conjecture. personal communication.

[17] U.-W. Rickert. *"Uber eine Vermutung in der additiven Zahlentheorie.* PhD thesis, Tech. Univ. Braunschweig, 1976.

[18] Ö. J. Rödseth. Sums of distinct residues mod p. *Acta Arith.*, 1994. to appear.

[19] L. Shatrovskii. A new generalization of Davenport's-Pillai's theorem on the addition of residue classes. *Doklady Akad. Nauk CCCR*, 45:315–317, 1944.

[20] R. Spigler. An application of group theory to matrices and to ordinary differential equations. *Linear Algebra and its Applications*, 44:143–151, 1982.

[21] D. Zeilberger. André's reflection proof generalized to the many-candidate ballot problem. *Discrete Math.*, 44:325–326, 1983.