Cross-Intersecting Sets of Vectors

János Pach[*] Gábor Tardos[†]

Abstract

Given a sequence of positive integers $p = (p_1, \ldots, p_n)$, let S_p denote the set of all sequences of positive integers $x = (x_1, \ldots, x_n)$ such that $x_i \leq p_i$ for all i. Two families of sequences (or vectors), $A, B \subseteq S_p$, are said to be r-*cross-intersecting* if no matter how we select $x \in A$ and $y \in B$, there are at least r distinct indices i such that $x_i = y_i$. We show that for any pair of 1-cross-intersecting families, $A, B \subseteq S_p$, we have $|A| \cdot |B| \leq |S_p|^2/k^2$, where $k = \min_i p_i$. We also determine the minimal value of $|A| \cdot |B|$ for any pair of r-cross-intersecting families and characterize the extremal pairs for $r > 1$, provided that $k > r+1$. The case $k \leq r+1$ is quite different. We have a conjecture for this case, which we can verify under additional assumptions. Our results generalize and strengthen several previous results by Berge, Borg, Frankl, Füredi, Livingston, and Moon, and answers a question of Zhang.

1 Introduction

The Erdős-Ko-Rado theorem [EKR61] states that for $n \geq 2k$, every family of pairwise intersecting k-element subsets of an n-element set consists of at most $\binom{n-1}{k-1}$ subsets, as many as the star-like family of all subsets containing a fixed element of the underlying set. This was the starting point of a whole new area within combinatorics: extremal set theory; see [GK78], [Bol86], [DeF83], [F95]. The Erdős-Ko-Rado theorem has been extended and generalized to other structures: to multisets, divisors of an integer, subspaces of a vector space, sets of permutations, etc. It was also generalized to "cross-intersecting" families, i.e., to families A and B with the property that every element of A intersects all elements of B; see Hilton [Hi77], Moon [Mo82], and Pyber [Py86].

For any positive integer k, we write $[k]$ for the set $\{1, \ldots, k\}$. Given a sequence of positive integers $p = (p_1, \ldots, p_n)$, let

$$S_p = [p_1] \times \cdots \times [p_n] = \{(x_1, \ldots, x_n) \ : \ x_i \in [p_i] \text{ for } i \in [n]\}.$$

[*]EPFL, Lausanne and Rényi Institute, Budapest. Supported by NSF Grant CCF-08-30272, by OTKA under EUROGIGA projects GraDR and ComPoSe 10-EuroGIGA-OP-003, and by Swiss National Science Foundation Grants 200020-144531 and 20021-137574. Email: pach@cims.nyu.edu

[†]Rényi Institute, Budapest. Supported by the grant OTKA NN-102029, the NSERC Discovery grant and a visit at EPFL, Lausanne. Email: tardos@renyi.hu

We will refer to the elements of S_p as *vectors*. The *Hamming distance* between the vectors $x, y \in S_p$ is $|\{i \in [n] \ : \ x_i \neq y_i\}|$ and is denoted by $d(x, y)$. Let $r \geq 1$ be an integer. Two vectors $x, y \in S_p$ are said to be *r-intersecting* if $d(x, y) \leq n - r$. (This term originates in the observation that if we represent a vector $x = (x_1, \dots, x_n) \in S_p$ by the set $\{(i, x_i) \ : \ i \in [n]\}$, then x and $y \in S_p$ are r-intersecting if and only if the sets representing them have at least r common elements.) Two families $A, B \subseteq S_p$ are *r-cross-intersecting*, if every pair $x \in A$, $y \in B$ is r-intersecting. If (A, A) is an r-cross-intersecting pair, we say A is r-intersecting. We simply say *intersecting* or *cross-intersecting* to mean 1-intersecting or 1-cross-intersecting, respectively.

The investigation of the maximum value for $|A| \cdot |B|$ for cross-intersecting pairs of families $A, B \subseteq S_p$ was initiated by Moon [Mo82]. She proved, using a clever induction argument, that in the special case when $p_1 = p_2 = \cdots = p_n = k$ for some $k \geq 3$, every cross-intersecting pair $A, B \subseteq S_p$ satisfies

$$|A| \cdot |B| \leq k^{2n-2},$$

with equality if and only if $A = B = \{x \in S_p \ : \ x_i = j\}$, for some $i \in [n]$ and $j \in [k]$. In the case $A = B$, Moon's theorem had been discovered by Berge [Be74], Livingston [Liv79], and Borg [Bo08]. See also Stanton [St80]. In his report on Livingston's paper, published in the *Mathematical Reviews*, Kleitman gave an extremely short proof for the case $A = B$, based on a shifting argument. Zhang [Zh13] established a somewhat weaker result, using a generalization of Katona's circle method [K72]. Note that for $k = 2$, we can take $A = B$ to be any family of 2^{n-1} vectors without containing a pair $(x_1, \dots, x_n), (y_1, \dots, y_n)$ with $x_i + y_i = 3$ for every i. Then A is an intersecting family with $|A|^2 = 2^{2n-2}$, which is not of the type described in Moon's theorem.

Moon also considered r-cross-intersecting pairs in S_p with $p_1 = p_2 = \cdots = p_n = k$ for some $k > r + 1$, and characterized all pairs for which $|A| \cdot |B|$ attains its maximum, that is, we have

$$|A| \cdot |B| = k^{2(n-r)}.$$

The assumption $k > r + 1$ is necessary.

Zhang [Zh13] suggested that Moon's results may be extended to arbitrary sequences of positive integers $p = (p_1, \dots, p_n)$. The aim of this note is twofold: (1) to establish such an extension under the assumption $\min_i p_i > r + 1$, and (2) to formulate a conjecture that covers essentially all other interesting cases. We verify this conjecture in several special cases.

Theorem 1. *Let $p = (p_1, \dots, p_n)$ be a sequence integers and let $A, B \subseteq S_p$ be two cross-intersecting families of vectors.*

We have $|A| \cdot |B| \leq |S_p|^2/k^2$, where $k = \min_i p_i$. Equality holds for the case $A = B = \{x \in S_p \ : \ x_i = j\}$, whenever $i \in [n]$ satisfies $p_i = k$ and $j \in [k]$. For $k \neq 2$, there are no other extremal cross-intersecting families.

We say that a coordinate $i \in [n]$ is *irrelevant* for a set $A \subseteq S_p$ if, whenever two elements of S_p differ only in coordinate i and A contains one of them, it also contains the other. Otherwise, we say that i is *relevant* for A.

Note that no coordinate i with $p_i = 1$ can be relevant for any set. Each such coordinate forces an intersection between every pair of vectors. So, if we delete it, every r-cross-intersecting pair becomes $(r-1)$-cross-intersecting. Therefore, from now on we will always assume that we have $p_i \geq 2$ for every i.

We call a sequence of integers $p = (p_1, \dots, p_n)$ a *size vector* if $p_i \geq 2$ for all i. The *length* of p is n. We say that an r-cross-intersecting pair $A, B \subseteq S_p$ *maximal* if it maximizes the value $|A| \cdot |B|$.

Using this notation and terminology, Theorem 1 can be rephrased as follows.

Theorem 1'. *Let $p = (p_1, \dots, p_n)$ be a sequence positive integers with $k = \min_i p_i > 2$.*

For any maximal pair of cross-intersecting families, $A, B \subseteq S_p$, we have $A = B$, and there is a single coordinate which is relevant for A. The relevant coordinate i must satisfy $p_i = k$.

See Section 5 for a complete characterization of maximal cross-intersecting pairs in the $k = 2$ case. Here we mention that only the coordinates with $p_i = 2$ can be relevant for them, but for certain pairs, *all* such coordinates are relevant simultaneously. For example, let n be odd, $p = (2, \dots, 2)$, and let $A = B$ consist of all vectors in S_p which have at most $\lfloor n/2 \rfloor$ coordinates that are 1. This makes (A, B) a maximal cross-intersecting pair.

Let $H \subseteq [n]$ be a subset of the coordinates, let $x_0 \in S_p$ be an arbitrary vector, and let k be an integer satisfying $0 \leq k \leq |H|$. The *Hamming ball* of radius k around x_0 in the coordinates H is defined as the set

$$B_k = \{x \in S_p \ : \ |\{i \in H \ : \ x_i \neq (x_0)_i\}| \leq k\}.$$

Note that the pair (B_k, B_l) is $(|H| - k - l)$-cross-intersecting. We use the word *ball* to refer to any Hamming ball without specifying its center, radius or its set of coordinates. A Hamming ball of radius 0 in coordinates H is said to be obtained by *fixing* the coordinates in H.

For the proof of Theorem 1, we need the following statement, which will be established by induction on n, using the idea in [Mo82].

Lemma 2. *Let $1 \leq r < n$, let $p = (p_1, \dots, p_n)$ be a size vector satisfying $3 \leq p_1 \leq p_2 \leq \dots \leq p_n$ and let $A, B \subseteq S_p$ form a pair of r-cross-intersecting families. If*

$$\frac{2}{p_{r+1}} + \sum_{i=1}^{r} \frac{1}{p_i} \leq 1,$$

then $|A| \cdot |B| \leq \prod_{i=r+1}^{n} p_i^2$. *In case of equality, we have* $A = B$ *and this set can be obtained by fixing* r *coordinates in* S_p.

By fixing any r coordinates, we obtain an r-intersecting family with r relevant coordinates. As was observed by Frankl and Füredi [FF80], not all maximal r-intersecting families can be constructed in this way. For instance, a Hamming ball of radius 1 in $r+2$ coordinates is r-intersecting, it has $r+2$ relevant coordinates and, depending on the vector p, it may have strictly more elements than the set obtained by fixing r coordinates. We have the following general conjecture.

Conjecture 3. *Let* $2 \leq r \leq n$ *and let* p *be a size vector of length* n. *If* $A, B \subseteq S_p$ *form a maximal pair of* r*-cross-intersecting families, then they are balls.*

It is not hard to narrow down the range of possibilities for maximal r-cross-intersecting pairs that are formed by two balls, A and B. In fact, the following simple lemma implies that all such pairs are determined up to isomorphism by the radii of A and B. Assuming that Conjecture 3 is true, finding all maximal pairs of r-cross-intersecting families in S_p boils down to making numeric comparisons for pairs of balls obtained by all possible radii.

Lemma 4. *Let* $1 \leq r \leq n$ *and let* $p = (p_1, \ldots, p_n)$ *be a size vector. If* $A, B \subseteq S_p$ *form a maximal pair of* r*-cross-intersecting sets, then either of them determines the other. In particular,* A *and* B *have the same set of relevant coordinates. Moreover, if* A *is a ball of radius* l *around* $x_0 \in S_p$ *in a set of coordinates* $H \subseteq [n]$, *then* $|H| \geq l + r$, B *is a ball of radius* $|H| - l - r$ *around* x_0 *in the same set of the coordinates, and we have* $p_i \leq p_j$ *for* $i \in H$ *and* $j \in [n] \setminus H$.

We cannot prove Conjecture 3 in its full generality, but we can prove it in several interesting special cases. We will proceed in two steps. First we argue, using *entropies*, that the number of relevant coordinates in a maximal r-cross-intersecting family is bounded. Then we apply combinatorial methods to prove the conjecture under the assumption that the number of relevant coordinates is small.

Using entropies, we can show that neither of the families in a maximal cross-intersecting pair can have arbitrarily many elements.

Theorem 5. *Let* $1 \leq r \leq n$, *let* $p = (p_1, \ldots, p_n)$ *be a size vector, let* $A, B \subseteq S_p$ *form a maximal pair of* r*-cross-intersecting families and let* H *be the set of coordinates that are relevant for* A *or* B. *Then neither the size of* A *nor the size of* B *can exceed*

$$\frac{|S_p|}{\prod_{i \in H} (p_i - 1)^{1-2/p_i}}.$$

We use this theorem to bound the number of relevant coordinates i with $p_i > 2$. The number of relevant coordinates i with $p_i = 2$ can be unbounded, provided that $r = 1$

(in which case, the statement of the conjecture is false.) However, we believe that for $r > 1$, that *total* number of relevant coordinates is also bounded from above.

.

Theorem 6. *Let $1 \leq r \leq n$, let $p = (p_1, \ldots, p_n)$ be a size vector and let $A, B \subseteq S_p$ form a maximal pair of r-cross-intersecting families.*

For the set of coordinates H relevant for A or B, we have

$$\prod_{i=1}^{r} p_i \geq \prod_{i \in H} (p_i - 1)^{1-2/p_i},$$

which implies that $|\{i \in H \,:\, p_i > 2\}| < 5r$.

We can characterize the maximal r-cross-intersecting pairs for all size vectors p satisfying $\min p_i > r + 1$, and in many other cases.

Theorem 7. *Let $2 \leq r \leq n$, let $p = (p_1, \ldots, p_n)$ be a size vector with $p_1 \leq p_2 \leq \cdots \leq p_n$ and let $A, B \subseteq S_p$ form a pair of r-cross-intersecting families.*

1. *If $p_1 > r + 1$, we have $|A| \cdot |B| \leq \prod_{i=r+1}^{n} p_i^2$. In case of equality, $A = B$ holds and this set can be obtained by fixing r coordinates in S_p.*
2. *If $p_1 = r + 1 > 7$, we have $|A| \cdot |B| \leq \prod_{i=r+1}^{n} p_i^2$. In case of equality, $A = B$ holds and this set can be obtained either by fixing r coordinates in S_p or by taking a Hamming ball of radius 1 in $r + 2$ coordinates, all satisfying $p_i = r + 1$.*
3. *If $p_1 \geq 2r/3 + t(r)$ and (A, B) is a maximal r-cross-intersecting pair, then the sets A and B are balls of radius 0 or 1 in at most $r + 2$ coordinates. Here the function $t(r)$ satisfies $t(r) = o(r)$.*

The proof of Theorem 7 relies on the following result.

Theorem 8. *Let $2 \leq r \leq n$ and let p be a size vector. If $A, B \subseteq S_p$ is a maximal pair of r-cross-intersecting families and at most $r + 2$ coordinates are relevant for them, then A and B are balls of radius 0 or 1.*

With an involved case analysis, Theorem 8 can be extended to pairs with $r + 4$ relevant coordinates (or possibly even further). Any such an improvement carries over to Theorem 7.

All of our results remain meaningful in the symmetric case where $A = B$. For instance, in this case, Theorem 1 states that every intersecting family $A \subseteq S_p$ has at most $|S_p|/k$ members, where $k = \min_i p_i$. In case $k > 2$, equality can be achieved only by fixing some coordinate i with $p_i = k$.

2 Proof of Theorem 1

The aim of this section is to establish Theorem 1. First, we verify Lemma 4 and another technical lemma (see Lemma 9 below), which generalizes the corresponding result in [Mo82]. Our proof is slightly simpler. Lemma 9 will enable us to deduce Lemma 2, the main ingredient of the proof of Theorem 1, presented at the end of the section.

Proof of Lemma 4. The first statement is self evident: if $A, B \subseteq S_p$ form a maximal r-cross-intersecting, then

$$B = \{x \in S_p \ : \ x \ r\text{-intersects } y \text{ for all } y \in A\}.$$

If a coordinate is irrelevant for A, then it is also irrelevant for B defined by this formula. Therefore, by symmetry, A and B have the same set of relevant coordinates.

If A is the Hamming ball around x_0 of radius l in coordinates H, then we have $B = \emptyset$ if $|H| < l+r$, which is not possible for a maximal cross-intersecting family. If $|H| \geq l+r$, we obtain the ball claimed in the lemma. For every $i \in H$, $j \in [n] \setminus H$, consider the set $H' = (H \setminus \{i\}) \cup \{j\}$ and the Hamming balls A' and B' of radii l and $|H| - l - r$ around x_0 in the coordinates H'. These balls form an r-cross-intersecting pair and in case $p_i > p_j$, we have $|A'| > |A|$ and $|B'| > |B|$, contradicting the maximality of the pair (A, B) pair. □

The following lemma will also be used in the proof of Theorem 5, presented in the next section.

Lemma 9. *Let $1 \leq r \leq n$, let $p = (p_1, \ldots, p_n)$ be a size vector, and let $A, B \subseteq S_p$ form a maximal pair of r-cross-intersecting families.*

If $i \in [n]$ is a relevant coordinate for A or B, then there exists a value $l \in [p_i]$ such that

$$|\{x \in A \ : \ x_i \neq l\}| \leq |A|/p_i,$$
$$|\{y \in B \ : \ y_i \neq l\}| \leq |B|/p_i.$$

Proof. Let us fix r, n, p, i, A and B as in the lemma. By Lemma 4, if a coordinate is irrelevant for A, then it is also irrelevant for B and vice versa.

In the case $n = r$, we have $A = B$ and this set must be a singleton, so that the lemma is trivially true. From now on, we assume that $n > r$ and hence the notion of r-cross-intersecting families is meaningful for $n - 1$ coordinates.

Let $q = (p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n)$. For any $l \in [p_i]$, let

$$A'_l = \{x \in A \ : \ x_i = l\},$$

$$B'_l = \{y \in B \ : \ y_i = l\},$$

and let A_l and B_l stand for the sets obtained from A'_l and B'_l, respectively, by dropping their ith coordinates. By definition, we have $A_l, B_l \subseteq S_q$, and $|A| = \sum_l |A_l|$ and $|B| = \sum_l |B_l|$. Furthermore, for any two distinct elements $l, m \in [p_i]$, the families A_l and B_m are r-cross-intersecting, since the vectors in A'_l differ from the vectors in B'_l in the ith coordinate, and therefore the r indices where they agree must be elsewhere.

Let Z denote the maximum product $|A^*|\cdot|B^*|$ of an r-cross-intersecting pair $A^*, B^* \subseteq S_q$. We have $|A_l| \cdot |B_m| \leq Z$ for all $l, m \in [p_i]$ with $l \neq m$. Adding an irrelevant ith coordinate to the maximal r-cross-intersecting pair $A^*, B^* \subseteq S_q$, we obtain a pair $A^{*\prime}, B^{*\prime} \subseteq S_p$ with $|A^{*\prime}| \cdot |B^{*\prime}| = p_i^2 Z$. Thus, using the maximality of A and B, we have $|A| \cdot |B| \geq p_i^2 Z$. Let l_0 be chosen so as to maximize $|A_{l_0}| \cdot |B_{l_0}|$, and let $c = |A_{l_0}| \cdot |B_{l_0}|/Z$.

Assume first that $c \leq 1$. Then we have

$$p_i^2 Z \leq |A| \cdot |B| = \sum_{l,m \in [p_i]} |A_l| \cdot |B_m| \leq \sum_{l,m \in [p_i]} Z = p_i^2 Z.$$

Hence, we must have equality everywhere. This yields that $c = 1$ and that A_l and B_m form a maximal r-intersecting pair for all $l, m \in [p_i]$, $l \neq m$. This also implies that $|A_l| = |A_m|$ for $l, m \in [p_i]$, from where the statement of the lemma follows, provided that $p_i = 2$.

If $p_i \geq 3$, then all sets A_l must be equal to one another, since one member in a maximal r-cross-intersecting family determines the other. This contradicts our assumption that the ith coordinate was relevant for A.

Thus, we may assume that $c > 1$.

For $m \in [p_i]$, $m \neq l_0$, we have $|A_{l_0}| \cdot |B_m| \leq Z = |A_{l_0}| \cdot |B_{l_0}|/c$. Thus,

$$|B_m| \leq |B_{l_0}|/c, \tag{1}$$

which yields that $|B| = \sum_{m \in [p_i]} |B_m| \leq (1 + (p_i - 1)/c)|B_{l_0}|$. By symmetry, we also have

$$|A_m| \leq |A_{l_0}|/c \tag{2}$$

for $m \neq l_0$ and $|A| \leq (1 + (p_i - 1)/c)|A_{l_0}|$. Combining these inequalities, we obtain

$$p_i^2 Z \leq |A| \cdot |B| \leq (1 + (p_n - 1)/c)^2 |A_{l_0}| \cdot |B_{l_0}| = (1 + (p_i - 1)/c)^2 cZ.$$

We solve the resulting inequality $p_i^2 \leq c(1 + (p_i - 1)/c)^2$ for $c > 1$ and conclude that $c \geq (p_i - 1)^2$. This inequality, together with Equations (1) and (2), completes the proof of Lemma 9. □

Proof of Lemma 2. We proceed by induction on n.

Let A and B form a maximal r-cross-intersecting pair. It is sufficient to show that they have only r relevant coordinates. Let us suppose that the set H of their relevant coordinates satisfies $|H| > r$, and choose a subset $H' \subseteq H$ with $|H'| = r+1$. By Lemma 9, for every $i \in H$ there exists $l_i \in [p_i]$ such that the set $X_i = \{x \in B \ : \ x_i \neq l_i\}$ has cardinality $|X_i| \leq |B|/p_i$.

If we assume that

$$\frac{2}{p_{r+1}} + \sum_{i=1}^{r} \frac{1}{p_i} < 1$$

holds (with strict inequality), then this bound of $|X_i|$ would suffice. In order to be able to deal with the case

$$\frac{2}{p_{r+1}} + \sum_{i=1}^{r} \frac{1}{p_i} = 1,$$

we show that $|X_i| = |B|/p_i$ is not possible. Considering the proof of Lemma 9, equality here would mean that the sets $A_l = \{x \in A \ : \ x_i = l\}$ and $B_l = \{y \in B \ : \ y_i = l\}$ satisfy the following condition: by dropping the ith coordinate from the pairs (A_{l_i}, B_m) and (A_m, B_{l_i}), we obtain maximal r-cross-intersecting pairs for $m \neq l_i$. By the induction hypothesis, this would imply that $A_{l_i} = B_m$ and $A_m = B_{l_i}$, contradicting that $|A_m| < |A_{l_i}|$ and $|B_m| < |B_{l_i}|$ (see (1), in view of $c > 1$). Therefore, we have $|X_i| < |B|/p_i$.

Let $C = \{x \in S_p \ : \ x_i = 1 \text{ for all } i \in [r]\}$ be the r-intersecting set obtained by fixing r coordinates in S_p. In the set $D = B \setminus (\bigcup_{i \in H'} X_i)$, the coordinates in H' are fixed. Thus, we have

$$|D| \leq \prod_{i \in [n] \setminus H'} p_i \leq \prod_{i=r+2}^{n} p_i = |C|/p_{r+1}.$$

On the other hand, we have

$$|D| = |B| - \sum_{i \in H'} |X_i| > |B|(1 - \sum_{i \in H'} 1/p_i) \geq |B|(1 - \sum_{i=1}^{r+1} 1/p_i).$$

Comparing the last two inequalities, we obtain

$$|B| < \frac{|C|}{p_{r+1}(1 - \sum_{i=1}^{r+1} 1/p_i)}.$$

By our assumption on p, the denominator is at least 1, so that we have $|B| < |C|$. By symmetry, we also have $|A| < |C|$. Thus, $|A| \cdot |B| < |C|^2$ contradicting the maximality of the pair (A, B). This completes the proof of Lemma 2. $\square$

Now we can quickly finish the proof of Theorem 1.

Proof of Theorem 1. Notice that Lemma 2 implies Theorem 1, whenever $k = \min_i p_i \geq 3$. It remains to verify the statement for $k = 1$ and $k = 2$. For $k = 1$, it follows from the fact that all pairs of vectors in S_p are intersecting, thus the only maximal cross-intersecting pair is $A = B = S_p$.

Suppose next that $k = 2$. For $x \in S_p$, let $x' \in S_p$ be defined by $x'_i = (x_i + 1 \bmod p_i)$ for $i \in [n]$. Note that $x \mapsto x'$ is a permutation of S_p. Clearly, x and x' are not intersecting, so we either have $x \notin A$ or $x' \notin B$. As a consequence, we obtain that $|A| + |B| \leq |S_p|$, which, in turn, implies that $|A| \cdot |B| \leq |S_p|^2/4$, as claimed. It also follows that all maximal pairs satisfy $|A| = |B| = |S_p|/2$. □

3 Using entropy: Proofs of Theorems 5 and 6

Proof of Theorem 5. Let r, n, p, A, B and H be as in the theorem. Let us write y for a randomly and uniformly selected element of B. Lemma 9 implies that, for each $i \in H$, there exists a value $l_i \in [p_i]$ such that

$$Pr[y_i = l_i] \geq 1 - 1/p_i. \tag{3}$$

We bound the *entropy* $H(y_i)$ of y_i from above by the entropy of the indicator variable of the event $y_i = l_i$ plus the contribution coming from the entropy of y_i assuming $y_i \neq l_i$:

$$H(y_i) \leq h(1 - 1/p_i) + (1/p_i)\log(p_i - 1) = \log p_i - (1 - 2/p_i)\log(p_i - 1),$$

where $h(z) = -z\log z - (1 - z)\log(1 - z)$ is the entropy function, and we used that $1 - 1/p_i \geq 1/2$.

For any $i \in [n] \setminus H$, we use the trivial estimate $H(y_i) \leq \log p_i$. By the subadditivity of the entropy, we obtain

$$\log|B| = H(y) \leq \sum_{i\in[n]} H(y_i) \leq \sum_{i\in H}(\log p_i - (1 - 2/p_i)\log(p_i - 1)) + \sum_{i\in[n]\setminus H}\log p_i,$$

or, equivalently,

$$|B| \leq \prod_{i\in H}\frac{p_i}{(p_i - 1)^{1-2/p_i}}\prod_{i\in[n]\setminus H} p_i = \frac{|S_p|}{\prod_{i\in H}(p_i - 1)^{1-2/p_i}}$$

as required. The bound on $|A|$ follows by symmetry and completes the proof of the theorem. □

Theorem 6 is a simple corollary of Theorem 5.

Proof of Theorem 6. Fixing the first r coordinates, we obtain the set

$$C = \{x \in S_p \ : \ x_i = 1 \text{ for all } i \in [r]\}.$$

This set is r-intersecting. Thus, by the maximality of the pair (A, B), we have

$$|A| \cdot |B| \geq |C|^2 = (\prod_{i=r+1}^{n} p_i)^2. \tag{4}$$

Comparing this with our upper bounds on $|A|$ and $|B|$, we obtain the first inequality claimed in the theorem.

To prove the required bound on the number of relevant coordinates i with $p_i \neq 2$, we assume that the coordinates are ordered, that is $p_1 \leq p_2 \leq \cdots \leq p_n$. Applying the above estimate on $\prod_{i\in[r]} p_i$ and using $(p_i - 1)^{1-2/p_i} > p_i^{1/5}$ whenever $p_i \geq 3$, the theorem follows. $\square$

4 Monotone families: Proofs of Theorems 8 and 7

Given a vector $x \in S_p$, the set $\mathrm{supp}(x) = \{i \in [n] : x_i > 1\}$ is called the *support* of x. A family $A \subseteq S_p$ is said to be *monotone*, if for any $x \in A$ and $y \in S_p$ satisfying $\mathrm{supp}(y) \subseteq \mathrm{supp}(x)$, we have $y \in A$.

For a family $A \subseteq S_p$, let us define its *support* as $\mathrm{supp}(A) = \{\mathrm{supp}(x) : x \in A\}$. For a monotone family A, its support is clearly subset-closed and it uniquely determines A, as $A = \{x \in S_p : \mathrm{supp}(x) \in \mathrm{supp}(A)\}$.

The next result shows that if we want to prove Conjecture 3, it is sufficient to prove it for monotone families. This will enable us to establish Theorems 8 and 7, that is, to verify the conjecture for maximal r-cross-intersecting pairs with a limited number of relevant coordinates.

Lemma 10. *Let $1 \leq r \leq n$ and let p be a size vector of length n.*

There exists a maximal pair of r-cross-intersecting families $A, B \subseteq S_p$ such that both A and B are monotone.

If $r \geq 2$ and $A, B \subseteq S_p$ are maximal r-cross-intersecting families that are not *balls, then there exists a pair of maximal r-cross-intersecting families that consists of monotone sets that are not balls and have no more relevant coordinates than A or B.*

Proof of Lemma 10. Consider the following *shift operations.* For any $i \in [n]$ and $j \in [p_i] \setminus \{1\}$, for any family $A \subseteq S_p$ and any element $x \in A$, we define

$$\phi_i(x) = (x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n),$$
$$\phi_{i,j}(x, A) = \begin{cases} \phi_i(x) & \text{if } x_i = j \text{ and } \phi_i(x) \notin A \\ x & \text{otherwise,} \end{cases}$$
$$\phi_{i,j}(A) = \{\phi_{i,j}(x, A) : x \in A\}.$$

Clearly, we have $|\phi_{i,j}(A)| = |A|$ for any set $A \subseteq S_p$. We claim that for any pair of r-cross-intersecting sets $A, B \subseteq S_p$, the sets $\phi_{i,j}(A)$ and $\phi_{i,j}(B)$ are also r-cross-intersecting. Indeed, if $x \in A$ and $y \in B$ are r-intersecting vectors, then $\phi_{i,j}(x, A)$ and $\phi_{i,j}(y, B)$ are also r-intersecting, unless x and y have exactly r common coordinates, one of them is $x_i = y_i = j$, and this common coordinate gets ruined as $\phi_{i,j}(x, A) = x$ and $\phi_{i,j}(y, B) = \phi_i(y)$ (or vice versa). However, this is impossible, because this would imply that the vector $\phi_i(x)$ belongs to A, in spite of the fact that $\phi_i(x)$ and $y \in B$ are not r-intersecting.

If (A, B) is a *maximal* r-cross-intersecting pair, then so is $(\phi_{i,j}(A), \phi_{i,j}(B))$. When applying one of these shift operations does change the sets A or B, then the total sum of all coordinates of all elements decreases. Therefore, after shifting a finite number of times we arrive at a maximal pair of r-intersecting families that cannot be changed by further shifting. We claim that this pair (A, B) is monotone. Let $y \in B$ and $y' \in S_p \setminus B$ be arbitrary. We show that B is monotone by showing that $\mathrm{supp}(y')$ is not contained in $\mathrm{supp}(y)$. Indeed, by the maximality of the pair (A, B) and using the fact that $y' \notin B$, there must exist $x' \in A$ such that x' and y' are not r-cross-intersecting, and hence $|\mathrm{supp}(x') \cup \mathrm{supp}(y')| > n - r$. Applying "projections" ϕ_i to x' in the coordinates $i \in \mathrm{supp}(x') \cap \mathrm{supp}(y)$, we obtain x with $\mathrm{supp}(x) = \mathrm{supp}(x') \setminus \mathrm{supp}(y)$. The shift operations $\phi_{i,j}$ do not change the set A, thus A must be closed for the projections ϕ_i and we have $x \in A$. The supports of x and y are disjoint. Thus, their Hamming distance is $|\mathrm{supp}(x) \cup \mathrm{supp}(y)|$, which is at most $n - r$, as they are r-intersecting. Therefore, $\mathrm{supp}(x) \cup \mathrm{supp}(y) = \mathrm{supp}(x') \cup \mathrm{supp}(y)$ is smaller than $\mathrm{supp}(x') \cup \mathrm{supp}(y')$, showing that $\mathrm{supp}(y') \not\subseteq \mathrm{supp}(y)$. This proves that B is monotone. By symmetry, A is also monotone, which proves the first claim of the lemma.

To prove the second claim, assume that $r \geq 2$ and let $A, B \subseteq S_p$ form a maximal r-cross-intersecting pair. By the previous paragraph, this pair can be transformed into a monotone pair by repeated applications of the shift operations $\phi_{i,j}$. Clearly, these operations do not introduce new relevant coordinates. It remains to check that the shifting operations do not produce balls from non-balls, that is, if $A, B \subseteq S_p$ are maximal r-cross-intersecting families, and $A' = \phi_{i,j}(A)$ and $B' = \phi_{i,j}(B)$ are balls, then so are A and B. In fact, by Lemma 4 it is sufficient to prove that one of them is a ball.

We saw that A' and B' must also form a maximal r-cross-intersecting pair. Thus, by Lemma 4, there is a set of coordinates $H \subseteq [n]$, a vector $x_0 \in S_p$, and radii l and m satisfying $|H| = r + l + m$ and that A' and B' are the Hamming balls of radius l and m in coordinates H around the vector x_0. We can assume that $i \in H$, because otherwise $A = A'$ and we are done. We also have that $(x_0)_i = 1$, otherwise $A' = \phi_{i,j}(A)$ is impossible. The vectors $x \in S_p$ such that $x_i = j$ and

$$|\{k \in H : x_k \neq (x_0)_k\}| = l + 1$$

are called A-*critical.* Analogously, the vectors $y \in S_p$ such that $y_i = j$ and

$$|\{k \in H \,:\, y_k \neq (x_0)_k\}| = m + 1$$

are said to be B-*critical.* By the definition of $\phi_{i,j}$, the set A differs from A' by including some A-critical vectors x and losing the corresponding vectors $\phi_i(x)$. Symmetrically, $B \setminus B'$ consists of some B-critical vectors y and $B' \setminus B$ consists of the corresponding vectors $\phi_i(y)$. Let us consider the bipartite graph G whose vertices on one side are the A-critical vectors x, the vertices on the other side are the B-critical vectors y (considered as disjoint sets, even if $l = m$), and x is adjacent to y if and only if $|\{k \in [n] \,:\, x_k = y_k\}| = r$. If x and y are adjacent, then neither the pair $(x, \phi_i(y))$, nor the pair $(\phi_i(x), y)$ is r-intersecting. As A and B are r-cross-intersecting, for any pair of adjacent vertices x and y of G, we have $x \in A$ if and only if $y \in B$.

The crucial observation is that, since $r > 1$, the graph G is connected and thus either all A-critical vectors belong to A or none of them does. In the latter case, we have $A = A'$, in the former case A is the Hamming ball of radius l in coordinates H around the vector x'_0, where x'_0 agrees with x_0 in all coordinates but in $(x'_0)_i = j$. In either case, A is a ball as required. □

Proof of Theorem 8. By Lemma 10, it is enough to restrict our attention to monotone sets A and B. We may also assume that all coordinates are relevant (simply drop the irrelevant coordinates), and thus we have $n \leq r + 2$.

We denote by U_l the Hamming ball of radius l around the all-1 vector in the entire set of coordinates $[n]$. Notice that the monotone sets A and B are r-cross-intersecting if and only if for $a \in \mathrm{supp}(A)$ and $b \in \mathrm{supp}(B)$ we have $|a \cup b| \leq n - r$. We consider all possible values of $n - r$, separately.

If $n = r$, both sets A and B must coincide with the singleton set U_0.

If $n = r + 1$, it is still true that either A or B is U_0. Otherwise both $\mathrm{supp}(A)$ and $\mathrm{supp}(B)$ have to contain at least one non-empty set, and these sets has to be the same singleton set (otherwise their union has more than $n - r = 1$ elements). So we have $\mathrm{supp}(A) = \mathrm{supp}(B) = \{\emptyset, \{i\}\}$ for some $i \in [n]$, but this contradicts our assumption that the coordinate i is relevant for A.

If $n = r + 2$, we are done if $A = B = U_1$. Otherwise, we must have a two-element set either in $\mathrm{supp}(A)$ or in $\mathrm{supp}(B)$. Let us assume that a two-element set $\{i, j\}$ belongs to $\mathrm{supp}(A)$. Then each set $b \in \mathrm{supp}(B)$ must satisfy $b \subseteq \{i, j\}$. This leaves five possibilities for a non-empty monotone set B, as $\mathrm{supp}(B)$ must be one of the following sets:

1. $\{\emptyset\}$,
2. $\{\emptyset, \{i\}\}$,
3. $\{\emptyset, \{j\}\}$,

4. $\{\emptyset, \{i\}, \{j\}\}$, and

5. $\{\emptyset, \{i\}, \{j\}, \{i,j\}\}$.

Cases 2, 3, and 5 are not possible, because either i or j would not be relevant for B.

In case 1, we have $B = U_0$, and thus $A = U_2$. In this case, A and B are balls, but the radius of A is 2. This is impossible, as U_1 is r-intersecting and $|U_1|^2 > |U_0| \cdot |U_2|$ always holds, so (A, B) is not maximal.

It remains to deal with case 4. Here $\mathrm{supp}(A)$ consists of the sets of size at most 1 and the two-element set $\{i,j\}$. Define

$$C = \{x \in S_p \ : \ x_k = 1 \text{ for all } k \in [n] \setminus \{i,j\}\}.$$

Note that $|A| + |B| = |U_1| + |C|$, because each vector in S_p appears in the same number of sets on both sides. Thus, we have either $|A| + |B| \leq 2|U_1|$ or $|A| + |B| \leq 2|C|$. Since $|A| > |B|$, the above inequalities imply $|A| \cdot |B| < |U_1|^2$ or $|A| \cdot |B| < |C|^2$. This contradicts the maximality of the pair (A, B), because both U_1 and C are r-cross-intersecting. The contradiction completes the proof of Theorem 8. □

Now we can prove our main theorem, verifying Conjecture 3 in several special cases.

Proof of Theorem 7. The statement about the case $p_1 > r + 1$ readily follows from Lemma 2, as in this case the condition

$$\frac{2}{p_{r+1}} + \sum_{i=1}^{r} \frac{1}{p_i} \leq 1$$

holds.

We can assume that A and B form a maximal r-cross-intersecting pair. We also assume without loss of generality that all coordinates are relevant for both sets (simply drop the irrelevant coordinates).

By Theorem 6, we have $\prod_{i=1}^{r} p_i \geq \prod_{i=1}^{n} (p_i - 1)^{1-2/p_i}$, and thus

$$\prod_{i=1}^{r} \frac{p_i}{(p_i - 1)^{1-2/p_i}} \geq \prod_{i=r+1}^{n} (p_i - 1)^{1-2/p_i}.$$

Here the function $x/(x-1)^{1-2/x}$ is decreasing for $x \geq 3$, while $(x-1)^{1-2/x}$ is increasing, and we have $p_i \geq p_1 \geq 3$. Therefore, we also have

$$\prod_{i=1}^{r} \frac{p_1}{(p_1 - 1)^{1-2/p_1}} \geq \prod_{i=r+1}^{n} (p_1 - 1)^{1-2/p_1},$$

$$p_1^r \geq (p_1 - 1)^{n(1-2/p_1)},$$

$$n \leq \frac{r \log p_1}{(1-2/p_1)\log(p_1-1)}.$$

It can be shown by simple computation that the right-hand side of the last inequality is strictly smaller than $r+3$ if $p_1 \leq 2r/3 + t(r)$ for some function $t(r) = O(r/\log r)$ and, in particular, for $p_1 = r+1 \geq 8$. In this case, we have $n \leq r+2$ relevant coordinates. Thus, Theorem 8 applies, yielding that A and B are balls. This proves the last statement of Theorem 7.

For the middle statement, we use Lemma 4 to calculate the sizes of A in B in the three possible cases. The product $|A| \cdot |B|$ is $z_1 = \prod_{i=r+1}^{n} p_i^2$ if A and B are balls of radius 0. The same product is $z_2 = (\sum_{i=1}^{r+1} p_i - r) \prod_{i=r+2}^{n} p_i^2$ if one of them is a ball of radius 0 while the other is a ball of radius 1. Finally, the product is $z_3 = (\sum_{i=1}^{r+2} p_i - r - 1)^2 \prod_{i=r+3}^{n} p_i^2$ if both sets are balls of radius 1. Note that we have $A = B$ in the first and third cases. Using the condition $p_i \geq r+1$, it is easy to verify that $z_2 < z_1$ and $z_3 \leq z_1$. Furthermore, we have $z_3 = z_1$ if and only if $p_i = r+1$ for all $i \in [r+2]$. This completes the proof of Theorem 7. □

5 Concluding remarks

For the simple characterization of the cases of equality in Theorem 1, the assumption $k \neq 2$ is necessary. Here we characterize the maximal cross-intersecting pairs in the $k = 2$ case.

Let $p = (p_1, \ldots, p_n)$ be a size vector of positive integers with $k = \min_i p_i = 2$ and let $I = \{i \in [n] \ : \ p_i = 2\}$. For any set W of functions from $I \to [2]$, define the sets

$$A_W = \{x \in S_p \ : \ \exists f \in W \text{ such that } x_i = f(i) \text{ for every } i \in I\},$$

$$B_W = \{y \in S_p \ : \ \nexists f \in W \text{ such that } y_i \neq f(i) \text{ for every } i \in I\}.$$

The sets A_W and B_W are cross-intersecting for any W. Moreover, if $|W| = 2^{I-1}$, we have $|A| \cdot |B| = |S_p|^2/4$, so they form a maximal cross-intersecting pair. Note that these include more examples than just the pairs of families described in Theorem 1, provided that $|I| > 1$.

We claim that all maximal cross-intersecting pairs are of the form constructed above. To see this, consider a maximal pair $A, B \subseteq S_p$. We know from the proof of Theorem 1 that $x \in A$ if and only if $x' \notin B$, where x' is defined by $x'_i = (x_i + 1 \bmod p_i)$ for all $i \in [n]$. Let $j \in [n]$ be a coordinate with $p_j > 2$. By the same argument, we also have that $x \in A$ holds if and only if $x'' \notin B$, where $x''_i = x'_i$ for $i \in [n] \setminus \{j\}$ and $x''_j = (x_j + 2 \bmod p_j)$. Thus, both x' and x'' belong to B or neither of them does. This holds for every vector x', implying that j is irrelevant for the set B and thus also for A.

As there are no relevant coordinates for A and B outside the set I of coordinates with $p_i = 2$, we can choose a set W of functions from I to $[2]$ such that $A = A_W$. This makes

$$B = \{y \in S_p \ : \ y \text{ intersects all } x \in A\} = B_W.$$

We have $|A| + |B| = |S_p|$ and $|A| \cdot |B| = |S_p|^2/4$ if and only if $|W| = 2^{|I|-1}$.

In [FF80], the case $p = (k, \ldots, k)$, $n = r + 2$ was considered and it was noted that the radius 1 ball C in all the coordinates is an r-intersecting. The cardinality of C is $|C| = (r + 2)(k - 1) + 1$. The "trivial" r-intersecting family obtained by fixing r coordinates has k^2 elements, thus it does not provide an extremal example if $r \geq k$. If $k = r+1$, then the trivial example and the family C are of the same size. This means that fixing r coordinates is definitely not the only way to obtain maximal r-cross-intersecting pairs, but it is not clear whether both examples are maximal or not. The $r = 1$ special case of Theorem 1 suggests that the trivial example may still be maximal in the case $k = r + 1$, although clearly not the only maximal example. Theorem 7 establishes this for $r > 6$. The intermediate cases $2 \leq r \leq 6$ are still open but could possibly be handled by computer search.

Finally, we mention that there is a simple connection between the problem discussed in this paper and a question related to communication complexity. Consider the following two-person communication game: Alice and Bob each receive a vector from S_p, and they have to decide whether the vectors are r-intersecting. In the *communication matrix* of such a game, the rows are indexed by the possible inputs of Alice, the columns by the possible inputs of Bob, and an entry of the matrix is 1 or 0 corresponding to the "yes" or "no" output the players have to compute for the corresponding inputs. In the study of communication games, the submatrices of this matrix in which all entries are equal play a special role. The largest area of an all-1 submatrix is the maximal value of $|A| \cdot |B|$ for r-cross-intersecting families $A, B \subseteq S_p$.

Acknowledgment. We are indebted to G. O. H. Katona, R. Radoičić, and D. Scheder for their valuable remarks.

References

[Be74] C. Berge: Nombres de coloration de l'hypegraphe h-parti complet, in *Hypergraph Seminar, Lecture Notes in Math.* **411**, Springer-Verlag, Heidelberg, 1974 Page 13–20.

[Bol86] B. Bollobás: *Combinatorics. Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability*, Cambridge University Press, Cambridge, 1986.

[Bo08] P. Borg: Intersecting and cross-intersecting families of labeled sets, *Electron. J. Combin.* **15** (2008), N9.

[DeF83] M. Deza and P. Frankl: Erdős-Ko-Rado theorem – 22 years later, *SIAM J. Alg. Disc. Methods* **4** (1983), 419–431.

[EKR61] P. Erdős, C. Ko and R. Rado: Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser.* **2 (12)** (1961), 313–318.

[F95] P. Frankl: Extremal set systems, in *Handbook of Combinatorics* (R. Graham et al., eds.), Elsevier, Amsterdam, 1995, 1293–1329.

[FF80] P. Frankl and Z. Füredi: The Erdős-Ko-Rado theorem for integer sequences, *SIAM J. on Discrete and Alg. Methods* **1** (1980), 376–381.

[GK78] C. Greene and D. J. Kleitman: Proof techniques in the ordered sets, in *Studies in Combinatorics*, Math. Assn. America, Washington DC, 1978, 22–79.

[Hi77] A. J. W. Hilton: An intersection theorem for a collection of families of subsets of a finite set, *J. London Math. Soc.* **2** (1977), 369–384.

[K72] G. O. H. Katona: A simple proof of the Erdős-Ko-Rado theorem, *J. Combin. Theory Ser. B* **13** (1972), 183–184.

[Liv79] M. L. Livingston: An ordered version of the Erdős-Ko-Rado theorem, *J. Combin. Theory Ser. A* **26** (1979), 162–165.

[Mo82] A. Moon: An analogue of the Erdős-Ko-Rado theorem for the Hamming schemes H(n,q), *J. Combin. Theory Ser. A* **32** (1982), no. 3, 386–390.

[Py86] L. Pyber: A new generalization of the Erdős-Rado-Ko theorem, *J. Combin. Theory Ser. A* **43** (1986), 85–90.

[St80] D. Stanton: Some Erdős-Ko-Rado theorems for Chevalley groups,*SIAM J. Algebraic Discrete Methods* **1** (1980), no. 2, 160–163.

[Zh13] H. Zhang: Cross-intersecting families of labeled sets, *Electron. J. Combin.* bf 20(1) (2013), P17.