Maximum hitting for n sufficiently large

Ben Barber*

October 3, 2014

Abstract

For a left-compressed intersecting family $\mathcal{A} \subseteq [n]^{(r)}$ and a set $X \subseteq [n]$, let $\mathcal{A}(X) = \{A \in \mathcal{A} : A \cap X \neq \emptyset\}$. Borg asked: for which X is $|\mathcal{A}(X)|$ maximised by taking $\mathcal{A}$ to be all r-sets containing the element 1? We determine exactly which X have this property, for n sufficiently large depending on r.

1 Introduction

Write $[n] = \{1, 2, \ldots, n\}$ and $[m, n] = \{m, m + 1, \ldots, n\}$. Denote the set of r-sets from a set S by $S^{(r)}$. A *family* of sets is a subset of $[n]^{(r)}$ for some n and r. We think of a set A as an increasing sequence of elements $a_1 a_2 \ldots a_r$. The *compression order* on $[n]^{(r)}$ has $A \leq B$ if and only if $a_i \leq b_i$ for $1 \leq i \leq r$. A family $\mathcal{A}$ is *left-compressed* if $A \in \mathcal{A}$ whenever $A \leq B$ for some $B \in \mathcal{A}$. The corresponding notion of left-compression is described in Section 2.

We call a family *intersecting* if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{A}$. (If $n < 2r$ then every family is intersecting.) The most basic result about intersecting families is the Erdős-Ko-Rado Theorem. For any n and r, write $\mathcal{S} = \{A \in [n]^{(r)} : 1 \in A\}$ for the *star* at 1.

Theorem 1 (Erdős-Ko-Rado [3])**.** *If $n \geq 2r$ and $\mathcal{A} \subseteq [n]^{(r)}$ is intersecting, then $|\mathcal{A}| \leq |\mathcal{S}|$.*

Borg considered a variant problem where we only count members that meet some fixed set X. For a family $\mathcal{A}$ and a non-empty set X, write

$$\mathcal{A}(X) = \{A \in \mathcal{A} : A \cap X \neq \emptyset\}.$$

Theorem 1 tells us that we can maximise $|\mathcal{A}(X)|$ by taking $\mathcal{A}$ to consist of all r-sets containing some fixed element of X. To avoid this trivial case we insist that $\mathcal{A}$ be left-compressed, which rules out stars centred anywhere but 1. The star at 1 remains the optimal family if $1 \in X$, so we assume further that $X \subseteq [2, n]$.

Question 2. *For which X do we have $|\mathcal{A}(X)| \leq |\mathcal{S}(X)|$ for all left-compressed intersecting families $\mathcal{A}$?*

Borg asked this question in [2], giving a complete answer for the case $|X| \geq r$ and a partial answer for the case $|X| < r$. Call X *good* (for n and r) if for every left-compressed intersecting family $\mathcal{A} \subseteq [n]^{(r)}$ we have $|\mathcal{A}(X)| \leq |\mathcal{S}(X)|$.

*Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, CB3 0WB, UK. b.a.barber@dpmms.cam.ac.uk

Theorem 3 (Borg [2])**.** *Let $r \geq 2$, $n \geq 2r$ and $X \subseteq [2, n]$.*

(a) *If $|X| > r$, then X is good.*

(b) *If X is good and $X \leq X'$, then X' is good.*

(c) *For any $k \leq r$, $\{2k, 2k+2, \ldots, 2r\}$ is good.*

(d) *If $n = 2r$ and $|X| = r$, then X is good if and only if $\{2, 4, \ldots, 2r\} \leq X$.*

(e) *If $n > 2r$, $|X| = r$ and either*

  (i) *$r \geq 4$ and $X \neq [2, r+1]$,*

  (ii) *$r = 3$ and $\{2, 3\} \not\subseteq X$, or*

  (iii) *$r = 2$ and $\{2, 3\} \neq X$,*

  then X is good. Otherwise, X is not good.

It is not true that all X are good. For example, consider the *Hilton-Milner* family $\mathcal{T} = \mathcal{S}([2, r+1]) \cup \{[2, r+1]\}$. The family $\mathcal{T}$ is left-compressed and for any $X \subseteq [2, r+1]$, $|\mathcal{T}(X)| = |\mathcal{S}(X)| + 1$, so X is not good.

Our main result is that, surprisingly, for large n and $|X| \geq 4$ this turns out to be the only obstruction.

Theorem 4. *Let $r \geq 3$, $n \geq 2r$ and $X \subseteq [2, n]$ with $|X| \leq r$. If $X \not\subseteq [2, r+1]$ and either*

(i) *$|X| \geq 4$,*

(ii) *$|X| = 3$ and $\{2, 3\} \not\subseteq X$,*

(iii) *$|X| = 2$ and $2, 3 \notin X$, or*

(iv) *$|X| = 1$,*

then, for n sufficiently large, X is good. Otherwise, X is not good.

For $r = 2$, condition (iii) needs to be replaced by $X \neq \{2, 3\}$. The result can then be checked easily by hand or read out of Theorem 3 in conjunction with the Hilton-Milner example, so we assume $r \geq 3$ for simplicity.

Our proof uses Ahlswede and Khachatrian's notion of generating sets to express the sizes of maximal left-compressed intersecting families, and their restrictions under X, as polynomials in n. It turns out to be sufficient to consider only leading terms, reducing a question about intersecting families of r-sets to a question about intersecting families of 2-sets, which have a very simple structure.

Section 2 sets out the basic properties of compressions and generating sets that we shall use. Section 3 describes a way of thinking about maximal left-compressed intersecting families and proves the lemma that allows us to compare coefficients of polynomials instead of set sizes. Section 4 completes the proof of Theorem 4. Section 5 discusses possible improvements and generalisations.

2 Compressions and generating sets

In this section we describe the notion of left-compression corresponding to $\leq$ on $[n]^{(r)}$ and the use of generating sets.

2.1 Compressions

For a set A, and $i < j$, the *ij-compression* of A is

$$C_{ij}(A) = \begin{cases} A - j + i & \text{if } j \in A, i \notin A, \\ A & \text{otherwise;} \end{cases}$$

that is, replace j by i if possible. Observe that $A \leq B$ if and only if A can be obtained from B by a sequence of ij-compressions.

For a set family $\mathcal{A}$, define

$$C_{ij}(\mathcal{A}) = \{C_{ij}(A) : A \in \mathcal{A} \text{ and } C_{ij}(A) \notin \mathcal{A}\} \cup \{A : A \in \mathcal{A} \text{ and } C_{ij}(A) \in \mathcal{A}\};$$

that is, compress A if possible. Observe that $\mathcal{A}$ is left-compressed if and only if $C_{ij}(\mathcal{A}) = \mathcal{A}$ for all $i < j$. We will use the following basic result.

Lemma 5. *If $\mathcal{A}$ is intersecting then $C_{ij}(\mathcal{A})$ is intersecting.*

Proof. The proof is an easy case check. Details, and a further introduction to compressions, can be found in Frankl's survey article [4]. □

Lemma 5 means that we can always compress an intersecting family to a left-compressed intersecting family of the same size by repeatedly applying ij-compressions. We eventually reach a left-compressed family as $\sum_{A \in \mathcal{A}} \sum_{i=1}^{r} a_i$ is positive and strictly decreases with each successful compression.

2.2 Generating sets

For any r and n, and a collection $\mathcal{G}$ of sets, the family *generated* by $\mathcal{G}$ is

$$\mathcal{F}(r, n, \mathcal{G}) = \{A \in [n]^{(r)} : A \supseteq G \text{ for some } G \in \mathcal{G}\}.$$

Generating sets were introduced by Ahlswede and Khachatrian [1], and are useful for the study of intersecting families because they give a restricted number of sets on which all the intersecting actually happens.

Lemma 6 ([1])**.** *For $n \geq 2r$, $\mathcal{F}(r, n, \mathcal{G})$ is intersecting if and only if $\mathcal{G}$ is.*

Proof. If $\mathcal{G}$ is intersecting then certainly $\mathcal{F}(r, n, \mathcal{G})$ is. Conversely, if $\mathcal{G}$ contains two disjoint sets then (since $n \geq 2r$) they can be completed to disjoint r-sets in $\mathcal{F}(r, n, \mathcal{G})$. □

If $\mathcal{G}$ generates a left-compressed intersecting family then

$$\mathcal{G}' = \{G' : G' \leq G \text{ for some } G \in \mathcal{G}\}$$

generates the same family, so we may assume that $\mathcal{G}$ is 'left-compressed' (overlooking non-uniformity) and can therefore be described by listing its maximal elements. It is convenient to take

$$\mathcal{F}(r, n, \mathcal{G}) = \{A \in [n]^{(r)} : A \prec G \text{ for some } G \in \mathcal{G}\},$$

where $A \prec G$ ('A is generated by G') if and only if $|G| \leq |A|$ and $a_i \leq g_i$ for $1 \leq i \leq |G|$. We can think of $\prec$ as an extension of $\leq$ to the non-uniform case, where 'missing' elements are assumed to take the value ∞. Thus

$$123 \prec 12 \ (= 12\infty);$$
$$(12\infty =) \ 12 \not\prec 123.$$

The following weaker form of Lemma 6 is better suited to our new definition and is sufficient for our purposes.

Corollary 7. *Let $n \geq 2r$ and $\mathcal{G}$ be a collection of subsets of $[2s]$ of size at most s. If $\mathcal{F}(s, 2s, \mathcal{G})$ is intersecting, then so is $\mathcal{F}(r, n, \mathcal{G})$.* □

3 Maximal left-compressed intersecting families

We say an intersecting family $\mathcal{A} \subseteq [n]^{(r)}$ is *maximal* if no other set can be added to $\mathcal{A}$ while preserving the intersecting property. The maximal objects in the set of left-compressed intersecting families are maximal intersecting families (otherwise an extension could be compressed to a left-compressed extension), so the ordering of 'maximal' and 'left-compressed' is unimportant.

The maximal left-compressed intersecting subfamilies of $[n]^{(2)}$ are $\{12, 13, \ldots, 1n\}$ and $\{12, 13, 23\}$, and we can already distinguish between these families when $n = 4$. In fact, the same phenomenon occurs for all r.

Lemma 8. *Let $\mathcal{A} \subseteq [2r]^{(r)}$ be a maximal left-compressed intersecting family and $n \geq 2r$. Then $\mathcal{A}$ extends uniquely to a maximal left-compressed intersecting subfamily of $[n]^{(r)}$. Moreover, every maximal left-compressed intersecting subfamily of $[n]^{(r)}$ arises in this way.*

Proof. Since $\mathcal{A}$ is left-compressed, it can be completely described by listing its $\leq$-maximal elements $A_1, \ldots, A_k$. Some of these sets might contain final segments of [2r]. The idea is that the elements of these final segments would take larger values if they were allowed to, so we obtain a generating set by 'replacing them by ∞'.

For $A = A_i$, take s greatest with $a_s < r + s$ (s exists since $[r + 1, 2r]$ is not a member of any left-compressed intersecting family), and let $A' = a_1 \ldots a_s$. Then $\mathcal{G} = \{A'_1, \ldots, A'_k\}$ generates $\mathcal{A}$, as the sets generated by A'_i are precisely those lying below A_i. Since $\mathcal{G}$ is a collection of subsets of $[2r]$ of size at most r and $\mathcal{A} = \mathcal{F}(r, 2r, \mathcal{G})$ is intersecting, Corollary 7 tells us that $\mathcal{F}(r, n, \mathcal{G})$ is a left-compressed intersecting family for every n.

Now let $\mathcal{B}$ be any extension of $\mathcal{A}$ to a left-compressed intersecting subfamily of $[n]^{(r)}$. We will show that $\mathcal{B} \subseteq \mathcal{F}(r, n, \mathcal{G})$. Indeed, if $\mathcal{B} \not\subseteq \mathcal{F}(r, n, \mathcal{G})$ then there is a $B \in \mathcal{B} \setminus \mathcal{F}(r, n, \mathcal{G})$. We claim that there is a $B' \in [2r]^{(r)}$ with $B' \leq B$ and $B' \notin \mathcal{F}(r, 2r, \mathcal{G})$, contradicting the maximality of $\mathcal{A}$.

We obtain B' from B by compressing as little as possible to get $B' \subseteq [2r]$; that is, we take $B' = (B \cap [2r]) \cup [q, 2r]$ with q chosen such that $|B'| = r$. Explicitly, $b'_i = \min(b_i, r + i)$. Now take $G \in \mathcal{G}$. Since $B \notin \mathcal{F}(r, n, \mathcal{G})$, there is an i with $b_i > g_i$. By construction, $r + i > g_i$. So $b'_i = \min(b_i, r + i) > g_i$, and G does not generate B'. Hence $\mathcal{A}$ extends uniquely to a maximal left-compressed intersecting subfamily of $[n]^{(r)}$.

It remains to show that every maximal left-compressed intersecting subfamily of $[n]^{(r)}$ arises in this way. So suppose $\mathcal{C} \subseteq [n]^{(r)}$ is a maximal left-compressed intersecting family with $\mathcal{C} \cap [2r]^{(r)}$ not maximal. Let $\mathcal{D}_0$ be an extension of $\mathcal{C} \cap [2r]^{(r)}$ to a maximal left-compressed intersecting subfamily of $[2r]^{(r)}$, and let $\mathcal{D}$ be the unique maximal extension of $\mathcal{D}_0$ to $[n]^{(r)}$. Since $\mathcal{C}$ is maximal and $\mathcal{D} \setminus \mathcal{C} \neq \emptyset$, there is a $C \in \mathcal{C} \setminus \mathcal{D}$. As above, we obtain $C' \in [2r]^{(r)}$ with $C' \leq C$ and $C' \notin \mathcal{D}_0$. But then $C' \notin \mathcal{C}$, contradicting the assumption that $\mathcal{C}$ is left-compressed. $\square$

Lemma 8 allows a compact description of maximal left-compressed intersecting families. For example, $\{1\}$ generates the star and $\{1(r+1), [2, r+1]\}$ generates the Hilton-Milner family. Enumerating the generating sets using a computer is feasible for small r; for $r = 3$ they are $\{1\}, \{23\}, \{345\}, \{14, 234\}, \{13, 235, 145\}$ and $\{12, 245\}$.

In view of Lemma 8, our key tool is the following.

Lemma 9. *Let $n \geq 2$, $X \subseteq [2, 2r]$. Then*

$$|\mathcal{F}(r, n, \mathcal{G})(X)| = \sum_{i=1}^{r} |\mathcal{F}(i, 2r, \mathcal{G})(X)| \binom{n-2r}{r-i}.$$

Proof. How do we construct a member of $\mathcal{F}(r, n, \mathcal{G})(X)$? We first choose an initial segment for our set that is contained in $[2r]$ and witnesses the membership of $\mathcal{F}(r, n, \mathcal{G})(X)$ (i.e. meets X and is $\prec$ some $G \in \mathcal{G}$). We then complete our set by taking as many elements as we need from outside $[2r]$. This gives rise to the size claimed. $\square$

4 Proof of Theorem 4

We first show that X is not good if the given conditions do not hold. We have already seen that for $X \subseteq [2, r+1]$ the Hilton-Milner family shows that X is not good for any n. In each of the remaining cases we claim that the family generated by $\{23\}$ shows that X is not good for any n.

So take $X = 23k$ with $k \geq r+2$. We have

$$|\mathcal{F}(r, n, \{1\})(23k)| = \binom{n-2}{r-2} + \binom{n-3}{r-2} + \binom{n-4}{r-2},$$

where the first term counts the sets containing 1 and 2, the second term the sets containing 1 and 3 but not 2, and the third term the sets containing 1 and k but neither 2 nor 3. Similarly,

$$|\mathcal{F}(r, n, \{23\})(23k)| = \binom{n-2}{r-2} + \binom{n-3}{r-2} + \binom{n-3}{r-2},$$

where the terms count the sets containing 1 and 2, the sets containing 1 and 3 but not 2, and the sets containing 2 and 3 but not 1 respectively. Since $r \geq 3$, $|\mathcal{F}(r, n, \{23\})(23k)| > |\mathcal{F}(r, n, \{1\})(23k)|$ and $23k$ is not good.

Next take $X = 3j$ with $j \geq r+2$. We have

$$|\mathcal{F}(r, n, \{1\})(3j)| = \binom{n-2}{r-2} + \binom{n-3}{r-2},$$

where the terms count the sets containing 1 and 3, and the sets containing 1 and j but not 3 respectively. Similarly,

$$|\mathcal{F}(r, n, \{23\})(3j)| = \binom{n-2}{r-2} + \binom{n-3}{r-2} + \binom{n-4}{r-3},$$

where the terms count the sets containing 1 and 3, the sets containing 2 and 3 but not 1, and the sets containing 1, 2 and j but not 3 respectively. Again, since $r \geq 3$, $|\mathcal{F}(r, n, \{23\})(3j)| > |\mathcal{F}(r, n, \{1\})(3j)|$ and $3j$ is not good. It follows from Theorem 3(b) that $2j$ is not good either.

Now we take X satisfying the conditions of the theorem and show that X is good for n sufficiently large. We will show that, for any $\mathcal{G} \neq \{1\}$, $|\mathcal{F}(2, 2r, \mathcal{G})(X)| < |\mathcal{F}(2, 2r, \{1\})(X)| = |X|$. Note that, for any $\mathcal{G}$, $|\mathcal{F}(1, 2r, \mathcal{G})(X)| = 0$ as the only possible singleton generator is 1, which does not meet X. So by Lemma 9, $\mathcal{F}(r, n, \mathcal{G})(X)$ has size polynomial in n with leading coefficient $|\mathcal{F}(2, 2r, \mathcal{G})(X)|$, from which the result will follow.

There are two maximal left-compressed intersecting families of 2-sets, and $\mathcal{F}(2, 2r, \mathcal{G})(X)$ must be contained in one of them. We handle each case separately.

Suppose first that $\mathcal{F}(2, 2r, \mathcal{G})(X) \subseteq \{12, 13, 23\}$. Then it is enough to show that

$$|\{12, 13, 23\}(X)| < |X|.$$

This is clearly true for $|X| \geq 4$. If $|X| = 3$, then it is true because one of 2 or 3 is missing from X so that $|\{12, 13, 23\}(X)| \leq 2$. If $|X| = 2$, then it is true because both 2 and 3 are missing from X, so that $|\{12, 13, 23\}(X)| = 0$. Finally, if $|X| = 1$, then it is true because $X = \{i\}$ with $i \geq r + 2 \geq 4$.

Next suppose that $\mathcal{F}(2, 2r, \mathcal{G})(X) \subseteq \{12, 13, \ldots, 1(2r)\}$. Since $\mathcal{F}(r, 2r, \mathcal{G})$ is left-compressed and has a member not containing the element 1, it has $[2, r+1]$ as a member. Hence by the intersecting property of the generators, $\mathcal{F}(2, 2r, \mathcal{G})(X)$ cannot contain $1j$ for any $j \geq r + 2$. But $X \not\subseteq [2, r + 1]$, so there is such a $j \in X \setminus [2, r + 1]$ and $|\mathcal{F}(2, 2r, \mathcal{G})(X)| < |X|$. □

5 Improvements and generalisations

What happens for small n? Theorem 3(c) tells us that our characterisation cannot be correct for all $n \geq 2r$.

Question 10. *How large is 'sufficiently large' for n in Theorem 4?*

For $2 \leq r \leq 5$, computational results suggest that $n \geq 2r + 2$ is large enough for our characterisation to be correct. It would be particularly nice to show that $n \geq 2r + c$ is sufficient for some constant c independent of r.

A natural conjecture is that for $n = 2r$, $[2k, 2k + 2, \ldots, 2r]$ is the unique minimal good set of its size. However, this is false; computational results give that $\{7, 10\}$ and $\{5, 8, 10\}$ are unique minimal good sets of their size when $r = 5$.

Question 11. *Is there a 'nice' characterisation of the good sets for $n = 2r$ when r is sufficiently large?*

It seems unlikely that a good explicit description exists for intermediate values of r and n. The following may be easier.

Question 12. *Is there a short list of families, one of which maximises $|\mathcal{A}(X)|$ for any X?*

Versions of Lemma 8 hold for any property that is preserved under left-compression and can be detected on generating sets. The most obvious candidate is that of being t-intersecting (a family $\mathcal{A}$ is *t-intersecting* if $|A \cap B| \geq t$ for all $A, B \in \mathcal{A}$). Indeed, an identical argument gives the corresponding result that, for large n, a set $X \subseteq [t+1, n]$ with $|X| \geq t+3$ is good if and only if $X \not\subseteq [t+1, r+1]$. (For smaller X the form of good X is again decided by the need to prevent problems caused when $\mathcal{F}(t+1, 2r-t+1, \mathcal{G})(X) \subseteq [t+2]^{(t+1)}$.)

In the context of t-intersecting families it may be more natural to consider

$$\mathcal{A}(s, X) = \{A \in \mathcal{A} : |A \cap X| \geq s\}.$$

For $s = 1$ the argument relies on the fact that maximal left-compressed t-intersecting families of $(t+1)$-sets have one of two very simple forms. For $s = 2$, even the $t = 1$ case is complicated by the larger number of structures of intersecting families of 3-sets (more generally, $(t+s)$-sets); this problem seems likely to get worse for larger s and t.

Acknowledgements. I would like to thank the anonymous referees for carefully reading an earlier draft of this paper and making a number of helpful comments.

References

[1] Rudolf Ahlswede and Levon H. Khachatrian. The complete intersection theorem for systems of finite sets. *European J. Combin.*, 18(2):125–136, 1997.

[2] Peter Borg. Maximum hitting of a set by compressed intersecting families. *Graphs and Combinatorics*, 27(6):785–797, 2011.

[3] P. Erdős, Chao Ko, and R. Rado. Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford Ser. (2)*, 12:313–320, 1961.

[4] P. Frankl. The shifting technique in extremal set theory. In C. Whitehead, editor, *Surveys in Combinatorics*, volume 123 of *London Math. Soc. Lecture Notes Series*, pages 81–110. Cambridge University Press, Cambridge, 1987.