Order shattering and Wilson's theorem

Katalin Friedl Lajos Rónyai

Computer and Automation Institute

Hungarian Academy of Sciences

and

Budapest University of Technology and Economy

October 16, 2002

Abstract

The notion of order shattering was introduced in [1]. Here we pursue further the algebraic interpretation that was established there. With this tool we give a new proof and a generalization for Wilson's theorem on the diagonal form for the incidence matrices of t-subsets vs. k-subsets [5]. This allows a generalization of the corresponding rank formula modulo p, where p is an arbitrary prime.

Keywords. Shattered set, inclusion matrix, Wilson's rank formula.

1 Introduction

Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. For a set S let $\binom{S}{k}$ denote the family of all k element subsets of S, and $\binom{S}{\leq k} = \binom{S}{0} \cup \cdots \cup \binom{S}{k}$ the subsets of size at most k.

For families $\mathcal{F}, \mathcal{G} \subseteq 2^{[n]}$ the *inclusion matrix* $I(\mathcal{F}, \mathcal{G})$ is a (0,1) matrix of size $|\mathcal{F}| \times |\mathcal{G}|$ whose rows and columns are indexed by the elements of $\mathcal{F}$ and $\mathcal{G}$, respectively. The entry at position (F, G) is 1 if $G \subseteq F$ and 0 otherwise $(F \in \mathcal{F}, G \in \mathcal{G})$.

[0]Research supported in part by OTKA and NWO-OTKA grants, and the EU-COE Grant of MTA SZTAKI.

Wilson's theorem [5, Theorem 2] states that if $d \le t \le n-d$ then the matrix $\mathcal{I} := I(\binom{[n]}{t}, \binom{[n]}{d})$ has a simple diagonal form. In fact it is row-column equivalent over $\mathbb{Z}$ to a diagonal matrix with diagonal entries $\binom{t-i}{d-i}$ with multiplicity $\binom{n}{i} - \binom{n}{i-1}$ for $0 \le i \le d$. As a corollary, the rank of $\mathcal{I}$ over finite prime fields $\mathbb{F}_p$ was established. A new, simpler proof for the rank formula was found by Frankl [4], and his method was extended by Bier [3] to obtain the theorem on the diagonal form.

Building on the notion of order shattering and its connection with polynomial functions, we give a new interpretation of the Frankl–Bier approach. This provides a natural explicit basis of the column space of $\mathcal{I}$ and allows a generalization to $I(\binom{[n]}{t}, \mathcal{H})$ where $\mathcal{H} = \binom{[n]}{d_1} \cup \cdots \cup \binom{[n]}{d_r}$.

Following [1] we say that the set $S = \{s_1, s_2, \ldots, s_d\} \subseteq [n]$ is *order shattered* by a given family $\mathcal{F} \subseteq 2^{[n]}$ if the following holds: in the case $S = \emptyset$ the family $\mathcal{F}$ has to contain a set; when $|S| > 0$ and $s_1 < s_2 < \cdots < s_d$ then there are 2^d sets in $\mathcal{F}$ that can be divided into two families $\mathcal{F}_0$ and $\mathcal{F}_1$ such that $s_d \notin F$ for all $F \in \mathcal{F}_0$, $s_d \in F$ for all $F \in \mathcal{F}_1$, and both $\mathcal{F}_0, \mathcal{F}_1$ order shatter the set $S \setminus \{s_d\}$, furthermore $T \cap F_0 = T \cap F_1$ holds for $T = \{s_d+1, s_d+2, \ldots, n\}$ and all $F_0 \in \mathcal{F}_0, F_1 \in \mathcal{F}_1$.

Let

$$\mathrm{osh}\,(\mathcal{F}) = \{S \subseteq [n] : \mathcal{F} \text{ order shatters } S\}.$$

Notice that $\mathrm{osh}\,(\mathcal{F})$ is a downset: if $S \in \mathrm{osh}\,(\mathcal{F})$ and $T \subseteq S$, then $T \in \mathrm{osh}\,(\mathcal{F})$ as well.

Examples. 1. If $\mathcal{F} = 2^S$ then clearly $\mathrm{osh}\,(\mathcal{F}) = 2^S$ for $S \subseteq [n]$. More generally, $\mathrm{osh}\left(\binom{[n]}{\le d}\right) = \binom{[n]}{\le d}$.

2. The family $\mathcal{G}$ consisting of all even subsets $F \subseteq [n]$ does not order shatter $\{1\}$. Indeed, there are no $F_0, F_1 \in \mathcal{G}$ such that $1 \in F_1$, $1 \notin F_0$ and $F_1 \cap \{2, \ldots, n\} = F_0$. On the other hand this $\mathcal{G}$ order shatters all subsets of $\{2, \ldots, n\}$.

It was proved in [1] that $|\mathrm{osh}\,(\mathcal{F})\,| = |\mathcal{F}|$ holds for arbitrary family $\mathcal{F} \subseteq 2^{[n]}$.

We denote the field of rational numbers, the ring of integers, and the p-element field (p is a prime) by $\mathbb{Q}, \mathbb{Z}, \mathbb{F}_p$, resp. Let R be a commutative ring with identity. In the sequel R will be one of $\mathbb{Q}$, $\mathbb{Z}$, or $\mathbb{F}_p$. Let $R[x_1, \ldots, x_n]$ denote the ring of polynomials in variables $x_1, \ldots, x_n$ over R. For a subset $F \subseteq [n]$ we write $x_F = \prod_{j \in F} x_j$, in particular, $x_\emptyset = 1$.

Let $v_F \in \{0,1\}^n$ denote the *characteristic vector* of a set $F \subseteq [n]$. For a family of subsets $\mathcal{F} \subseteq 2^{[n]}$, let $V(\mathcal{F}) = \{v_F : F \in \mathcal{F}\} \subseteq \{0,1\}^n \subseteq R^n$.

For a family $\mathcal{F} \subseteq 2^{[n]}$ and a ring R let $\mathsf{func}(\mathcal{F}, R)$ denote the set of all functions from $V(\mathcal{F})$ to R. It is clear that $\mathsf{func}(\mathcal{F}, R)$ is a free module of rank $|\mathcal{F}|$ over R (addition and multiplication by a scalar is pointwise).

In [1] it was proved that if $\mathcal{F} \subseteq 2^{[n]}$ is a nonempty family and K is a field then $S = \{x_F : F \in \mathrm{osh}\,(\mathcal{F})\}$ is a K-basis of $\mathsf{func}(\mathcal{F}, K)$. In fact it was proved there that S is the set of lexicographic standard monomials of the ideal of all polynomials $f \in K[x_1, \ldots, x_n]$ which vanish on $V(\mathcal{F}) \subseteq K^n$. We use this algebraic interpretation of $\mathrm{osh}\,()$ to prove and generalize Wilson's theorem.

To simplify notation we write $\mathrm{osh}\binom{[n]}{d}$ and $V\binom{[n]}{d}$, etc. instead of the more precise variants with double parentheses.

2 Uniform families

Let $\mathcal{F} \subseteq 2^{[n]}$ and $H \subseteq [n]$. Assume that $|F \cup H| \neq [n]$ holds for each $F \in \mathcal{F}$. Denote by $F^+(H)$ the set $F \cup \{j\}$ where j is the largest element of $[n] \setminus (F \cup H)$. Let $\mathcal{F}^+(H)$ denote the family of these sets, i.e. $\mathcal{F}^+(H) = \{F^+(H) : F \in \mathcal{F}\}$.

Lemma 2.1 (Going up) *Suppose that a family $\mathcal{F} \subseteq 2^{[n]}$ order shatters the set $S = \{s_1, s_2, \ldots, s_d\} \subseteq [n]$. Let H be a superset of S, $S \subseteq H \subseteq [n]$. Assume that $F \cup H \neq [n]$ holds for each $F \in \mathcal{F}$. Then $\mathcal{F}^+(H)$ also order shatters S.*

Proof. We proceed by induction on d.

Assume first that $d = 0$. Then $\mathcal{F}$ is not empty, let $F \in \mathcal{F}$ for some $F \subset [n]$. Then $F^+(H)$ obviously exists and order shatters $S = \emptyset$.

Let $d > 0$ and $s_1 < s_2 < \ldots < s_d$. We may assume without loss of generality that $\mathcal{F}$ order shatters S and $|\mathcal{F}| = 2^d$. Let $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$ be the partition of $\mathcal{F}$, where $s_d \notin F$ for $F \in \mathcal{F}_0$ and $s_d \in F$ for $F \in \mathcal{F}_1$. The families $\mathcal{F}_0$ and $\mathcal{F}_1$ both order shatter the set $S' = \{s_1, \ldots, s_{d-1}\}$. Clearly $F \cup H \neq [n]$ whenever $F \in \mathcal{F}_i$ and $S' \subseteq H \subseteq [n]$. By the induction hypothesis the families $\mathcal{F}_0^+(H)$ and $\mathcal{F}_1^+(H)$ both order shatter S'. Notice that $\mathcal{F}^+(H) = \mathcal{F}_0^+(H) \cup \mathcal{F}_1^+(H)$.

To finish the proof it is sufficient to verify that

$$F_0^+(H) \cap [s_d + 1, n] = F_1^+(H) \cap [s_d + 1, n] \tag{1}$$

holds whenever $F_i \in \mathcal{F}_i$. We know that $F_0 \cap [s_d + 1, n] = F_1 \cap [s_d + 1, n]$ whenever $F_i \in \mathcal{F}_i$ because $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$ order shatters S. Denote this common intersection by D, so $D = F_0 \cap [s_d+1, n]$. Now if $[s_d+1, n] \subseteq D \cup H$ then the intersections on both sides of (1) will be D. On the other hand, if $[s_d + 1, n] \not\subseteq D \cup H$ and j is the largest element of $[s_d + 1, n] \setminus (D \cup H)$ then the intersection on both sides of (1) is $D \cup \{j\}$. So we have seen that (1) holds and consequently $\mathcal{F}^+(H)$ order shatters S. □

The above method can be used also to decrease the sets of a shattering family $\mathcal{F}$. For this notice that when the family $\mathcal{F}$ order shatters the set S then the family $co\mathcal{F} = \{T \subseteq [n] : [n] \setminus T \in \mathcal{F}\}$ also order shatters S.

Corollary 2.2 *If $S = \{s_1, \ldots, s_d\} \subseteq [n]$ is order shattered by some k-uniform family $\mathcal{F}$ then for each $d \le t \le n - d$ there exists a t-uniform family that order shatters S.*

Proof. Clearly $d \le k$ because there exists an $F \in \mathcal{F}$ with the property $S \subseteq F$ and $k \le n - d$ also holds because there exists an $F' \in \mathcal{F}$ such that $F' \cap S = \emptyset$. (This shows that if S is uniformly order shattered then $d \le n/2$ and $d \le n - d$ as it is required.)

We may assume that $k \le n/2$ because if $S \in \operatorname{osh}(\mathcal{F})$ for a k-uniform $\mathcal{F}$ then the $(n - k)$-uniform $co\mathcal{F}$ also order shatters S.

For any t-uniform shattering family we can apply the Going up lemma with the choice $H = S$ if $t + d < n$. Therefore there is an order shattering t-uniform family for every t that satisfies $k \le t \le n - d$.

Since for an order shattering t-uniform family $\mathcal{F}$ the $n - t$-uniform $co\mathcal{F}$ also order shatters S, there are order shattering t-uniform families for every $d \le t \le n - k$.

These two intervals cover the interval $[d, n - d]$ (recall that $k \le n/2$), so the proof is finished. □

Proposition 2.3 *Let $d \le t \le n - d$. Then*

$$\operatorname{osh}\binom{[n]}{t} \bigcap \binom{[n]}{\le d} = \operatorname{osh}\binom{[n]}{d} \tag{2}$$

Proof. Let $\mathcal{F} = \mathrm{osh}\binom{[n]}{d}$. Obviously $\mathcal{F} \subseteq \binom{[n]}{<d}$ and if $F \in \mathcal{F}$ then $|F| \leq t \leq n - |F|$ holds. By corollary 2.2 F is order shattered by the t-sets as well, hence F belongs to the left hand side of (2).

Conversely, assume that $G \subseteq [n]$ is order shattered by t-sets and $|G| \leq d$. Then $|G| \leq d \leq t \leq n - d \leq n - |G|$, hence (by Corollary 2.2) $G \in \mathrm{osh}\binom{[n]}{d}$. $\square$

From this we can determine the number of i-sets order shattered by $\binom{[n]}{t}$.

Corollary 2.4 *Let $0 \leq i \leq t \leq n - i$. Then the number of i-element subsets of $[n]$ order shattered by $\binom{[n]}{t}$ is $\binom{n}{i} - \binom{n}{i-1}$.*

(A binomial coefficient with a negative integer at the bottom is understood as 0.)

Proof. By the preceding proposition

$$\binom{[n]}{i} \bigcap \mathrm{osh}\binom{[n]}{t} = \mathrm{osh}\binom{[n]}{i} \setminus \mathrm{osh}\binom{[n]}{i-1}$$

and since $\left|\mathrm{osh}\binom{[n]}{i}\right| = \binom{n}{i}$, the proof is complete. $\square$

Remark. In [1] the same formula is obtained with appeal to ballot sequences. It was also shown there that $S \subseteq [n]$, $S = \{s_1 < s_2 < \cdots < s_i\}$ is in $\mathrm{osh}\binom{n}{t}$ if and only if $s_j \geq 2j$ for $j = 1, 2, \ldots, i$. This means that $S \in \mathrm{osh}\binom{[n]}{t}$ if and only if the Frankl-rank of S is $|S|$ (see [3, 4]).

3 Order shattered sets and polynomials

Recall that for a family $\mathcal{F} \subseteq 2^{[n]}$ and a ring R the set of all functions from $V(\mathcal{F})$ to R is denoted by $\mathsf{func}(\mathcal{F}, R)$, where $V(\mathcal{F}) = \{v_F : F \in \mathcal{F}\} \subseteq \{0,1\}^n \subseteq R^n$ and v_F denotes the characteristic vector of the set $F \subseteq [n]$.

Proposition 3.1 *The monomials $\{x_F : F \in \mathcal{F}\}$, when viewed as functions on $V(\mathcal{F})$, form a basis over the field K of the space of functions $\mathsf{func}(\mathcal{F}, K)$. (Here $\{0,1\}^n$ is considered as a subset of K^n.)*

Proof. As the space of functions $\mathsf{func}(\mathcal{F}, K)$ has dimension $|\mathcal{F}|$ over K, it suffices to verify that the x_F for $F \in \mathcal{F}$ are linearly independent functions on $V(\mathcal{F})$. Suppose for contradiction that

$$\sum_{F \in \mathcal{F}} \alpha_F x_F = 0 \tag{3}$$

is a nontrivial linear dependence with $\alpha_F \in K$. Let $H \in \mathcal{F}$ be a minimal subset with respect to $\subseteq$ with $\alpha_H \neq 0$. Upon substitution of v_H into (3) we obtain that $\alpha_H = 0$ which is a contradiction. $\square$

Proposition 3.2 *Let $\mathcal{F}, \mathcal{G} \subseteq 2^{[n]}$. Suppose that the set of monomials $S = \{x_G : G \in \mathcal{G}\}$ is a basis of $\mathsf{func}(\mathcal{F}, K)$ for every field K. Then S is a basis of the lattice $\mathsf{func}(\mathcal{F}, \mathbb{Z})$.*

Proof. The functions $\{x_G : G \in \mathcal{G}\}$ are independent over $\mathbb{Z}$, because they are independent over $\mathbb{Q}$. It suffices to show that they span the lattice of all integer-valued functions on $V(\mathcal{F})$. Consider the characteristic functions f_H for $H \in \mathcal{F}$ for which $f_H(v_H) = 1$ and $f_H(v_{H'}) = 0$ for every $H' \neq H$, $H' \in \mathcal{F}$. Clearly these functions span the lattice and it follows from the assumption that they can be expressed as a linear combination of the x_G,

$$f_H = \sum_{G \in \mathcal{G}} \alpha_G x_G$$

with $\alpha_G \in \mathbb{Q}$. We show that here actually $\alpha_G \in \mathbb{Z}$ as well. For otherwise we obtain a relation

$$m \cdot f_H = \sum_{G \in \mathcal{G}} m_G x_G \tag{4}$$

with $m, m_G \in \mathbb{Z}$, where $m > 1$, and no prime divisor of m divides all the m_G. Now take a prime factor p of m and interpret the relation (4) modulo p. It gives a nontrivial linear dependence over $\mathbb{F}_p$ among the monomials x_G as elements of $\mathsf{func}(\mathcal{F}, \mathbb{F}_p)$, which is impossible by our assumption. $\square$

This result, with Proposition 3.1 implies the following:

Corollary 3.3 *Let* $\mathcal{F} \subseteq 2^{[n]}$*. The monomials* $\{x_F : F \in \mathcal{F}\}$ *form a* $\mathbb{Z}$*-basis of the lattice* $\mathsf{func}(\mathcal{F}, \mathbb{Z})$*.* $\square$

In Theorem 4.3 of [1] it was proved that the family of monomials $S = \{x_G : G \in \operatorname{osh}(\mathcal{F})\}$ is a basis of $\mathsf{func}(\mathcal{F}, K)$ for every field K. In fact it was proved there that S is the set of lexicographic standard monomials of the ideal of all polynomials $f \in K[x_1, \ldots, x_n]$ which vanish on $V(\mathcal{F}) \subseteq K^n$. This result together with Proposition 3.2 immediately gives the following:

Corollary 3.4 *Let* $\mathcal{F} \subseteq 2^{[n]}$*. Then* $\{x_G : G \in \operatorname{osh}(\mathcal{F})\}$ *is a basis of the lattice* $\mathsf{func}(\mathcal{F}, \mathbb{Z})$ $\square$

4 Inclusion matrices

Recall that $I(\mathcal{F}, \mathcal{G})$ denotes the inclusion matrix of the families $\mathcal{F}, \mathcal{G} \subseteq 2^{[n]}$.

It is known (see [2, Section 7.3]) that if $d \le t \le n-d$ are natural numbers then

$$\operatorname{rank}_{\mathbb{Q}} I\left(\binom{[n]}{t}, \binom{[n]}{\le d}\right) = \binom{n}{d}. \tag{5}$$

This implies that the monomials $\{x_F : |F| \le d\}$ span a subspace of dimension $\binom{n}{d}$ in the space $\mathsf{func}(\binom{[n]}{t}, \mathbb{Q})$.

Let R be a commutative ring. For $d \le t \le n-d$ let $L_{d,t,R}$ denote the R-module of functions from $V\binom{[n]}{t}$ to R spanned by the monomials $\{x_F : |F| \le d\}$. By (5) we have

$$\dim_{\mathbb{Q}} L_{d,t,\mathbb{Q}} = \binom{n}{d}$$

and hence

$$\operatorname{rank}_{\mathbb{Z}} L_{d,t,\mathbb{Z}} = \binom{n}{d}.$$

The following is a refinement of Wilson's formula [5, Lemma 1] for the value of $\dim_K L_{d,t,K}$. It can also be interpreted as a polynomial version of what Bier calls Frankl's Spanning Lemma (cf. [4]).

Theorem 4.1 *Assume that $d \leq t \leq n-d$ and put $\mathcal{G} = \operatorname{osh}\binom{[n]}{d}$. Then $\{x_G : G \in \mathcal{G}\}$ is a $\mathbb{Z}$-basis of $L_{d,t,\mathbb{Z}}$, hence $\dim_K L_{d,t,K} = \binom{n}{d}$ for every field K.*

Proof. Let L' be the sublattice of $L = L_{d,t,\mathbb{Z}}$ spanned by the monomials $\{x_G : G \in \mathcal{G}\}$. By Proposition 2.3 we have $\operatorname{osh}\binom{[n]}{d} \subseteq \operatorname{osh}\binom{[n]}{t}$. This, together with Corollary 3.4 applied to $\mathcal{F} = \binom{[n]}{t}$ implies that the set of functions $\{x_G : G \in \mathcal{G}\} \subset L'$ is independent over $\mathbb{Z}$. We have $\operatorname{rank}_{\mathbb{Z}} L' = |\mathcal{G}| = \binom{n}{d} = \operatorname{rank}_{\mathbb{Z}} L$, hence L/L' is a finite group. Suppose for contradiction that $L \neq L'$. Then there is an element $w \in L \setminus L'$ and a prime p such that $pw \in L'$. We have then

$$p \cdot w = \sum_{G \in \mathcal{G}} \alpha_G \, x_G \tag{6}$$

with $\alpha_G \in \mathbb{Z}$ and not all divisible by p. Taking (6) modulo p, we infer that x_G, $G \in \mathcal{G}$ are linearly dependent functions from $V\binom{[n]}{t}$ to $\mathbb{F}_p$, a contradiction. $\square$

Let $d \leq t \leq n-d$ and let $M_{d,t}$ denote the lattice of all functions from $V\binom{[n]}{t}$ to $\mathbb{Z}$ spanned by the monomials x_F, $|F| = d$. We intend to establish a nice basis of $M_{d,t}$.

For $G \in \operatorname{osh}\binom{[n]}{d}$ we define the expression

$$y_G := \sum_{\substack{H \supseteq G \\ |H| = d}} x_H .$$

When viewed as functions on $V\binom{[n]}{t}$, the polynomials y_G belong to $M_{d,t}$.

Theorem 4.2 *Suppose that $d \leq t \leq n-d$. The set $B := \{y_G : G \in \operatorname{osh}\binom{[n]}{d}\}$ is a $\mathbb{Z}$-basis of $M_{d,t}$. Moreover $y_G = \binom{t-|G|}{d-|G|} \cdot x_G$ as functions on $V\binom{[n]}{t}$.*

Proof. First we prove the statement in the case $t = d$. Let $\mathsf{func} = \mathsf{func}(\binom{[n]}{d}, \mathbb{Z})$. By Corollary 3.3 we have $M_{d,d} = \mathsf{func}$ and hence $L_{d,d,\mathbb{Z}} = \mathsf{func}$. Moreover, by Corollary 3.4 $\{x_G : G \in \operatorname{osh}\binom{[n]}{d}\}$ is a basis of this lattice.

Let v_F be the incidence vector of a d-subset of $[n]$. It is immediate that for any $G \in \operatorname{osh}\binom{[n]}{d}$ we have

$$x_G(v_F) = y_G(v_F). \tag{7}$$

Indeed, the value on both sides is 1 iff $G \subseteq F$ and 0 otherwise. It follows that the additive map $\alpha : \mathsf{func} \to \mathsf{func}$ defined by $\alpha(x_G) = y_G$ is actually the identity map of func. We infer that B is indeed a basis of func.

In fact, we can say a little bit more. Let N_d denote the sublattice of $\mathbb{Z}[x_1, \ldots, x_n]$ spanned by the monomials x_F with $F \subseteq [n]$, $|F| = d$. Corollary 3.3 implies that the substitution map from N_d to $M_{d,d} = \mathsf{func}$ is an isomorphism. This, together with the fact that B is a basis of func implies that *the polynomials y_G, $G \in \mathrm{osh}\binom{[n]}{d}$, viewed now as formal linear combinations, form a basis of N_d.*

Turning now to the proof of the general case (i.e. $d < t \leq n - d$), it is immediate that

$$y_G(v_F) = \binom{t - |G|}{d - |G|} \cdot x_G(v_F) \tag{8}$$

for any t-set $F \subseteq [n]$. This with the $\mathbb{Z}$-independence of the functions x_G implies the $\mathbb{Z}$-independence of the functions in B as well. By the emphasized statement the polynomials y_G span the lattice N_d, hence as functions they span $M_{d,t}$. We obtained that B is a $\mathbb{Z}$-basis of $M_{d,t}$ and this finishes the proof. $\square$

Remark. This theorem gives an explicit "diagonal form" in the sense of Wilson's theorem [5, Theorem 2.]: we have that $\{x_G : G \in \mathrm{osh}\binom{[n]}{d}\}$ is a subset of a basis of $\mathsf{func}(\binom{[n]}{t}, \mathbb{Z})$, and the integer multiples $\binom{t-|G|}{d-|G|} \cdot x_G$ form a basis $\{y_G\}$ of the lattice of those elements from $\mathsf{func}(\binom{[n]}{t}, \mathbb{Z})$, which can be represented as integer linear combinations of x_H, $|H| = d$.

Corollary 4.3 (Wilson's $\bmod p$ **rank formula)** *Let p be a prime and suppose that $d \leq t \leq n - d$. A basis of the space $P_{d,t}$ of $\mathbb{F}_p$-valued functions on $V\binom{[n]}{t}$, which are $\mathbb{F}_p$-linear combinations of monomials x_H, $|H| = d$ is*

$$D := \left\{ y_G : G \in \mathrm{osh}\binom{[n]}{d}, p \not| \binom{t - |G|}{d - |G|} \right\}.$$

In particular,

$$\dim_{\mathbb{F}_p} P_{d,t} = |D| = \sum_{\substack{0 \leq i \leq d \\ p \not| \binom{t-i}{d-i}}} \binom{n}{i} - \binom{n}{i-1}.$$

Proof. The reduction modulo p map which sends a $\mathbb{Z}$-valued function to a $\mathbb{F}_p$-valued function is obviously surjective, hence by Theorem 4.2 the $\mathbb{F}_p$-linear combinations of y_G, $G \in \mathrm{osh}\binom{[n]}{d}$ span the space of functions $P_{d,t}$. In the light of relations (8) we see that y_G is the 0 function if $p|\binom{t-|G|}{d-|G|}$, and a nonzero scalar multiple of x_G otherwise. Thus, every function from $P_{d,t}$ is a linear combination of some elements from D. Moreover, they are independent over $\mathbb{F}_p$ as the corresponding x_G are ([1, Theorem 4.3]). To finish the proof, it suffices to note that for $0 \le i \le d$ there are exactly $\binom{n}{i} - \binom{n}{i-1}$ i-sets in $\mathrm{osh}\binom{[n]}{d}$ by Corollary 2.4. The proof is complete. □

5 A generalization

An interesting feature of Theorem 4.2 is that a diagonal form for $M_{d,t}$ exists with respect to a basis of $\mathsf{func}(\binom{[n]}{t}, \mathbb{Z})$ that is independent of d, namely $\{x_G : G \in \mathrm{osh}\binom{[n]}{t}\}$. (Of course the functions x_G for $|G| > d$ contributed 0 to the diagonal form.) This coincidence allows an immediate generalization of Theorem 4.2 as well as the rank formula.

Let $0 \le d_1 < d_2 < \cdots < d_r$ be integers such that $d_r \le t \le n - d_r$ and let $M = M_{d_1,t} + M_{d_2,t} + \cdots M_{d_r,t}$ be the lattice of all functions from $V\binom{[n]}{t}$ to $\mathbb{Z}$ which can be expressed as $\mathbb{Z}$-linear combinations of monomials x_G with $|G| \in \{d_1, \ldots, d_r\}$.

Theorem 5.1 *There is a set of polynomials*

$$B^* = \{z_G : G \in \mathrm{osh}\binom{[n]}{d_r}\} \subseteq M$$

such that B^* is a $\mathbb{Z}$-basis of M and $z_G = n_G \cdot x_G$ as functions on $V\binom{[n]}{t}$. The value of n_G is $\gcd\left(\binom{t-|G|}{d_1-|G|}, \ldots, \binom{t-|G|}{d_r-|G|}\right)$.

For the proof we need the following easy fact:

Proposition 5.2 *Let A be an Abelian group with generators $v_1, \ldots, v_k$. Suppose that the subgroups $A_i \le A$ for $1 \le i \le r$ are generated by the elements $m_{i,1}v_1, \ldots, m_{i,k}v_k$, with $m_{i,j} \in \mathbb{Z}$. Then the group $A_1 + A_2 + \ldots + A_r$ (i.e. the subgroup generated by all A_i's) is generated by $n_1v_1, n_2v_2, \ldots, n_kv_k$, where $n_j = \gcd(m_{1,j}, m_{2,j}, \ldots, m_{r,j})$.*

Proof. Clearly $m_{i,j}v_j$ is an integer multiple of n_jv_j, hence $\langle n_1v_1,\ldots,n_kv_k\rangle \geq A_1+\cdots+A_r$. Conversely, as n_j is an integer linear combination of the numbers $m_{1,j},\ldots,m_{r,j}$, we see that $n_jv_j \in A_1+\cdots+A_r$, hence $\langle n_1v_1,\ldots,n_kv_k\rangle = A_1+\cdots+A_r$. $\square$

Proof of Theorem 5.1. By Theorem 4.2 the space of functions $M_{d_i,t}$ is generated by the functions $\left\{\binom{t-|G|}{d_i-|G|}\cdot x_G : G \in \mathrm{osh}\binom{[n]}{t}\right\}$. We apply Proposition 5.2 with $A_i = M_{d_i,t}$ and $v_j = x_G$. We infer that the functions $\{n_Gx_G : G \in \mathrm{osh}\binom{[n]}{t}\}$ generate M. Clearly we have $n_G = 0$ if $|G| > d_r$ and $n_G \neq 0$ otherwise. Using that the x_G are linearly independent, we obtain that B^* is an independent set of functions. The proof is complete. $\square$

As in the case $r=1$, the diagonal form leads to a rank formula modulo p:

Corollary 5.3 *Suppose that $0 \leq d_1 < d_2 \cdots < d_r \leq t \leq n-d_r$ and let p be a prime. The dimension of the linear space of $\mathbb{F}_p$-valued functions on $V\binom{[n]}{t}$ which are $\mathbb{F}_p$-linear combinations of x_F, where $|F| \in \{d_1,\ldots,d_r\}$ is*

$$\sum_{\substack{0\leq i\leq d_r\\ p \nmid n_i}} \binom{n}{i} - \binom{n}{i-1}, \tag{9}$$

where $n_i = \gcd\left(\binom{t-i}{d_1-i},\binom{t-i}{d_2-i},\ldots,\binom{t-i}{d_r-i}\right)$.

Proof. Similarly to the case $r=1$ (Corollary 4.3), we obtain from Theorem 5.1 that the functions $\{n_G\cdot x_G : G \in \mathrm{osh}\binom{[n]}{d_r}\}$ generate the space of functions under consideration. Using that the x_G are independent over $\mathbb{F}_p$, we see that

$$\{n_G\cdot x_G : G \in \mathrm{osh}\binom{[n]}{d_r}, p \nmid n_G\}$$

is a basis of the space. By noting that n_i is the common value of the n_G with $|G| = i$, we conclude as in Corollary 4.3. $\square$

Remark. In terms of inclusion matrices, the quantity in (9) is the rank modulo p of the inclusion matrix

$$I\left(\binom{[n]}{t},\mathcal{H}\right)$$

where $\mathcal{H} = \binom{[n]}{d_1} \cup \binom{[n]}{d_2} \cup \ldots \cup \binom{[n]}{d_r}$.

For a family $\mathcal{F} \subseteq 2^{[n]}$ one can consider the family of sets complementary to the elements of $\mathcal{F}$:

$$co\mathcal{F} := \{[n] \setminus F : F \in \mathcal{F}\}.$$

Let $\mathcal{F}, \mathcal{G} \subseteq 2^{[n]}$ be two families. It is immediate that the transpose of the matrix $I(\mathcal{F}, \mathcal{G})$ can be identified with $I(co\mathcal{G}, co\mathcal{F})$. In particular, the matrices $I(\mathcal{F}, \mathcal{G})$ and $I(co\mathcal{G}, co\mathcal{F})$ have the same rank over any field. We have the following simple consequence of Corollary 5.3.

Corollary 5.4 *Suppose that* $0 \le d_1 < d_2 \cdots < d_r \le t \le n - d_r$ *and let* p *be a prime. Let* $\mathcal{G} := \binom{[n]}{n-d_1} \cup \binom{[n]}{n-d_2} \cup \ldots \cup \binom{[n]}{n-d_r}$. *Then the rank over* $\mathbb{F}_p$ *of the inclusion matrix* $I(\mathcal{G}, \binom{[n]}{n-t})$ *is*

$$\sum_{\substack{0 \le i \le d_r \\ p \not| n_i}} \binom{n}{i} - \binom{n}{i-1},$$

where $n_i = \gcd\left(\binom{t-i}{d_1-i}, \binom{t-i}{d_2-i}, \ldots, \binom{t-i}{d_r-i}\right)$. □

Acknowledgement. We are grateful to Tibor Szabó for useful discussions on the subject.

References

[1] R.P. Anstee, L. Rónyai, A. Sali, Shattering news, *Graphs and Combinatorics* **18** (2002), 59–73.

[2] L. Babai, P. Frankl *Linear algebra methods in combinatorics,* September 1992.

[3] T. Bier, Remarks on recent formulas of Wilson and Frankl, *Europ. J. Combin.* **14**(1993), 1–8.

[4] P. Frankl, Intersection theorems and mod p rank of inclusion matrices, *J. Combin. Theory A* **54**(1990), 85–94.

[5] R.M. Wilson, A diagonal form for the incidence matrices of t-subsets vs. k-subsets, *Europ. J. Combin.* **11**(1990), 609–615.

Corresponding author:
Lajos Rónyai
Computer and Automation Institute
Hungarian Academy of Sciences
H-1111 Budapest, Lágymányosi u. 11.
Hungary
E-mail: `ronyai@sztaki.hu`