Proceedings of the 12th European Conference on Combinatorics, Graph Theory and Applications
EUROCOMB'23
Prague, August 28 - September 1, 2023

Avoiding intersections of given size in finite affine spaces $\mathrm{AG}(2, q)$

(Extended abstract)

Benedek Kovács* Zoltán Lóránt Nagy†

Abstract

We study the set of intersection sizes of a k-dimensional affine subspace and a point set of size $m \in [0, 2^n]$ of the n dimensional binary affine space $\mathrm{AG}(n, 2)$.

DOI: `https://doi.org/10.5817/CZ.MUNI.EUROCOMB23-093`

1 Introduction

In this paper, we address the following general problem. Let S be a subset of size m of the affine space $\mathrm{AG}(n, q)$. Does there always exist a k-dimensional affine subspace which contains exactly t points of S? A k-dimensional affine subspace will be referred to as a *k-flat*, and a k-flat which contains exactly t points of a given set S will be called a *$[k, t]$-flat (induced by S)*. If for every m-element set S in the n-dimensional affine space, there is a k-flat containing exactly t points of S, we say that the pair $[n, m]$ forces the pair $[k, t]$, or that t-sets are unavoidable in k-flats. We use the notation $[n, m]_q \to [k, t]$ for this concept.

Our main focus will be the case $q = 2$, and we will omit the index q except when we wish to refer to arbitrary finite fields. The graph theoretic analogue of this problem was initiated by Erdős, Füredi, Rothschild and T. Sós [7]. Let $G(n, m)$ denote a graph on n vertices and m edges. Fix a positive integer k and a pair of integers (n, m) such that $0 \le m \le \binom{n}{2}$. For

*ELTE Linear Hypergraphs Research Group, Eötvös Loránd University, Budapest, Hungary. The author is partially supported by the ÚNKP, New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

†ELTE Linear Hypergraphs Research Group, Eötvös Loránd University, Budapest, Hungary. The author is supported by the Hungarian Research Grant (NKFI) No. PD 134953.

which t does it hold that any n-vertex graph with m edges contains an induced subgraph on k vertices having exactly t edges? Equivalently, we are seeking pairs (k, t) such that k-vertex induced subgraphs with t edges are unavoidable in graphs of the form $G(n, m)$. Erdős, Füredi, Rothschild and T. Sós introduced the notation $(n, m) \to (k, t)$ for the case when this is true.

Their main result showed that forced pairs (k, t) are rare in the following sense. Consider the set $Sp(n; k, t)$ of all edge cardinalities m such that $(n, m) \to (k, t)$. Its density is the ratio $\frac{|Sp(n;k,t)|}{\binom{n}{2}}$. They proved that the limit superior of this density is bounded from above by $2/3$ apart from a handful cases, and is 0 for the majority of the pairs (k, t) for fixed and large enough k. Erdős, Füredi, Rothschild and T. Sós conjectured that in fact it is bounded from above by $1/2$ apart from finitely many pairs (k, t). This was confirmed recently by He, Ma and Zhao [9]. Several related problems have been studied in the last couple of years [1, 4]. Our problem can be viewed as the q-analogue of this problem, when we investigate whether all subspaces of given dimension k can avoid to have an intersection of size t with an m-element set of the space $\mathrm{AG}(n, q)$, which corresponds to $\mathbb{F}_q^n$.

Considering the case $q = 3$, $k = 1$ and $t = 3$, we in turn get the famous cap set problem, which asks for the maximum number of points in $\mathrm{AG}(n, 3)$ without creating a line, or in other words, without containing a 3-term arithmetic progression. There has been a recent breakthrough due to Ellenberg and Gijswijt [6], building upon the ideas of Croot, Lev and Pach [5], which showed that to avoid complete lines, $|S|/3^n$ has to be exponentially small. This connection highlights the complexity of the problem.

There has been significant interest in the case when we want to forbid each intersection of size larger than or equal to f instead of avoiding only f-sets in k-flats. If there exists a set $S \in AG(n, q)$ for which all k-dimensional affine subspaces contain at most c points of S, then S is called (k, c)-subspace evasive.The importance of such sets relies on its connections to explicit contructions to bipartite Ramsey graphs by Pudlák and Rödl [10] and with list-decodable codes by Guruswami [8].

By a standard application of the first moment method, Guruswami obtained random subsets of $\mathbb{F}^n$ of large size which are (k, c)-subspace evasive. ($\mathbb{F}$ denotes a finite field.)

Theorem 1.1 (Guruswami, [8])**.** *For any fixed pair (k, c), there exists a (k, c)-evasive set in $\mathbb{F}^n$ of size at least $C \cdot |\mathbb{F}|^{n(1-\frac{2k}{c})}$, where $C > 0$ is a constant independent of n.*

Corollary 1.2. *For any fixed pair (k, t) with $t > 1$, there exists a constant $C > 0$ for which the following holds: if $m \leq C \cdot 2^{n(1-\frac{2 \cdot k}{t-1})}$, then t-sets are avoidable in the k-flats of $\mathrm{AG}(n, 2)$, i.e., in this case $[n, m] \not\to [k, t]$.*

Surprisingly it turned out that the obtained bounds are sharp in a weak sense.

Theorem 1.3 (Sudakov and Tomon, [11])**.** *Let $\mathbb{F}$ be a field, $k \in \mathbb{Z}^+$ and $\varepsilon \in (0, 0.05)$. If n is sufficiently large with respect to k, and $S \subseteq \mathbb{F}^n$ has size $m \geq |\mathbb{F}|^{n(1-\varepsilon)}$, then S is not $(k, \frac{k-\log_2(1/\varepsilon)}{8\varepsilon})$-subspace evasive.*

Note however that S being non-(k,c)-evasive does not necessarily imply that S contains a k-flat with $c+1$ points, except when $c = |\mathbb{F}|^n - 1$. The latter case on the other hand is not covered by the theorem, as $\frac{k-\log_2(1/\varepsilon)}{8\varepsilon} < 2^{k-3}$ holds for every $\varepsilon \in (0, 0.05)$.

2 Our main results

Similarly to the investigation in [7], we define the set of forcing sizes m with respect to $[k,t]$, which are the sizes for which a $[k,t]$-flat is unavoidable, as follows.

Definition 2.1.

$$Sp(n;k,t) := \{m : [n,m]_q \to [k,t]\}$$

is called the *set of forcing sizes with respect to* n,k,t, and we refer to it as the $(n;k,t)$-*spectrum.*

$$\rho(n;k,t) := \frac{Sp(n;k,t)}{|\mathbb{F}|^n}$$

is the *density of the spectrum.*

Our aim is to characterize the spectra or at least bound the density of the spectra for various values of k and f. Note that from now on, $\mathbb{F}$ is considered to be the binary field. When it is not confusing, we use the notation $[a,b]$ for the integers in the interval. For a set of integers $H \subseteq \mathbb{Z}$ and $c \in \mathbb{Z}$, let $c - H = \{c - h : h \in H\}$. For any set $X \in \mathbb{F}_2^n$, $\overline{X}$ will denote the complement of X. Observe first that determining the spectrum $Sp[n,k,f]$ or its density is essentially the same problem as determining $Sp[n,k,2^k - f]$ or its density.

Lemma 2.2. *If* $n,k \geq 1$ *are integers and* $0 \leq f \leq 2^k$ *then* $Sp[n,k,f] = 2^n - Sp[n,k,2^k - f]$.

2.1 Some exact results on the spectrum

We start with investigating $Sp[n,k,f]$ in cases where k is small.

Proposition 2.3. *(i)* $Sp[n,1,0] = [0, 2^n - 2]$, *(ii)* $Sp[n,1,1] = [1, 2^n - 1]$,
(iii) $Sp[n,2,1] = [0,2^n] \setminus (\{2^n\} \cup \{2^n - 2^d : d \in [0,n]\})$,
(iv) $Sp[n,2,2] = [2, 2^n - 2]$, *(v)* $Sp[n,2,3] = [0,2^n] \setminus (\{0\} \cup \{2^d : d \in [0,n]\})$.

The determination of the spectrum $Sp[n,2,4]$ for the full 2-dimensional flat is still a challenging problem. This has been studied under the name of Sidon sets in binary vector spaces. A subset S of an Abelian group is a *Sidon set* if the only solutions to the equation $a + b = c + d$ with $a,b,c,d \in A$ are the trivial solutions when (a,b) is a permutation of (c,d). Observe that for $A = \mathbb{F}_2^n$, S contains a $(2,4)$-flat if and only if it is not a Sidon set. There are known results on Sidon sets in this setting which imply the following.

Proposition 2.4 (([3], also see [2, 12]))**.**

1. *There exists a constant* $C > 0$ *such that* $[n,m] \to [2,4]$ *for every* $m \geq C \cdot 2^{\frac{1}{2}n}$.
2. *The explicit construction* $\{(x, x^3) : x \in \mathbb{F}_{2^{n/2}}\}$ *shows that* $[n, 2^{\frac{1}{2}n}] \not\to [2,4]$.

The complete characterization of the spectrum for case $k = 3, f = 4$ requires a combination of various tools, including probabilistic methods.

Proposition 2.5. *For every $n \geq 3$, $Sp[n, 3, 4] = [4, 2^n - 4]$.*

2.2 Bounds on the density of the spectrum - unavoidable elements

Our main results are concerning the case when f is a power of 2.

Theorem 2.6. *Suppose that $k > \ell > 0$. Then there exist absolute constants $C, D > 0$ such that $[n, m] \to [k, 2^{k-\ell}]$ for $m \in \left(C \cdot 2^{n\left(1-\frac{1}{2^{k-\ell-1}}\right)}, D \cdot 2^n\right)$. Moreover, for each $\varepsilon > 0$ and sufficiently large n, $\rho(n; k, 2^{k-\ell}) \geq \frac{1-\varepsilon}{2^{\ell-1}}$.*

If ℓ is small, we can prove even stronger results.

Theorem 2.7. *Suppose that $\ell \in \{0, 1\}$. Then there exists an absolute constant C such that $[n, m] \to [k, 2^{k-\ell}]$ for $m \in \left(C \cdot 2^{n\left(1-\frac{1}{2^{k-\ell-1}}\right)}, 2^n - C \cdot 2^{n\left(1-\frac{1}{2^{k-\ell-1}}\right)}\right)$.*

Finally we discuss a case when t is the sum of two consecutive powers of 2. This case is significantly more involved compared to the case of $t = 2^{k-\ell}$.

Theorem 2.8. *For every pair (k, ℓ) of integers with $2 \leq \ell \leq k - 1$, the density of integer values m within the interval $\left[0, \frac{1}{2^{\ell-1}} \cdot 2^n\right]$ for which $[n, m] \to [k, 3 \cdot 2^{k-\ell}]$ holds, tends to 1 as $n \to \infty$. Hence $\rho(n; k, 3 \cdot 2^{k-\ell}) \geq \frac{1}{2^{\ell-1}}$.*

The above described results are relying on bounds on the size of cut in hypercube, induction and supersaturation results, and bounds on the additive energy.

If f is not a power of 2, then we must expect that the spectrum is *scattered*, several values of m are missing from it.

2.3 Missing elements from the spectrum

The random construction of Guruswami showed that we can obtain a (k, c)-evasive set on at least $2^{n(1-\frac{2k}{c})}/2^{k+1}$ points. Below we refine his argument using alteration.

Theorem 2.9. *Let $k, c \in \mathbb{Z}^+$ with $c \geq k + 1$. Then for $n > k$, there exists a (k, c)-evasive set in $\mathbb{F}_2^n$ of size at least*

$$\lfloor K \cdot 2^{n\left(1-\frac{k}{c}\right)} \rfloor - 1 \quad \textit{where} \quad K = K(k, c) := \frac{c}{c+1} \cdot 2^{k(k+1)/c} \cdot \left(2e^{2/3}(c+1)\binom{2^k}{c+1}\right)^{-\frac{1}{c}}.$$

Thus if $t > k$, then the smallest element in the spectrum of $Sp[n, k, t]$ is exponential in n. There is another reason why some elements are missing from a spectrum. The lexicographic construction below shows that if m can be obtained as a sum of few powers of 2, then the intersections with k-flats avoids many possible sizes.

Definition 2.10. Let m be an integer in $[0, 2^n]$. The *lexicographic construction of m points* is defined as follows. Take every point $P = (x_1, x_2, \ldots, x_n)$ in $\mathbb{F}_2^n$ for which the binary representation of the n-digit binary number $\overline{x_1 x_2 \ldots x_n}$ represents a number smaller than m.

Proposition 2.11. *Let m be a positive integer and let $M(m)$ denote the number of nonzero digits in the binary representation of m. If $M(t) > M(m)$ then $[n, m] \not\to [k, t]$.*

Corollary 2.12. *The lexicographic construction guarantees that for a given pair of integers k, t, the number of avoidable intersection sizes with respect to k-flats is at least a polynomial in n having degree $M(t)$.*

The theorem above can be combined with the former results on (k, c)-subspace-evasive sets. Indeed, suppose that we have a point set S_0 which avoids k-flats having intersection of size r for each $r \in [t_1, t_2]$ and $c < t_2 - t_1$. Then the union of S_0 and a (k, c)-subspace-evasive set S_1 will avoid k-flats with intersections of size $t \in [t_1 + c, t_2]$.

In fact the lexicographic construction shows further values of m which are avoidable. Suppose that m is a difference of two powers of 2. While $M(m)$ might be large, the lexicographic construction of m points shows that $[k, t]$ is avoidable if t cannot be written as a power of 2 or a difference of two powers of 2. It is easy to deduce a statement similar to Proposition 2.11. Still, we believe that in general the following conjecture holds.

Conjecture 2.13. *Let $t > k$ integers. Then* $\lim_{n\to\infty} \rho[n, k, t] = 1$.

References

[1] Axenovich, M., Weber, L. (2021). Absolutely avoidable order-size pairs for induced subgraphs. arXiv preprint arXiv:2106.14908.

[2] Bennett, M. (2018). Bounds on sizes of general caps in $AG(n, q)$ via the Croot-Lev-Pach polynomial method. arXiv preprint arXiv:1806.05303.

[3] Bose, R. C., Ray-Chaudhuri, D.K. (1960). On a Class of Error Correcting Binary Group Codes. Information and Control 3, pp. 68–79.

[4] Caro, Y., Lauri, J., Zarb, C. (2021). The feasibility problem for line graphs. arXiv preprint arXiv:2107.13806.

[5] Croot, E., Lev, V. F., Pach, P. P. (2017). Progression-free sets in are exponentially small. Annals of Mathematics, 331-337.

[6] Ellenberg, J. S., Gijswijt, D. (2017). On large subsets of with no three-term arithmetic progression. Annals of Mathematics, 339-343.

[7] Erdős, P., Füredi, Z., Rothschild, B. L., Sós, V. T. (1999). Induced subgraphs of given sizes. Discrete mathematics, 200(1-3), 61-77.

[8] Guruswami, V. (2011). Linear-algebraic list decoding of folded reed-solomon codes, in Proceedings of the 26th IEEE Conference on Computational Complexity.

[9] He, J., Ma, J., Zhao, L. (2021). Improvements on induced subgraphs of given sizes. arXiv preprint arXiv:2101.03898.

[10] Pudlák, P., Rödl, V. (2004). Pseudorandom sets and explicit constructions of Ramsey graphs. Quaderni di Matematica, 13, 327–346.

[11] Sudakov, B., Tomon, I. (2022). The extremal number of tight cycles. International Mathematics Research Notices, 2022(13), 9663-9684.

[12] Tait, M., Won, R. (2021). Improved Bounds on Sizes of Generalized Caps in AG(n,q). SIAM Journal on Discrete Mathematics, 35(1), 521-531.