Discrete Mathematics 84 (1990) 135–141
North-Holland

AN INTERSECTION PROBLEM FOR CODES

P. FRANKL
CNRS, Paris, France

Received 22 June 1988

Let $[s]^n$ denote all sequences $\vec{a} = (a_1, \ldots, a_n)$ of integers with $1 \le x_i \le s$. Consider a subset A of $[s]^n$. It is called $(t_1, \ldots, t_s)$-intersecting if for any two members $\vec{a}, \vec{b} \in A$ and any $1 \le i \le s$ there are at least t_i positions j, where both $\vec{a}$ and $\vec{b}$ have entry i, that is, $a_j = b_j = i$. The problem of determining $\max |A|$ for A being $(t_1, \ldots, t_s)$-intersecting is considered. In particular, the case $t_1 = t_2 = \cdots = t_s = 1$ is solved completely.

1. Introduction

Let $A = \{a_1, \ldots, a_s\}$ be an alphabet of size s and consider $X = A^n$, the set of all words $(x_1, \ldots, x_n)$ of length n over A, i.e. $x_i \in A$ for all i.

A set $C \subset X$ is called a code. It is called $(t_1, \ldots, t_s)$-*intersecting* if for all $1 \le i \le s$ and for any two members of C there are at least t_i coordinate places where both have a_i.

Setting $\vec{t} = (t_1, \ldots, t_s)$ we shall speak of $\vec{t}$-intersecting codes. To avoid trivialities we suppose that $n \ge t_1 + \cdots + t_s$.

Definition 1.1. Let $m(n, \vec{t})$ denote the maximum size of a $\vec{t}$-intersecting code in A^n.

For $1 \le i \le s$ set $\vec{t}_i = (0, 0, \ldots, 0, t_i, 0, \ldots, 0)$. Recently, Winkler [13] has formulated the following conjecture.

Conjecture 1.2.

$$m(n, \vec{t}) = \max_{n_1 + \cdots + n_s = n} \prod_{i=1}^{s} m(n_i, \vec{t}_i).$$

This conjecture would reduce the determination of $m(n, \vec{t})$ to the special case when all but one of the t_i's are equal to zero.

We shall discuss this special case in Section 2.

Let us mention that in the case $s = 2$ one can formulate Conjecture 1.2 in terms of families of sets. Namely, define for $\vec{x} \in C$ the set $F(\vec{x}) = \{i : x_i = a_1\}$; $\mathscr{F}(C) = \{F(\vec{x}) : \vec{x} \in C\}$.

Then (t_1, t_2)-intersecting means $|F_1 \cap F_2| \ge t_1$ and $|F_1 \cup F_2| \le n - t_2$ for all $F_1, F_2 \in \mathscr{F}(C)$.

For this special case the conjecture was formulated by Bang et al. [1] and in a more general form by the author [5].

The special case $t_1 = 1$ was conjectured by Katona [11] and settled by the author [6].

In this note we shall prove Conjecture 1.2 in the special case $t_i < s$ for all i.

Theorem 1.3. *Suppose that $t_i < s$ for $1 \leq i \leq s$. Then*

$$m(n, \vec{t}) = s^{n-t_1-\cdots-t_s} \quad \text{holds for all } n \geq t_1 + \cdots + t_s.$$

The proof of this result will be given in Section 4 after some preparations in Section 3.

2. Convex hulls of $\vec{f}$-vectors of t-intersecting families

Let $\mathcal{F} \subset 2^{[n]}$ be a family of subsets of $[n] = \{1, 2, \ldots, n\}$, $\mathcal{F}$ is called t-intersecting if $|F \cap F'| \geq t$ holds for all $F, F' \in \mathcal{F}$.

Example 2.1. For $0 \leq i \leq (n-t)/2$ define $\mathcal{B}_i = \{B \subset [n]: |B \cap [t+2i]| \geq t+i\}$.

Clearly, $\mathcal{B}_i$ is t-intersecting and Katona [10] proved that among all t-intersecting families $\mathcal{B}\lfloor(n-t)/2\rfloor$ has the largest size.

For $i + t < k \leq (n+t)/2$ define also

$$\mathcal{B}_i^k = \{B \subset B_i : |B| \geq k\} \cup \{A \subset X : |A| \geq n + t - k\}.$$

For a family, $\mathcal{F} \subset 2^{[n]}$ its *f-vector* $\vec{f}(\mathcal{F}) = (f_0, \ldots, f_n)$ is defined by

$$f_i = |\{F \in \mathcal{F} : |F| = i\}|, \quad 0 \leq i \leq n.$$

Definition 2.2. A set $Z = \{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$ of t-intersecting families $\mathcal{F}_i \subset 2^{[n]}$ is called *dominating* if for every t-intersecting family $\mathcal{F} \subset 2^{[n]}$ there exist nonnegative reals $\alpha_1, \ldots, \alpha_m$ with $\alpha_1 + \cdots + \alpha_m = 1$, such that $\vec{f}(\mathcal{F}) \leq \sum \alpha_i \vec{f}(\mathcal{F}_i)$ holds coordinatewise, that is $f_j(\mathcal{F}) \leq \sum \alpha_i f_j(\mathcal{F}_i)$ for $0 \leq j \leq n$.

Conjecture 2.3 (Cooper [2]). $\{\mathcal{B}_i^k : 0 \leq i \leq (n-t)/2,\ t + i < k \leq (n+t)/2\}$ is a dominating set for $1 \leq t \leq n$.

Let us mention that the case $t = 1$ was solved by Erdös et al. [3].

Conjecture 2.3 would have several important corollaries, e.g. it would imply that the largest size of a t-intersecting family of k-element sets is $f_k(\mathcal{B}_i)$ for some $1 \leq i \leq (n-t)/2$.

Proposition 2.4.

$$m(n, (t, 0, \ldots, 0)) = \max_{\vec{f}} \sum_{0 \leq j \leq n} f_j (s-1)^{n-j} \tag{2.1}$$

where the maximum is over the f-vectors $(f_0, \ldots, f_n)$ of all t-intersecting families.

Proof. Let C be a $(t, 0, \ldots, 0)$-intersecting code of maximal size. For $\vec{x} = (x_1, \ldots, x_n) \in C$ defined $F(\vec{x}) = \{i : x_i = a_1\}$. Then $\mathscr{F}(C) = \{F(\vec{x}) : \vec{x} \in C\}$ is t-intersecting and the maximality of C implies that for given $F \in \mathscr{F}(C)$ all $(s-1)^{n-|F|}$ words $\vec{x} \in X$ with $F(\vec{x}) = F$ are in C.

On the other hand, if $\mathscr{F}$ is t-intersecting then $C(\mathscr{F}) = \{\vec{x} \in X^n : F(\vec{x}) \in \mathscr{F}\}$ is $(t, 0, \ldots, 0)$-intersecting of size $\sum_{0 \leq j \leq n} f_j(\mathscr{F})(s-1)^{n-j}$.

Consequently, the determination of $m(n, (t, 0, \ldots, 0))$ is equivalent to determine $\max \sum f_j(s-1)^{n-j}$ where the maximum is over all f-vectors $(f_0, \ldots, f_n)$ of t-intersecting families. □

Since by Conjecture 2.3 the $\mathscr{B}_i$'s form a dominating set and the coefficients $(s-1)^{n-j}$ are nonnegative, we infer

Corollary 2.5. *If Conjecture* 2.3 *is true, then*

$$m(n, (t, 0, \ldots, 0)) = \max_{i,k} \sum_{0 \leq j \leq n} f_j(\mathscr{B}_i^k)(s-1)^{n-j}.$$

3. A Kleitman-type result

For $\vec{x} = (x_1, \ldots, x_n) \in A^n$ define $S_i(\vec{x}) = \{j : x_j = a_i\}$. Clearly $[n] = \{1, \ldots, n\} = S_1(\vec{x}) \cup \cdots \cup S_s(\vec{x})$ is a partition.

Call a code $C \subset A^n$ *i-closed* if for $\vec{x}, \vec{y} \in A^n S_i(\vec{x}) \subseteq S_i(\vec{y})$ and $\vec{x} \in C$ imply $\vec{y} \in C$.

Theorem 3.1. *Suppose that $C_i \subset A^n$ is i-closed for $1 \leq i \leq s$. Then*

$$|C_1 \cap \cdots \cap C_s|/s^n \leq \prod_{1 \leq i \leq s} |C_i|/s^n \quad \textit{holds}. \tag{3.1}$$

Proof. For $n = 0$ each of C_i is empty or consists of the empty word. Thus (3.1) holds. Apply induction on n, $n \geq 1$. For $1 \leq i \leq s$ let $C_i(j)$ denote the code obtained from the codewords of C_i which have a_j in the last coordinate position by deleting this last position. Using the definition and i-closedness we have

$$|C_i(1)| + \cdots + |C_i(s)| = |C_i|,$$
$$|C_i(i)| \geq |C_i(j)| \quad \text{for } i \neq j \quad \text{and} \quad |C_i(j)| = |C_i(j')| \quad \text{for } i \notin \{j, j'\}. \tag{3.2}$$

Set $p_i = |C_i(i)|/|C_i|$, $q_i = |C_i(j)|/|C_i|$ for some $j \neq i$. In view of (3.2) we have

$$p_i + (s-1)q_i = 1, \quad q_i \leq 1/s \leq p_i. \tag{3.3}$$

Claim.

$$\sum_{1 \leq j \leq s} q_1 \cdot \cdots \cdot q_{j-1} p_j q_{j+1} \cdot \cdots \cdot q_s \leq s^{1-s}$$

holds with equality if and only if $q_i = p_i = 1/s$ holds for at least $s-1$ values of $i = 1, \ldots, s$.

Proof of the Claim. Set $p_i = 1 - (s-1)q_i$ and consider the partial derivative of the LHS with respect to q_i. It is the sum of s terms, one negative and $(s-1)$ positive. The negative term is $-(s-1)q_i \cdots q_s$, while each positive is a similar product, except that the coefficient is one, and one of the q_j's is replaced by p_j which is – by (3.3) – not less in value. This implies the nonnegativeness of the derivative. It is even strictly positive, unless $q_i = p_i$, i.e. $q_i = 1/s$.

Thus we increase the value of LHS by setting $q_1 = \cdots = q_s = 1/s$ and then its value is s^{1-s}. □

Now using $|C_1 \cap \cdots \cap C_s| = \sum_{1 \leq j \leq s} |C_1(j) \cap \cdots \cap C_s(j)|$, applying the induction hypothesis and using the claim we obtain:

$$|C_1 \cap \cdots \cap C_s|/s^n \leq \frac{1}{s} \prod_{i=1}^{s} |C_i|/s^{n-1} \sum_{j=1}^{s} q_1 \cdot \cdots \cdot q_{j-1} p_j q_{j+1} \cdot \cdots \cdot q_s \leq \prod_{1 \leq i \leq s} |C_i|/s^n.$$

□

Remark. Note that in the case $s = 2$ inequality (3.1) was already proved by Kleitman [9].

4. The main result

Let $C \subset A^n$ be a $\vec{t}$-intersecting code.

Let $\tilde{C} \subset A^{n+1}$ be obtained by adding for all codewords in C in all possible ways a $(n+1)$th coordinate. Clearly, $|\tilde{C}| = s\,|C|$ and $\tilde{C}$ is $\vec{t}$-intersecting. This implies:

Proposition 4.1. *$m(n, \vec{t})/s^n$ is monotone increasing and thus*

$$p(\vec{t}) = \lim_{n \to \infty} m(n, \vec{t})/s^n \quad \textit{exists.}$$

Let $\vec{t} = (t_1, \ldots, t_s)$ and set

$$p(t_i) = \lim_{n \to \infty} m(n, (0, 0, \ldots, t_i, \ldots, 0))/s^n.$$

Theorem 4.2. $p(\vec{t}) = p(t_1) \cdot \cdots \cdot p(t_s)$ *holds.*

Proof. $p(t) \geq p(t_1) \cdot \cdots \cdot p(t_s)$ follows from

$$m(n, \vec{t}) \geq \prod_{1 \leq i \leq s} m(n_i, t_i) \quad \text{for all } n_i \geq t_i \quad \text{with } n_1 + \cdots + n_t = n.$$

To prove the upper bound we show:

$$m(n, \vec{t})/s^n \leq \prod_{1 \leq i \leq s} m(n, t_i)/s^n. \tag{4.1}$$

To prove (4.1) let $C \subset A^n$ be $\vec{t}$-intersecting. Define the i-closed family C_i by

$$C_i = \{\vec{y} \subset A^n : \exists \vec{x} \in C,\ S_i(\vec{x}) \subseteq S_i(\vec{y})\}.$$

Then C_i is clearly $(0, \ldots, 0, t_i, 0, \ldots, 0)$-intersecting. Thus $|C_i| \leqslant m(n, t_i)$ holds.
Since $C \subset C_1 \cap \cdots \cap C_s$, an application of Theorem 3.1 yields (4.1).
Now $p(\vec{t}) \leqslant p(t_1) \cdot \cdots \cdot p(t_s)$ follows by taking limits of both sides. □

Even if Theorem 4.2 does not prove Winkler's conjecture (Conjecture 1.2), it shows that it is asymptotically correct. Slightly more would follow from Conjecture 2.3.

Proposition 4.3. *If* $s \geqslant 3$ *and Conjecture* 2.3 *is true then for* $n \geqslant n_0(\vec{t})$ *Conjecture* 1.2 *is true.*

Proof (sketch). Considering the RHS of (2.1) one sees that the maximum is attained for some $i < q(\vec{t}, s)$, where $q(\vec{t}, s)$ is independent of n (here we used $s \geqslant 3$). This on the other hand implies $m(n+1, \vec{t}_i) = sm(n, \vec{t}_i)$ for $n \geqslant 2q(\vec{t}, s) + t_i$. That is, $m(n, \vec{t}_i) = p(t_i)s^n$ holds for $n \geqslant 2q(\vec{t}, s) + t_i$. From (4.1) we infer

$$m(n, \vec{t}) \leqslant p(t_1) \cdot \cdots \cdot p(t_j)s^n, \quad \text{and for } n \geqslant \sum_{1 \leqslant i \leqslant s} (2q(\vec{t}, s) + t_i)$$

we can have equality here by the obvious product construction. □

Remark. What one needs for the proof is that in the maximum-sized $\vec{t}_i$-intersecting families the intersection property is assured by a set of bounded size of the coordinates. This might be easier to prove than Conjecture 2.3.

Proposition 4.4. $m(n, (t, 0, \ldots, 0)) = s^{n-t}$ *holds if* $s \geqslant t+1$ *and* $n \geqslant t$.

Proof. This result was proved by Frankl and Füredi [8] for $t \geqslant 15$ and Moon [12] gave a sharpening for $s \geqslant t+2$. To obtain the result for $2 \leqslant t \leqslant 14$ as well we have to go through the proof of [8] and do some modifications.

In view of $m(n+1, (t, 0, \ldots, 0)) \geqslant sm(n, (t, 0, \ldots, 0))$ if for some n one had $m(n, (t, 0, \ldots, 0)) > s^{n-t}$, then $p((t, 0, \ldots, 0)) > s^{-t}$ would follow. Thus, it is sufficient to show

$$p((t, 0, \ldots, 0)) \leqslant s^{-t} \tag{4.2}$$

To prove (4.2) we apply Proposition 2.4. First note that

$$\sum_{j > (1/s+\varepsilon)n} f_j(s-1)^{n-j} \leqslant \sum_{j > (1/s+\varepsilon)n} \binom{n}{j}(s-1)^{n-j} = o(s^n)$$

holds.

Thus to prove (4.2) we may suppose that $f_j = 0$ unless $j \leqslant (1/s + \varepsilon)n$.

For j satisfying $(j-t+1)(t+1) \leqslant n$ we may apply the exact form of the Erdős–Ko–Rado theorem (cf. [14]) to deduce

$$f_j \leqslant \binom{n-t}{j-t}. \tag{4.3}$$

If $s \geq t+2$ and $n > n_0(t)$, then (4.3) holds for all j and gives

$$m(n, (t, 0, \ldots, 0)) < \sum_j \binom{n-t}{j-t} s^{n-j} + o(s^n) = s^{n-t} + o(s^n)$$

as desired.

But in the case $s = t+1$ (4.3) does not necessarily hold for $n/s < j < n/s(1+\varepsilon)$. In this case we cannot just apply Wilson's result, but we have to use the following inequality, which follows from the actual proof.

$$f_j \leq (1+\delta)\binom{n-t}{j-t} \quad \text{holds for } j \leq \left(\frac{1}{s} + \varepsilon\right)n \quad \text{if } \varepsilon = \varepsilon(\delta) \tag{4.4}$$

is a sufficiently small positive constant.

Using this instead of (4.3) gives in the same way

$$m(n, (t, 0, \ldots, 0)) \leq (1 + \delta + o(1))s^{n-t}$$

for arbitrarily small δ, provided $n > n_0(\delta)$. This implies (4.2). □

Remark. Let us mention that we feel it is rather surprising that the exact result (Proposition 4.4) is deduced from an asymptotic result ((4.2)). In a sense this shows the strength of the Erdős–Ko–Rado Theorem. The original proof of [8] needed the condition $t \geq 15$ only because at that time the exact bound $(n \geq (k-t+1)(t+1))$ in the Erdős–Ko–Rado Theorem was known only for $t \geq 15$ (see [7]). That result has the advantage of showing that for $(k-t+1) > n/(t+1)$ but $k < 1.2n/(t+1) f_k \leq f_k(\mathcal{B}_1)$ holds (which is best possible), implying (4.4) in a stronger form. Actually, combining this result and Proposition 2.4 one can show that for $s = t \geq 15$

$$m(n, (t, 0, \ldots, 0)) = (s^2 + s - 1)s^{n-t-2}$$

holds.

References

[1] C. Bang, H. Sharp and P. Winkler, On families of finite sets with bounds on unions and intersections, Discrete Math. 45 (1983) 123–126.

[2] B. Cooper, unpublished manuscript, 1987.

[3] P.L. Erdős, P. Frankl and G.O.H. Katona, Extremal hypergraph problems and convex hulls, Combinatorica 5 (1985) 11–26.

[4] P. Erdős, C. Ko and R. Rado, Intersection theorems for systems of finite sets, Quart. J. Math. Oxford 12 (1961) 313–320.

[5] P. Frankl, Extremal set systems, Ph.D. Thesis, Budapest, 1976.

[6] P. Frankl, The proof of conjecture of G.O.H. Katona, J. Combin. Theory Ser. A 19 (1975) 208–219.

[7] P. Frankl, The Erdős–Ko–Rado Theorem is true for $n = ckt$, Coll. Math. Soc. J. Bolyai 18 (1978) 365–375.

[8] P. Frankl and Z. Füredi, The Erdős–Ko–Rado Theorem for integer sequences, SIAM J. Algebraic Discrete Methods 1 (1980) 376–381.
[9] G.O.H. Katona, Extremal problems for hypergraphs, in "Combinatorics" Vol. II. pp. 13–42 Math. Centre Tracts, Vol. 56 (Amsterdam, 1974).
[10] Gy. Katona, Intersection theorems for systems of finite sets, Acta Math. Hungar. 15 (1964) 329–337.
[11] D.J. Kleitman, Families of non-disjoint subsets, J. Combin. Theory Ser. A 1 (1966) 153–155.
[12] A. Moon, An analogue of the Erdős–Ko–Rado Theorem for the Hamming schemes, J. Combin. Theory Ser. A 32 (1982) 386–390.
[13] P. Winkler, personal communication, 1988.
[14] R.M. Wilson, The exact bound in the Erdős–Ko–Rado Theorem, Combinatorica 4 (1984) 247–257.