

Near-sunflowers and focal families

Noga Alon*

Ron Holzman†

Abstract

We present some problems and results about variants of sunflowers in families of sets. In particular, we improve an upper bound of the first author, Körner and Monti on the maximum number of binary vectors of length n so that every four of them are split into two pairs by some coordinate. We also propose a weaker version of the Erdős-Rado sunflower conjecture.

1 Introduction

Introduced by Erdős and Rado [10], sunflowers (also called strong Δ -systems) have a long history of study and applications in extremal combinatorics and theoretical computer science. Recall that a family $\mathcal{H}$ of r distinct subsets of $[n] = \{1, 2, \dots, n\}$ is called a *sunflower* of size r if every $i \in [n]$ belongs to either 0, 1 or r of the sets in $\mathcal{H}$.

Erdős and Rado famously conjectured that if $\mathcal{F}$ is a k -uniform family of sets (i.e., $|A| = k$ for every $A \in \mathcal{F}$) not containing a sunflower of size r , then $|\mathcal{F}| \leq C^k$, where C is a constant depending only on r . For many years, the best known upper bound was close to $k!$ for any fixed r . A recent breakthrough due to Alweiss, Lovett, Wu and Zhang [3] improved the bound to $(\log k)^{(1+o(1))k}$ for any fixed r , but the original conjecture is still open even for $r = 3$.

Seeking a bound that depends on the size n of the ground set rather than the uniformity k , Erdős and Szemerédi [11] conjectured that if $\mathcal{F}$ is a family of subsets of $[n]$ not containing a sunflower of size r , then $|\mathcal{F}| \leq c^n$, where $c < 2$ is a constant depending only on r . They showed (implicitly, made explicit by Deuber et al. [6]) that their conjecture would follow from the Erdős-Rado conjecture. The recent solution by Ellenberg and Gijswijt [8] of the cap set problem, confirmed the $r = 3$ case of the Erdős-Szemerédi conjecture (via a reduction due to the first author, Shpilka and Umans [2]); see also Naslund and Sawin [17] and Hegedűs [12] for explicit bounds. But the conjecture is still open for $r \geq 4$.

*Department of Mathematics, Princeton University, Princeton, NJ 08544, USA and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by NSF grant DMS-2154082 and BSF grant 2018267.

†Department of Mathematics, Technion – Israel Institute of Technology, Technion City, Haifa 3200003, Israel. Email: holzman@technion.ac.il. Research done during a visit at the Department of Mathematics, Princeton University, supported by the H2020-MSCA-RISE project CoSP-GA No. 823748.

In this paper we introduce a weaker variant of sunflowers.

Definition 1.1. A family $\mathcal{H}$ of r distinct subsets of $[n]$ is called a *near-sunflower* of size r if every $i \in [n]$ belongs to either 0, 1, $r - 1$ or r of the sets in $\mathcal{H}$.

The weakening consists in adding the option of belonging to $r - 1$ of the sets (this renders the property interesting only for $r \geq 4$). It is natural in that it makes the property symmetric: if $\mathcal{H}$ is a near-sunflower then so is $\{[n] \setminus A : A \in \mathcal{H}\}$. One may hope that when sunflowers are replaced by near-sunflowers, the notoriously difficult conjectures of Erdős-Rado and Erdős-Szemerédi will become easier. We show that this is indeed the case in the Erdős-Szemerédi setting (bound depending on n), but leave the question in the Erdős-Rado setting (bound depending on k) open.

The following variant of near-sunflowers, in which one member of the family plays a distinguished role, will be of interest. It is convenient to define it for binary vectors of length n instead of subsets of $[n]$ – henceforth we will pass freely between these two equivalent formalisms.

Definition 1.2. A family $x^{(0)}, x^{(1)}, \dots, x^{(r-1)}$ of r distinct vectors in $\{0, 1\}^n$ is *focal* with focus $x^{(0)}$ if for every coordinate $i \in [n]$ at least $r - 2$ of the $r - 1$ entries $x_i^{(1)}, \dots, x_i^{(r-1)}$ are equal to $x_i^{(0)}$.

Thus, a focal family is a near-sunflower with the additional property that one of the vectors – the focus – is always in the majority. Unlike near-sunflowers, focal families are interesting already for $r = 3$. While sunflowers and focal families are both special kinds of near-sunflowers, they are not logically comparable to each other.

The two extremal functions corresponding to our definitions are:

$$\begin{aligned} g_r^{\text{ns}}(n) &= \max\{|\mathcal{F}| : \mathcal{F} \subseteq \{0, 1\}^n \text{ contains no near-sunflower of size } r\} \\ g_r^{\text{ff}}(n) &= \max\{|\mathcal{F}| : \mathcal{F} \subseteq \{0, 1\}^n \text{ contains no focal family of size } r\} \end{aligned}$$

It follows from the definitions that $g_r^{\text{ns}}(n) \leq g_r^{\text{ff}}(n)$. Our main result gives upper and lower bounds for these functions.

Theorem 1.3. *For $r \geq 3$ we have:*

- (a) $g_r^{\text{ns}}(n) \leq g_r^{\text{ff}}(n) \leq (r - 1)2^{\lceil \frac{(r-2)n}{r-1} \rceil}$.
- (b) *There exist positive constants c_r^{ns} and c_r^{ff} so that*

$$\begin{aligned} g_r^{\text{ns}}(n) &\geq c_r^{\text{ns}} \left(\frac{2}{(r+1)^{\frac{1}{r-1}}} \right)^n, \\ g_r^{\text{ff}}(n) &\geq c_r^{\text{ff}} \left(\frac{2}{r^{\frac{1}{r-1}}} \right)^n. \end{aligned}$$

In particular, for $r = 4$, our bounds (ignoring constants) are $2^{\frac{2n}{3}}$ from above and $(\frac{8}{5})^{\frac{n}{3}}$ and $2^{\frac{n}{3}}$ from below, for near-sunflowers and focal families, respectively. Families without near-sunflowers of

size 4 were previously studied (with different terminology) by the first author, Körner and Monti [1], settling a problem suggested by Sós in the late 80's. While their lower bound was the same as ours, their upper bound was roughly $2^{0.773n}$, with a proof based on Sauer's lemma. It is remarkable that our short and elementary proof improves their bound. We note that they also extended their result to $r > 4$, but with a different definition. Whereas our near-sunflowers allow an element to belong to 0, 1, $r - 1$ or r of the sets, their definition allows everything except one forbidden value.

We also remark that a concept analogous to our focal families, where instead of requiring “at least $r - 2$ ” in Definition 1.2 one requires “at least 1,” was studied in coding theory under the names separating codes (Cohen and Schaathun [5]) and frameproof codes (Blackburn [4]). This case is somewhat simpler and the bounds obtained in those papers coincide with ours for the special case $r = 3$.

More generally, our paper follows a long line of literature in extremal combinatorics, information theory and coding theory. The common thread is bounding the largest possible cardinality of a family of vectors of length n , so that for any r of them there exist coordinates displaying certain desirable patterns. For $r = 2$, Sperner's [18] classical theorem on antichains is a prime example. For $r = 3$ we mention the theorem of Erdős, Frankl and Füredi [9] on families in which no set is covered by the union of two others; the problem of cancellative families solved by Tolhuizen [19]; and a variety of related problems described by Körner [13]. For $r = 4$ there is Lindström's [15] theorem on determining two vectors from their sum modulo 2, and Körner and Simonyi's [14] bounds for two-different quadruples. For general r , we refer to the study of disjunctive codes (Dyachkov and Rykov [7]). In all of these problems, and many others, the cardinality of the largest family grows exponentially in n , but (with few exceptions) the asymptotic growth rate is not known. Our problems are no exception.

The proof of Theorem 1.3 is given in the next section. In Section 3 we adapt the definition and the bounds for focal families to vectors over larger alphabets, noting that the upper bound becomes essentially tight when the size of the alphabet exceeds n . We show in Section 4 that the upper bound in Theorem 1.3 can be improved if the family of vectors is closed under addition modulo 2 (i.e., forms a linear code). In Section 5 we consider one-sided focal families, where 0 and 1 entries are treated asymmetrically, and obtain corresponding bounds. Finally, in Section 6 we discuss the challenge of obtaining an exponential upper bound in terms of the uniformity k , and prove such a bound under a stronger condition.

2 Proof of Theorem 1.3

The upper bound. Let $\mathcal{F} \subseteq \{0, 1\}^n$ have cardinality $|\mathcal{F}| > (r - 1)2^{\lceil \frac{(r-2)n}{r-1} \rceil}$. We have to show that $\mathcal{F}$ contains a focal family of size r . Fix a partition $A_1, \dots, A_{r-1}$ of $[n]$ into $r - 1$ parts of size $|A_j| \geq \lfloor \frac{n}{r-1} \rfloor$ each. For a subset $S \subseteq [r - 1]$ with $|S| = r - 2$, say that a vector $x \in \mathcal{F}$ is *S-unique* if there is no other vector in $\mathcal{F}$ with the same projection on $\bigcup_{j \in S} A_j$. Since $|\bigcup_{j \in S} A_j| \leq \lceil \frac{(r-2)n}{r-1} \rceil$, for

a given S the number of S -unique vectors in $\mathcal{F}$ is at most $2^{\lceil \frac{(r-2)n}{r-1} \rceil}$. It follows from our assumption on $|\mathcal{F}|$ that there exists a vector $x^{(0)} \in \mathcal{F}$ which is not S -unique for any $S \subseteq [r-1]$ with $|S| = r-2$. This means that we can find vectors $x^{(1)}, \dots, x^{(r-1)} \in \mathcal{F} \setminus \{x^{(0)}\}$ so that each $x^{(j)}$ agrees with $x^{(0)}$ except possibly on coordinates in A_j . Note that $x^{(1)}, \dots, x^{(r-1)}$ are pairwise distinct, because if two of them were equal they would have to coincide with $x^{(0)}$. By construction, the subfamily $x^{(0)}, x^{(1)}, \dots, x^{(r-1)}$ is focal with focus $x^{(0)}$.

The lower bounds. As is common in such problems, we use random choice with alterations. We describe the argument for near-sunflowers, later pointing out how to adapt it to focal families.

We start by forming a random family $\mathcal{G} \subseteq \{0,1\}^n$ to which each vector $x \in \{0,1\}^n$ belongs, independently, with probability p (to be determined later). Then $\mathbb{E}(|\mathcal{G}|) = 2^n p$. Let $N_{\mathcal{G}}$ be a random variable counting the number of near-sunflowers of size r contained in $\mathcal{G}$. By removing at most $N_{\mathcal{G}}$ vectors from $\mathcal{G}$, we obtain a family $\mathcal{F}$ of cardinality at least $|\mathcal{G}| - N_{\mathcal{G}}$ which contains no near-sunflower of size r . By linearity of expectation, $\mathbb{E}(|\mathcal{F}|) \geq 2^n p - N_r^{\text{ns}} p^r$, where N_r^{ns} is the number of near-sunflowers of size r in $\{0,1\}^n$.

To estimate N_r^{ns} , note that the number of $r \times n$ binary matrices so that the number of 1 entries in each column is 0, 1, $r-1$ or r is $(2r+2)^n$. Since near-sunflowers correspond to such matrices with distinct rows, and the order of the rows is immaterial, it follows that $N_r^{\text{ns}} \leq \frac{1}{r!}(2r+2)^n$. Thus,

$$\mathbb{E}(|\mathcal{F}|) \geq 2^n p - \frac{1}{r!}(2r+2)^n p^r,$$

and choosing $p = c(\frac{1}{(r+1)^{\frac{1}{r-1}}})^n$ for a suitable $c = c(r) > 0$ yields $\mathbb{E}(|\mathcal{F}|) \geq c_r^{\text{ns}}(\frac{2}{(r+1)^{\frac{1}{r-1}}})^n$ for some $c_r^{\text{ns}} > 0$. Hence, there is a realization of $\mathcal{F}$ having at least this cardinality.

Moving to focal families, the argument is similar, but now we have to estimate the number N_r^{ff} of focal families of size r in $\{0,1\}^n$. The number of $r \times n$ binary matrices so that in each column the first entry is repeated at least $r-2$ times among the other entries is $(2r)^n$. Since focal families correspond to such matrices with distinct rows, and the order of the last $r-1$ rows is immaterial, it follows that $N_r^{\text{ff}} \leq \frac{1}{(r-1)!}(2r)^n$. Thus,

$$\mathbb{E}(|\mathcal{F}|) \geq 2^n p - \frac{1}{(r-1)!}(2r)^n p^r,$$

and choosing $p = c(\frac{1}{r^{\frac{1}{r-1}}})^n$ for a suitable $c = c(r) > 0$ yields $\mathbb{E}(|\mathcal{F}|) \geq c_r^{\text{ff}}(\frac{2}{r^{\frac{1}{r-1}}})^n$ for some $c_r^{\text{ff}} > 0$, as required. $\square$

3 Focal families over larger alphabets

Given any integer $q \geq 2$, Definition 1.2 can be applied verbatim to vectors in $[q]^n$ to define q -ary focal families. Let $g_r^{q\text{-ff}}(n)$ be the corresponding extremal function. A straightforward adaptation of the proof above yields the following version of Theorem 1.3 for q -ary focal families.

Theorem 3.1. *For $q \geq 2$ and $r \geq 3$ we have*

$$c_r^{q\text{-ff}} \left(\frac{q}{((q-1)(r-1)+1)^{\frac{1}{r-1}}} \right)^n \leq g_r^{q\text{-ff}}(n) \leq (r-1)q^{\lceil \frac{(r-2)n}{r-1} \rceil}$$

for some positive constant $c_r^{q\text{-ff}}$.

When $q \geq n$ and q is a prime power, we can replace the probabilistic lower bound by a constructive one which matches (up to a constant factor depending on r) the upper bound.

Proposition 3.2. *If $q \geq n$ and q is a prime power then*

$$g_r^{q\text{-ff}}(n) \geq q^{\lceil \frac{(r-2)n}{r-1} \rceil}.$$

Proof. The Reed-Solomon code with suitable parameters gives the desired lower bound. For completeness, we describe the construction. We identify the elements of the finite field $\mathbb{F}_q$ with the q symbols in our alphabet. We choose and fix n distinct elements $a_1, a_2, \dots, a_n \in \mathbb{F}_q$, and identify them with the coordinates $1, 2, \dots, n$. There are $q^{\lceil \frac{(r-2)n}{r-1} \rceil}$ polynomials $p(x)$ of degree less than $\lceil \frac{(r-2)n}{r-1} \rceil$ over $\mathbb{F}_q$. With every such polynomial we associate the vector $(p(a_1), p(a_2), \dots, p(a_n))$, which gives a family $\mathcal{F}$ of q -ary vectors of length n , with $|\mathcal{F}| = q^{\lceil \frac{(r-2)n}{r-1} \rceil}$.

We claim that $\mathcal{F}$ contains no focal family of size r . Indeed, suppose that $x^{(0)}, x^{(1)}, \dots, x^{(r-1)} \in \mathcal{F}$ form such a family with focus $x^{(0)}$. Then by the pigeonhole principle, some $x^{(j)}$, $j \in [r-1]$, has to agree with $x^{(0)}$ on at least $\lceil \frac{(r-2)n}{r-1} \rceil$ coordinates. This means that the corresponding polynomials $p^{(j)}$ and $p^{(0)}$ agree on at least $\lceil \frac{(r-2)n}{r-1} \rceil$ elements of $\mathbb{F}_q$. But this is impossible, as they are distinct polynomials of degree less than $\lceil \frac{(r-2)n}{r-1} \rceil$. $\square$

4 Improved upper bound in the linear case

While Proposition 3.2 shows that our upper bound is essentially tight when $q \geq n$, we believe that for $q = 2$ and large n it is not. To support this belief, we show here that the upper bound can be significantly improved if we restrict attention to families of binary vectors which are linear codes (i.e., closed under addition modulo 2). This can be done for any value of r , but for simplicity and concreteness of the bound we do it for $r = 4$.

We are going to use a known bound on the tradeoff between cardinality and minimum Hamming distance in a family $\mathcal{F}$ of binary vectors of length n . Recall that, by the linear programming bound (McEliece, Rodemich, Rumsey and Welch [16]), if the Hamming distance between any two distinct vectors in $\mathcal{F}$ is greater than δn , then

$$|\mathcal{F}| \leq 2^{(h(\frac{1}{2} - \sqrt{\delta(1-\delta)}) + o(1))n},$$

where $h(x)$ is the binary entropy function defined by

$$h(x) = -x \log_2 x - (1-x) \log_2 (1-x).$$

We first prove the following theorem, which does not require linearity.

Theorem 4.1. *Let $\mathcal{F}$ be a family of at least $2^{0.44n}$ subsets of $[n]$, where n is large enough. Then there exist three pairs of distinct sets in $\mathcal{F}$ such that their symmetric differences $A \triangle B$, $C \triangle D$ and $E \triangle F$ are pairwise disjoint.*

Proof. We apply the above-mentioned bound repeatedly. First, a calculation shows that for $\delta = 0.213$, we have $h(\frac{1}{2} - \sqrt{\delta(1-\delta)}) < 0.44$. As $|\mathcal{F}| \geq 2^{0.44n}$ and n is large, the bound implies the existence of distinct sets $A, B \in \mathcal{F}$ with $|A \triangle B| \leq 0.213n$. Next, by the pigeonhole principle, we can find at least $\frac{|\mathcal{F}|}{2^{|A \triangle B|}}$ sets in $\mathcal{F}$ having the same intersection with $A \triangle B$. Let $\mathcal{F}'$ be the family obtained by restricting these sets to $[n] \setminus (A \triangle B)$. A calculation shows that for $\delta' = 0.287$, we have $h(\frac{1}{2} - \sqrt{\delta'(1-\delta')}) < 0.28$. As $|\mathcal{F}'| \geq 2^{0.44n - |A \triangle B|} > 2^{0.28(n - |A \triangle B|)}$ and n is large, the bound implies the existence of distinct sets $C, D \in \mathcal{F}$ such that $(C \triangle D) \cap (A \triangle B) = \emptyset$ and $|C \triangle D| \leq 0.287(n - |A \triangle B|)$. Now $|A \triangle B| + |C \triangle D| < 0.44n$, and again by the pigeonhole principle we can find two distinct sets $E, F \in \mathcal{F}$ having the same intersection with $(A \triangle B) \cup (C \triangle D)$, which completes the proof. $\square$

Corollary 4.2. *Let $\mathcal{F}$ be a linear subspace of $\{0, 1\}^n$ of dimension at least $0.44n$, where n is large enough. Then $\mathcal{F}$ contains a focal family of size 4.*

Proof. Viewing $\mathcal{F}$ as a family of subsets of $[n]$, it is closed under symmetric difference. Hence the theorem yields three pairwise disjoint non-empty sets $X^{(1)}, X^{(2)}, X^{(3)} \in \mathcal{F}$. Taking the empty set as the focus $X^{(0)}$, we obtain a focal family of size 4. $\square$

5 One-sided focal families

The requirement defining a focal family may be separated into two one-sided requirements as follows.

Definition 5.1. Let $b \in \{0, 1\}$. A family $x^{(0)}, x^{(1)}, \dots, x^{(r-1)}$ of r distinct vectors in $\{0, 1\}^n$ is b -focal with focus $x^{(0)}$ if for every coordinate $i \in [n]$ such that $x_i^{(0)} = b$, at least $r - 2$ of the $r - 1$ entries $x_i^{(1)}, \dots, x_i^{(r-1)}$ are equal to b .

The corresponding extremal functions for $b = 0, 1$ are:

$$g_r^{b\text{-ff}}(n) = \max\{|\mathcal{F}| : \mathcal{F} \subseteq \{0, 1\}^n \text{ contains no } b\text{-focal family of size } r\}$$

It will be convenient to study the extremal questions first for k -uniform families. Let $\binom{[n]}{k}$ be the family of all k -element subsets of $[n]$. For $b = 0, 1$ let:

$$g_r^{b\text{-ff}}(n, k) = \max\{|\mathcal{F}| : \mathcal{F} \subseteq \binom{[n]}{k} \text{ contains no } b\text{-focal family of size } r\}$$

Since $\mathcal{H}$ is 0-focal if and only if $\{[n] \setminus A : A \in \mathcal{H}\}$ is 1-focal, we have $g_r^{0\text{-ff}}(n) = g_r^{1\text{-ff}}(n)$ and $g_r^{0\text{-ff}}(n, k) = g_r^{1\text{-ff}}(n, n - k)$. So we only need to study these questions for one value of b .

Theorem 5.2. For $r \geq 3$ and $0 \leq k \leq n$ we have

$$g_r^{1\text{-ff}}(n, k) \leq (r-1) \frac{\binom{n}{\lceil \frac{(r-2)k}{r-1} \rceil}}{\binom{k}{\lceil \frac{(r-2)k}{r-1} \rceil}}.$$

Proof. Let $\mathcal{F}$ be a family of k -element subsets of $[n]$ containing no 1-focal family of size r . For a set $A \in \mathcal{F}$, we say that a set S is an *own-subset* of A if $S \subseteq A$ and $S \not\subseteq B$ for any $B \in \mathcal{F} \setminus \{A\}$.

Consider an arbitrary $(r-1)$ -tuple $A_1, \dots, A_{r-1}$ of pairwise disjoint $\lfloor \frac{k}{r-1} \rfloor$ -element subsets of a set $A \in \mathcal{F}$. If for every $j \in [r-1]$ there exists a set $B_j \in \mathcal{F} \setminus \{A\}$ such that $A \setminus A_j \subseteq B_j$, then the sets $A, B_1, \dots, B_{r-1}$ form a 1-focal family of size r with focus A , contradicting our assumption on the family $\mathcal{F}$. Hence there exists $j \in [r-1]$ so that $A \setminus A_j$ is an own-subset of A .

We claim that for a fixed set $A \in \mathcal{F}$, the probability that a uniformly random $\lceil \frac{(r-2)k}{r-1} \rceil$ -element subset S of A is an own-subset is at least $\frac{1}{r-1}$. Indeed, consider the following two-step random process. First, choose uniformly at random an $(r-1)$ -tuple $A_1, \dots, A_{r-1}$ of pairwise disjoint $\lfloor \frac{k}{r-1} \rfloor$ -element subsets of A . Second, choose uniformly at random a value $j \in [r-1]$ and let $S = A \setminus A_j$. Clearly, the resulting S is uniformly distributed over the $\lceil \frac{(r-2)k}{r-1} \rceil$ -element subsets of A . Conditional on the choice in the first step, the argument in the previous paragraph implies that the probability that S is an own-subset of A is at least $\frac{1}{r-1}$. As this holds for any outcome of the first step, it also holds unconditionally.

Thus, with each $A \in \mathcal{F}$ we can associate a family of at least $\frac{1}{r-1} \binom{k}{\lceil \frac{(r-2)k}{r-1} \rceil}$ own-subsets of A of size $\lceil \frac{(r-2)k}{r-1} \rceil$. The disjoint union of these families over all $A \in \mathcal{F}$ is contained in $\binom{[n]}{\lceil \frac{(r-2)k}{r-1} \rceil}$, implying that $|\mathcal{F}| \cdot \frac{1}{r-1} \binom{k}{\lceil \frac{(r-2)k}{r-1} \rceil} \leq \binom{n}{\lceil \frac{(r-2)k}{r-1} \rceil}$. It follows that $|\mathcal{F}| \leq (r-1) \frac{\binom{n}{\lceil \frac{(r-2)k}{r-1} \rceil}}{\binom{k}{\lceil \frac{(r-2)k}{r-1} \rceil}}$, as claimed. $\square$

Corollary 5.3. For $r \geq 3$ and $b = 0, 1$ we have

$$g_r^{b\text{-ff}}(n) \leq (r-1) \sum_{k=0}^n \frac{\binom{n}{\lceil \frac{(r-2)k}{r-1} \rceil}}{\binom{k}{\lceil \frac{(r-2)k}{r-1} \rceil}} = (1 + \frac{r-2}{(r-1)^{\frac{r-1}{r-2}}} + o(1))^n.$$

Proof. As pointed out above, it suffices to treat the case $b = 1$. Let $\mathcal{F}$ be a family of subsets of $[n]$ containing no 1-focal family of size r . Then $|\mathcal{F}| = \sum_{k=0}^n |\mathcal{F} \cap \binom{[n]}{k}|$, and applying the theorem to the families $\mathcal{F} \cap \binom{[n]}{k}$ yields the upper bound in summation form.

To obtain the asymptotic expression for the sum, we first use Stirling's formula to approximate $\binom{k}{\lceil \frac{(r-2)k}{r-1} \rceil}$ up to a factor of order $\sqrt{k}$ by $(\frac{(r-1)^{r-1}}{(r-2)^{r-2}})^{\frac{k}{r-1}}$. Plugging this approximation in the sum gives

$$\sum_{k=0}^n \left(\binom{n}{\lceil \frac{(r-2)k}{r-1} \rceil} \right) \left(\frac{r-2}{(r-1)^{\frac{r-1}{r-2}}} \right)^{\frac{(r-2)k}{r-1}}$$

which, by the binomial formula, is $\Theta((1 + \frac{r-2}{(r-1)^{\frac{r-1}{r-2}}})^n)$. $\square$

In the case $r = 3$, a 1-focal family is a triple of distinct sets satisfying $A \subseteq B \cup C$, and Corollary 5.3 reproduces the bound of $(\frac{5}{4} + o(1))^n$ obtained by Erdős, Frankl and Füredi [9] for the maximum possible cardinality of families not containing such triples. In the case $r = 4$, a 1-focal family is a 4-tuple of distinct sets satisfying $A \subseteq (B \cap C) \cup (B \cap D) \cup (C \cap D)$, or equivalently $A \subseteq (B \cup C) \cap (B \cup D) \cap (C \cup D)$. We get an upper bound of roughly $2^{0.47n}$ for the corresponding extremal problem.

Lower bounds on the extremal functions $g_r^{b\text{-ff}}(n, k)$ and $g_r^{b\text{-ff}}(n)$ may be obtained, as above, by random choice with alterations. Here, however, one should start with a random subfamily of $\binom{[n]}{k}$ instead of $\{0, 1\}^n$. Optimizing the bounds requires rather messy calculations, which we omit.

6 Bounds in terms of the set size

We turn our attention now to bounding the cardinality of a family of k -element sets (on a ground set of any size) not containing any near-sunflower of size r . Note that this question does not make sense for focal families, because we may take arbitrarily many pairwise disjoint k -element sets, avoiding focal families of size 3. With respect to near-sunflowers, however, any upper bound on the cardinality of a k -uniform family not containing a sunflower of size r automatically applies to our question, too; in particular, the recent bound of Alweiss, Lovett, Wu and Zhang [3] of order $(\log k)^{(1+o(1))k}$. Can this be improved for near-sunflowers?

Conjecture 6.1. *Let $r \geq 4$, and let $\mathcal{F}$ be a family of k -element sets which contains no near-sunflower of size r . Then $|\mathcal{F}| \leq C^k$, where C is a constant depending only on r .*

This is a weaker version of the Erdős-Rado sunflower conjecture. In view of the fame and difficulty of the latter, this weakening may turn out to be a more accessible goal. But we have not been able to make progress, even for $r = 4$.

We do have an upper bound of the desired exponential form under a stronger condition. Saying that a 4-tuple of distinct sets A, B, C, D is not a near-sunflower can be expressed as follows: there is a way to partition $\{A, B, C, D\}$ into two pairs with intersecting symmetric differences. A natural strengthening is to require this for every pairing of A, B, C, D . Let $\mathcal{F}$ be a family of sets so that for any (ordered) four distinct sets $A, B, C, D \in \mathcal{F}$ we have $(A \triangle B) \cap (C \triangle D) \neq \emptyset$. Körner and Simonyi [14] proved that if all sets are subsets of an n -element ground set then $|\mathcal{F}| \leq 1.217^n$ for large n . But here we are interested in such families that are k -uniform on any ground set. Fixing A, B , the condition implies that any C and D must differ within $A \triangle B$, which has at most $2k$ elements, easily giving $|\mathcal{F}| \leq 2^{2k}$. The following theorem improves this bound.

Theorem 6.2. *Let $\mathcal{F}$ be a family of k -element sets so that for any (ordered) four distinct sets $A, B, C, D \in \mathcal{F}$ we have $(A \triangle B) \cap (C \triangle D) \neq \emptyset$. Then $|\mathcal{F}| \leq 2.148^k$ for large enough k .*

Proof. Fix two sets $A, B \in \mathcal{F}$ so that $|A \cap B| = t$ maximizes the intersection size over all pairs of distinct sets in $\mathcal{F}$.

For any set E , denote by $[E]_t$ a set which is E itself if $|E| \leq t$, and otherwise it is an arbitrarily chosen $(t+1)$ -element subset of E . Let $\binom{A \triangle B}{\leq t+1}$ be the family of all subsets of $A \triangle B$ of size at most $t+1$. Define a mapping $f : \mathcal{F} \setminus \{A, B\} \rightarrow \binom{A \triangle B}{\leq t+1}$ by $f(C) = [C \cap (A \triangle B)]_t$.

We check that f is injective. Let C and D be two distinct sets in $\mathcal{F} \setminus \{A, B\}$. We have to show that $[C \cap (A \triangle B)]_t \neq [D \cap (A \triangle B)]_t$. If the sets on both sides have size at most t , then $C \cap (A \triangle B) \neq D \cap (A \triangle B)$ follows from $(A \triangle B) \cap (C \triangle D) \neq \emptyset$. If both $[C \cap (A \triangle B)]_t$ and $[D \cap (A \triangle B)]_t$ have size $t+1$, they cannot be equal since that would imply $|C \cap D| \geq t+1$, contradicting the maximality of t . Finally, if one of them has size at most t and the other has size $t+1$, they are obviously not equal. This implies that

$$|\mathcal{F}| - 2 \leq \sum_{j=0}^{t+1} \binom{2(k-t)}{j}.$$

For large k , we want to bound the right-hand side from above by C^k for some $C < 2.148$. Let us write $x = \frac{t}{2(k-t)}$. If $x \geq \frac{1}{2}$ then $2(k-t) \leq k$ and the sum is bounded by 2^k . Thus we may assume that $x < \frac{1}{2}$ and approximate the sum by $2^{2(k-t)h(x)} = 2^{\frac{2h(x)}{1+2x}k}$, where $h(x)$ is the binary entropy function. Routine calculations show that the maximum of $2^{\frac{2h(x)}{1+2x}}$ is attained when $x = (1-x)^3$ and its value is less than 2.148. $\square$

As in all these problems, the probabilistic method can be used to show the existence of a k -uniform family $\mathcal{F}$ with pairwise intersecting symmetric differences, so that $|\mathcal{F}|$ is exponential in k . Our argument gives $|\mathcal{F}| \approx 1.25^k$, we omit the details.

References

- [1] Noga Alon, János Körner, and Angelo Monti. String quartets in binary. *Combin. Probab. Comput.*, 9(5):381–390, 2000.
- [2] Noga Alon, Amir Shpilka, and Christopher Umans. On sunflowers and matrix multiplication. In *2012 IEEE 27th Conference on Computational Complexity—CCC 2012*, pages 214–223. IEEE Computer Soc., Los Alamitos, CA, 2012.
- [3] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. *arXiv preprint 1908.08483*, 2019.
- [4] Simon R. Blackburn. Frameproof codes. *SIAM J. Discrete Math.*, 16(3):499–510, 2003.
- [5] G. D. Cohen and H. G. Schaathun. New upper bounds on separating codes. In *Proc. 2003 Int. Conf. Telecommunications*, pages 1990–1994. 2003.
- [6] W. A. Deuber, P. Erdős, D. S. Gunderson, A. V. Kostochka, and A. G. Meyer. Intersection statements for systems of sets. *J. Combin. Theory Ser. A*, 79(1):118–132, 1997.

- [7] A. G. Dyachkov and V. V. Rykov. Bounds on the length of disjunctive codes. *Problemy Peredachi Informatsii*, 18(3):7–13, 1982.
- [8] Jordan S. Ellenberg and Dion Gijswijt. On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Ann. of Math. (2)*, 185(1):339–343, 2017.
- [9] P. Erdős, P. Frankl, and Z. Füredi. Families of finite sets in which no set is covered by the union of two others. *J. Combin. Theory Ser. A*, 33(2):158–166, 1982.
- [10] P. Erdős and R. Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.
- [11] P. Erdős and E. Szemerédi. Combinatorial properties of systems of sets. *J. Combinatorial Theory Ser. A*, 24(3):308–313, 1978.
- [12] G. Hegedűs. An improved upper bound for the size of a sunflower-free family. *Acta Math. Hungar.*, 155(2):431–438, 2018.
- [13] János Körner. On the extremal combinatorics of the Hamming space. *J. Combin. Theory Ser. A*, 71(1):112–126, 1995.
- [14] János Körner and Gábor Simonyi. Separating partition systems and locally different sequences. *SIAM J. Discrete Math.*, 1(3):355–359, 1988.
- [15] Bernt Lindström. Determination of two vectors from the sum. *J. Combinatorial Theory*, 6:402–407, 1969.
- [16] Robert J. McEliece, Eugene R. Rodemich, Howard Rumsey, Jr., and Lloyd R. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Inform. Theory*, IT-23(2):157–166, 1977.
- [17] Eric Naslund and Will Sawin. Upper bounds for sunflower-free sets. *Forum Math. Sigma*, 5:Paper No. e15, 10, 2017.
- [18] Emanuel Sperner. Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, 27(1):544–548, 1928.
- [19] Ludo M. Tolhuizen. New rate pairs in the zero-error capacity region of the binary multiplying channel without feedback. *IEEE Trans. Inform. Theory*, IT-46(3):1043–1046, 2000.