JOURNAL OF COMBINATORIAL THEORY, Series A **54**, 85–94 (1990)

Intersection Theorems and Mod p Rank of Inclusion Matrices

P. FRANKL*

CNRS, Paris, France

Communicated by the Managing Editors

Received October 4, 1988

Higher incidence matrices have proved an important tool both in design theory and extremal set theory. In the present paper some tight bounds on the rank over finite fields of some inclusion matrices are derived. In particular, a short proof of Wilson's mod p rank formula is given. A problem of Graham, Li, and Li concerning bases for so-called null t-designs is solved as well. © 1990 Academic Press, Inc.

1. INTRODUCTION

Let X be an n-element set and $\mathscr{F} \subset 2^X$ a family of its subsets. For $n \geqslant s \geqslant 0$ one defines the *inclusion matrix* $I(\mathscr{F}, s)$ as a $|\mathscr{F}|$ by $\binom{n}{s}$ matrix whose rows are indexed by $F \in \mathscr{F}$, the columns by $G \in \binom{X}{s}$ and the general entry is

$$i(F, G) = \begin{cases} 1 & \text{if } F \supset G \\ 0 & \text{if } F \not\supset G. \end{cases}$$

For matrices $M_1, ..., M_r$ having the same number of rows let $M_1 \mid M_2 \mid \cdots \mid M_r$ denote the matrix obtained by putting $M_1, ..., M_r$ next to each other. Set

$$I^*(\mathscr{F}, s) = I(\mathscr{F}, s) \mid I(\mathscr{F}, s-1) \mid \cdots \mid I(\mathscr{F}, 0).$$

Note the trivial inequality

$$\operatorname{rank}_p I^*(\mathscr{F}, s) \leqslant \sum_{0 \leqslant i \leqslant s} \binom{n}{i}, \tag{1}$$

where $\operatorname{rank}_p M$ denotes the rank of an integer matrix over $GF(p)$.

For sets K, $L \subset \{0, 1, ..., n\}$ the family $\mathscr{F}$ is called a (K, L)-*system* if $|F| \in K$ for all $F \in \mathscr{F}$ and $|F \cap F'| \in L$ for all distinct, F, $F' \in \mathscr{F}$.

* This research was done while the author was visiting AT&T Bell Laboratories, Murray Hill, NJ 07974.

0097-3165/90 $3.00

Inclusion matrices have proved useful in obtaining upper bounds for the maximum size, $m(n, K, L)$ of (K, L)-systems.

Let us recall the following result.

THEOREM 1.0 [FW]. *Suppose that $\mathscr{F} = \{F_1, ..., F_m\} \subset 2^X$, p is a prime and $q_1(x), ..., q_m(x)$ are integer-valued polynomials of degree at most s satisfying*

$$p \nmid q_i(|F_i|) \qquad \textit{for} \quad 1 \leqslant i \leqslant m$$

$$p \mid q_i(|F_i \cap F_j|) \qquad \textit{for} \quad 1 \leqslant i < j \leqslant m.$$

Then

$$\operatorname{rank}_p I^*(\mathscr{F}, s) = |\mathscr{F}|. \tag{2}$$

For example, if $\mu_0, \mu_1, ..., \mu_s$ are distinct residues mod p such that $k = \mu_0$ (mod p) for all $k \in K$ and $l \in \{\mu_1, ..., \mu_s\}$ (mod p) for all $l \in L$ then one can take $q_i(x) = \prod_{1 \leqslant j \leqslant s} (x - \mu_j)$ for $1 \leqslant i \leqslant m$.

In view of (1) this implies

$$|\mathscr{F}| \leqslant m(n, K, L) \leqslant \sum_{0 \leqslant j \leqslant s} \binom{n}{j}. \tag{3}$$

Very recently, Alon and Babai [AB] found a nice but complicated argument using spaces of polynomials to replace the RHS of (3) ny $\binom{n}{s}$. (Note that the case $s = 1$ was solved in [FR].) Here we show that, indeed, this can be derived from Theorem 1.0 as well.

Define $I^*(K, s)$ as $I^*(\mathscr{G}, s)$, where $\mathscr{G} = \bigcup_{k \in K} \binom{X}{k}$.

Our principal result is

THEOREM 1.1. *Suppose that $k \equiv k'$* (mod p) *for all $k, k' \in K$ and $0 \leqslant s < p$. Then*

$$\operatorname{rank}_p I^*(K, s) \leqslant \binom{n}{s}. \tag{4}$$

Since, $I^*(\mathscr{F}, s)$ is a submatrix of $I^*(K, s)$ in the above situation, Theorem 1.0 and (4) imply $m(n, K, L) \leqslant \binom{n}{s}$ as claimed.

Let us mention that in most cases, e.g., for $n \geqslant p + 2s$, equality holds in (4). For the proof we need two simple facts.

PROPOSITION 1.2. *Suppose that $0 \leqslant r < s < p$ and $r + s \leqslant n$. Then*

$$\operatorname{rank}_p I(s, r) = \binom{n}{r}. \tag{5}$$

PROPOSITION 1.3. *The general entry $m(F, A)$ of the matrix $I(\mathscr{F}, s)\, I(s, t)$ is*

$$m(F, A)=\begin{cases}\dbinom{|F|-t}{s-t} & \textit{if} \quad A\subset F\\ 0 & \textit{otherwise.}\end{cases}$$

This statement is both well known and trivial

The rank of inclusion matrices has aroused a lot of interest. The fact that

$$\operatorname{rank}_Q I(a, b)=\binom{n}{b} \qquad \text{for all} \quad a\geqslant b\geqslant 0,\ n\geqslant a+b, \tag{6}$$

was proved around the same time by Kantor [K], Graver-Jurkat [GJ], and Wilson [W1]; $\operatorname{rank}_2 I(a, a-1)=\binom{n-1}{a-1}$ follows easily, e.g., by using the exact sequence arising from the boundary operator for simplicial complexes. Linial and Rothschild [LR] succeeded in determining $\operatorname{rank}_2 I(a, b)$ in general.

Finally, Wilson [W2] found a beautiful but complex argument to compute the Smith norma form of $I(a, b)$ thereby determining $\operatorname{rank}_p I(a, b)$ for all p. The formula, which clearly implies (5), is

$$\operatorname{rank}_p I(a, b)=\sum\left\{\binom{n}{i}-\binom{n}{i-1}: 0\leqslant i\leqslant r,\ p\nmid\binom{a-i}{b-i}\right\}, \qquad n\geqslant a+b. \tag{7}$$

In Section 4 a simple, short proof of (7) is given. It is based on Corollary 3.4, which exhibits a special basis for the column space of $I^*(a, b)$. In Section 3 universal bases for the vector space of so-called null t-designs are constructed (Theorem 3.2) thereby solving a problem of Graham, Li, and Li [GLL]. The proof of Theorem 1.1 is given in Section 5. In Section 6 further problems are discussed. The paper is self-contained.

2. A PARTITION OF THE k-ELEMENT SETS

Let us represent every subset $F\subset[n]=\{1, 2, ..., n\}$ by a *walk*, $w(F)$ going from the origin to $(n-|F|, |F|)$ by steps of length one, the ith step is to the right or up according as $i\notin F$ or $i\in F$ holds.

The *rank*, $r(F)$ is defined as $|F|-j$, where j is largest integer such that $w(F)$ reaches the line $y=x+j$. E.g., $r([n])=0$, $r(\{1, 3, 4\})=1$. Define $\mathscr{G}(n, k, r)=\{G\subset[n],\ |G|=k,\ r(G)=r\}$.

CLAIM 2.1. *$|\mathcal{G}(n,k,r)| = \binom{n}{r} - \binom{n}{r-1}$ for $0 \leqslant r \leqslant k$ (where $\binom{n}{-1} = 0$).*

Proof. By the reflection principle the number of walks from $(0,0)$ to $(n-k,k)$ and hitting the line $y = x + k - r$ is the same as the number of walks from $(r-k,\ k-r)$ to $(n-k,k)$, that is $\binom{n}{r}$. Thus $\sum_{0 \leqslant j \leqslant r} |\mathcal{G}(n,k,r)| = \binom{n}{r}$ holds for $0 \leqslant r \leqslant k$, which in turn implies the claim. ▮

For a set G of rank r let us define a set $\tilde{G} = \{j_1, ..., j_r\}$ in the following way. Suppose that $G = \{i_1, ..., i_s\}$ with $i_1 > \cdots > i_s$. Choose $j_r \in ([n] - G)$ maximal with respect to $j_r < i_r$.

Once j_b is defined for $b > l$ choose $j_l \in ([n] - (G \cup \{j_{l+1}, ..., j_r\}))$ maximal with respect to $j_l < i_l$. The fact that we never get stuck is the content of the next proposition.

CLAIM 2.2. *The set $\tilde{G} \in \binom{[n]}{r}$ is well defined.*

Proof. Indeed, the only problem would be if we cannot choose some element, say j_l, that is $[i_l] \subset G \cup \{j_{l+1}, ..., j_r\}$. Since $|G \cap [i_l]| = k - l + 1$, $i_l \leqslant k - l + 1 + r - l$. Consequently, after i_l steps the walk $w(G)$ is on or above the line $y = x + k - r + 1$, contradicting the definition of $r = r(G)$. ▮

Remark 2.3. Note that if $r(G - \{i_1\}) = r - 1$ then $G - \{i_1\} = \{j_2, ..., j_r\}$ holds; this fact will be used in (3.1). ▮

3. UNIVERSAL BASES FOR INCLUSION MATRICES

Let us fix $1 \leqslant k \leqslant n$ and consider the *reverse lexicographic order* on $\binom{X}{k}$. This is a linear order for all $F, G \in \binom{[n]}{k}$ defined by $F < G$ iff $\max F - G < \max G - F$.

Let $x_1, ..., x_n$ be indeterminates and for a set F define $x_F = \prod_{i \in F} x_i$.

Let $V = V(n,k)$ be the vector space over some fixed field Γ of all formal linear combinations $\sum_{F \in \binom{[n]}{k}} \alpha(F)\, x_F$. In other words V is the vector space of all square-free, homogeneous polynomials of degree k.

For a set $G \in \binom{[n]}{k}$ of rank r we define the polynomial $p(G)$ in the following way. Let $G = \{i_1, ..., i_k\}$, $\tilde{G} = \{j_1, ..., j_r\}$ (cf. definition of $\tilde{G}$ in Section 2), where $i_1 > \cdots > i_k$. Define

$$p(G) = (x_{i_1} - x_{j_1}) \cdots (x_{i_r} - x_{j_r})\, x_{i_{r+1}} \cdots x_{i_k}.$$

Note that by Remark 2.3 we have

$$\frac{\partial p(G)}{\partial x_{i_1}} = p(G - \{i_1\}) \qquad \text{if} \quad i_1 > k + r. \tag{3.1}$$

For example in the case $k=1$ we obtain the polynomials x_1, $x_2-x_1, ..., x_j-x_1, ...$. It is easy to see that the first $\binom{j}{1}$ of them form a basis of $V(j, 1)$. We shall see that the $p(G)$ have this and some stronger properties as well.

DEFINITION 3.1. A polynomial $\sum_{F\in\binom{[n]}{k}} \alpha(F)\, x_F$ *is called a null t-design if*

$$\sum_{T\subset F} \alpha(F)=0 \qquad \text{holds for all} \quad |T|\leqslant t. \tag{3.2}$$

Note that for fields of characteristic zero this definition is equivalent to the one given in [GLL].

Let $V_t=V(n, k, t)$ be the subspace of V spanned by the null t-designs. For convenience we set $V_{-1}=V$.

THEOREM 3.2. *For all $d\geqslant k\geqslant r\geqslant 0$ the polynomials $p(G)$, $G\in \bigcup_{r\leqslant j\leqslant k}\mathscr{G}(d, k, j)$ form a basis of $V_{r-1}(d, k)$.*

Remark. Note that for $d<k+r$ all the families $\mathscr{G}(d, k, j)$ are empty, implying $V_{r-1}(d, k)=\langle 0\rangle$.

Proof. We apply induction on d and prove the statement simultaneously for all k and r. The case $d=k$ is trivial and can serve as the base step.

The proof is very simple in the case $r=0$ as well. Namely, G is the largest set F for which x_F occurs in the expansion of $p(G)$. Consequently, the $p(G)$ are linearly independent and their number is $\binom{d}{k}$.

Let $U_b=\langle p(G): G\in \bigcup_{i\leqslant b}\mathscr{G}(d, k, i)\rangle$. By what we have just proved, $\dim U_b=\binom{d}{b}$ holds for $0\leqslant b\leqslant d-k$.

CLAIM 3.3. $U_b\cap V_b=\langle 0\rangle$.

Proof. In the case $b=0$ one has $U_0=\langle x_1\cdot ...\cdot x_k\rangle$ and the statement is trivial. Thus we may suppose that $b\geqslant 1$ and apply induction on b. Consider $q(\mathbf{x})=\sum_G \alpha(G)\, p(G)\in U_b$.

Let G be the largest set (in the reverse lexicographic order) with $\alpha(G)\neq 0$. Set $r=r(G)$ and note $0\leqslant r\leqslant b$.

Let i_1 be the maximal element of G. We distinguish two cases.

(i) $i_1>k+r$. Consider the polynomial $q^*(\mathbf{x}) = \partial q(\mathbf{x})/\partial x_{i_1} \in U_{b-1}(i_1-1, k-1)$. By the induction hypothesis and (3.1) there exists some $S\in\binom{[i_1-1]}{b-1}$ with $\sum_{S\subset F}\alpha^*(F)\neq 0$ where

$$q^*(\mathbf{x})=\sum \alpha^*(F)\, x_F.$$

Since $\alpha^*(F) = \alpha(F \cup \{i_1\})$, we have

$$\sum_{(S \cup \{i_1\}) \subset H} \alpha(H) \neq 0;$$

i.e., $q(\mathbf{x}) \notin V_b$ as desired.

(ii) $i_1 \leqslant k + r$. In this case we claim that $q(\mathbf{x}) \notin V_r$. For $|A| \leqslant r$ define

$$B(A) = \sum_{A \subset F} \alpha(F).$$

Set $B = [k+r] - G$, $|B| = r$. Define

$$\delta(B) = \sum_{B \cap F = \varnothing} \alpha(F).$$

By inclusion-exclusion we have

$$\delta(B) = \sum_{A \subset B} (-1)^{|A|} B(A).$$

If $q(\mathbf{x}) \in V_r$ then $B(A) = 0$ and thus $\delta(B) = 0$ follows. However, the maximal choice of G implies $F \subset [k+r]$ whenever $\alpha(F) \neq 0$. Consequently, $\delta(B) = \alpha(G) \neq 0$, a contradiction. ▌

Now it is easy to conclude the proof of the theorem. By Claim 3.3 and dim $U_b = \binom{d}{b}$ we have dim $V_b \leqslant \binom{d}{k} - \binom{d}{b}$.

On the other hand, $\{p(G) : G \in \bigcup_{b < j \leqslant k} \mathscr{G}(d, k, j)\}$ are $\binom{d}{k} - \binom{d}{b}$ linearly independent polynomials in this space, consequently they form a basis. ▌

Corollary 3.3 [W2]. *The rank of $I^*(k, t)$ is $\binom{n}{t}$ over an arbitrary field, Γ for $n \geqslant k + t$.*

Proof. Note that V_t is the kernel of this matrix viewed as a linear transformation $\Gamma^{\binom{n}{k}} \to \Gamma^{\binom{n}{t} + \cdots + \binom{n}{0}}$. ▌

For a matrix M let $\mathrm{col}_\Gamma M$ be its column-space over Γ.

Corollary 3.4. *For every field Γ and for all $a \geqslant b$, $a + b \leqslant n$ one can choose a basis $v_1, ..., v_{\binom{n}{b}}$ of $\mathrm{col}_\Gamma I^*(a, b)$ such that for every $0 \leqslant j \leqslant b$ and $\binom{n}{j-1} < i \leqslant \binom{n}{j}$ the vector v_i is from $I(a, j)$.*

Proof. Apply induction on b. For $b = 0$ just take $v_1 = (1, ..., 1)$. Once a basis for $\mathrm{col}^*_\Gamma(a, b-1)$ is chosen, by $\mathrm{rank}_\Gamma I^*(a, b) - \mathrm{rank}_p I^*(a, b-1) = \binom{n}{b} - \binom{n}{b-1}$, we can extend it by $\binom{n}{b} - \binom{n}{b-1}$ vectors from $I(a, b)$ to a basis of $\mathrm{col}_\Gamma I^*(a, b)$. ▌

4. The Short Proof of Wilson's Rank Formula

For a prime p and fixed integers $a \geqslant b \geqslant 0$ define two sets

$$S = \left\{0 \leqslant i \leqslant a : p \nmid \binom{a-i}{b-i}\right\}, \qquad R = [0, a] - S.$$

Let us restate (7) with this notation.

Theorem 4.1 (Wilson [W2]). *Suppose that $n \geqslant a+b$ then*

$$\operatorname{rank}_p I(a, b) = \sum_{i \in S} \binom{n}{i} - \binom{n}{i-1}.$$

Clearly, this statement follows from Lemmas 4.2 and 4.3 below. For $B \subset \{0, 1, ..., b\}$ let $M(a, B)$ denote the matrix (with $\binom{n}{a}$ rows) formed by all the vectors v_i satisfying $v_i \in I(a, \tilde{b})$ with $\tilde{b} \in B$ (v_i is defined in Corollary 3.4).

Lemma 4.2.

$$\operatorname{rank}_p I(a, b) \leqslant \sum_{i \in S} \binom{n}{i} - \binom{n}{i-1}. \tag{4.1}$$

Proof. Consider the product $I(a, b)\, M(b, R)$. Its value is the all-zero matrix by Proposition 1.3 and the choice of R. By Corollary 3.4 it follows that

$$\operatorname{rank}_p M(b, R) = \sum_{i \in R} \binom{n}{i} - \binom{n}{i-1}.$$

Consequently,

$$\operatorname{rank}_p I(a, b) \leqslant \binom{n}{b} - \sum_{i \in R} \binom{n}{i} - \binom{n}{i-1} = \sum_{i \in S} \binom{n}{i} - \binom{n}{i-1}. \quad \blacksquare$$

Lemma 4.3.

$$\operatorname{rank}_p I(a, b) \geqslant \sum_{i \in S} \binom{n}{i} - \binom{n}{i-1}. \tag{4.2}$$

Proof. By Proposition 1.3 we have

$$\operatorname{col}_p I(a, i) \subset \operatorname{col}_p I(a, b) \qquad \text{for} \quad i \in S.$$

Consequently, $v_j \in \operatorname{col}_p I(a, b)$ for $\binom{n}{i-1} < j \leqslant \binom{n}{i}$ and $i \in S$, with the notation of Corollary 3.4. These vectors are independent, proving (4.2). $\blacksquare$

5. Proof of Theorem 1.1

Let us first prove an easy, kind of well-known statement. For a matrix M let $\operatorname{col}_p M$ denote its column space over $GF(p)$.

Proposition 5.1. *Suppose that* $0<d<p$ *and*

$$\binom{|F|-t}{s-t} \equiv d \pmod p \qquad \textit{for all} \quad F \in \mathcal{F}.$$

Then

$$\operatorname{col}_p(I(\mathcal{F}, t)) \subset \operatorname{col}_p(I(\mathcal{F}, s)) \qquad \textit{holds.} \tag{5.1}$$

Proof. By Proposition 1.3,

$$I(\mathcal{F}, t) = \frac{1}{d}\, I(\mathcal{F}, s)\, I(s, t)$$

holds, implying (5.1). ▮

Corollary 5.2. *Suppose that* $0<s\leqslant r<p$ *and* $|F| \equiv r \pmod p$ *for all* $F \in \mathcal{F}$. *Then*

$$\operatorname{col}_p(I(\mathcal{F}, t)) \subset \operatorname{col}_p(I(\mathcal{F}, s) \tag{5.2}$$

holds for all $0 \leqslant t \leqslant s$.

Note that (5.2) implies

$$\operatorname{rank}_p(I^*(s, \mathcal{F})) = \operatorname{rank}_p(I(s, \mathcal{F})) \leqslant \binom{n}{s}.$$

Another use of Proposition 1.3 is the following.

Proposition 5.3. *Suppose that* $p \mid \binom{|F|-t}{s-t}$ *for all* $F \in \mathcal{F}$. *Then*

$$\operatorname{rank}_p(I(\mathcal{F}, s)) \leqslant \binom{n}{s} - \operatorname{rank}_p(I(s, t)) \qquad \textit{holds.} \tag{5.3}$$

Proof. By Proposition 1.3, $I(\mathcal{F}, s)\, I(s, t)$ is the zero matrix over $GF(p)$. ▮

Now we can easily derive Theorem 1.1. Define $0 \leqslant r < p$ by $k \equiv r \pmod p$ for all $k \in K$. If $s \leqslant r$ then it follows from Corollary 5.2 that $\operatorname{col}_p(I^*(K, s)) = \operatorname{col}_p(I(K, s))$ and this latter matrix has only $\binom{n}{s}$ columns proving (4).

Let now $0 \leqslant r < s < p$. For $n < r + p$, $I^*(K, s) = I^*(r, s)$ and this latter arises from $I^*(r, r)$ by adding some all zero columns. Thus (4) follows. Again by Corollary 5.2 and the above case for $s = r$ we have

$$\operatorname{col}_p(I(K, s) \mid I(K, s-1) \mid \cdots \mid I(K, r+1)) = \operatorname{col}_p(I(K, s))$$

and

$$\operatorname{col}_p(I^*(K, r)) = \operatorname{col}_p(I(K, r)).$$

This yields using (5.3):

$$\operatorname{rank}_p I^*(K, s) \leqslant \operatorname{rank}_p I(K, s) + \operatorname{rank}_p I(K, r) \leqslant \binom{n}{s} + \binom{n}{r} - \operatorname{rank}_p I(s, r).$$

To prove (4), we need $\operatorname{rank}_p(I(s, r)) = \binom{n}{r}$. This, however, follows from Proposition 1.2 by $n \geqslant r + p > r + s$. ▮

6. More Rank Formulas

There are many more problems to be considered. For example, what happens if we drop the condition $s < p$ in Theorem 1.1? Or if the numbers $k \in K$ are allowed to belong to t residue classes modulo p.

We shall return to these problems in a forthcoming paper. Let us just announce the following extension of another result of Alon and Babai [AB].

THEOREM 6.1. *Suppose that $|K| = t$. Then*

$$\operatorname{rank} I^*(K, s) \leqslant \binom{n}{s} + \binom{n}{s-1} + \cdots + \binom{n}{s-t+1}.$$

References

[AB] N. Alon and L. Babai, Multilinear polynomials and the Frankl–Wilson intersection theorem, to appear.

[B] L. Babai, The non-uniform Ray-Chaudhuri–Wilson theorem, *Combinatorica* **8** (1988), 133–135.

[FR] P. Frankl and I. G. Rosenberg, Regularly intersecting families of finite sets, *Ars Combin.* **22** (1986), 97–105.

[FW] P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences, *Combinatorica* **1** (1981), 357–368.

[GLL] R. L. Graham, S. Y. R. Li, and W. C. Li, On the structure of t-designs, *SIAM J. Algebraic Discrete Meth.* **1** (1980), 8–14.

[GJ] J. E. Graver and W. B. Jurkat, The module structure of integral designs, *J. Combin. Theory Ser. A* **15** (1973), 75–90.

[K] W. M. Kantor, On incidence matrices of finite projective and affine spaces, *Math. Z.* **124** (1972), 315–318.

[LR] N. Linial and B. L. Rothschild, Incidence matrices of subsets—A rank formula, *SIAM J. Algebraic Discrete Meth.* **2** (1981), 333–340.

[RW] D. K. Ray-Chaudhuri and R. M. Wilson, On t-designs, *Osaka J. Math.* **12** (1975), 737–744.

[W1] R. M. Wilson, The necessary conditions for t-designs are sufficient for something, *Utilitas Math.* **4** (1973), 207–215.

[W2] R. M. Wilson, The Smith normal form of inclusion matrices, manuscript, 1980.