Sunflowers in set systems of bounded dimension

Jacob Fox, János Pach, Andrew Suk

Abstract

Given a family $\mathcal{F}$ of k-element sets, $S_1, \ldots, S_r \in \mathcal{F}$ form an *r-sunflower* if $S_i \cap S_j = S_{i'} \cap S_{j'}$ for all $i \neq j$ and $i' \neq j'$. According to a famous conjecture of Erdős and Rado (1960), there is a constant $c = c(r)$ such that if $|\mathcal{F}| \geq c^k$, then $\mathcal{F}$ contains an r-sunflower.

We come close to proving this conjecture for families of bounded *Vapnik-Chervonenkis dimension*, VC-dim$(\mathcal{F}) \leq d$. In this case, we show that r-sunflowers exist under the slightly stronger assumption $|\mathcal{F}| \geq 2^{10k(dr)^{2\log^* k}}$. Here, $\log^*$ denotes the iterated logarithm function.

We also verify the Erdős-Rado conjecture for families $\mathcal{F}$ of bounded *Littlestone dimension* and for some geometrically defined set systems.

1 Introduction

An *r-sunflower* is a collection of r sets whose pairwise intersections are the same. That is, r distinct sets $S_1, \ldots, S_r$ form an r-sunflower if $S_i \cap S_j = S_{i'} \cap S_{j'}$ for all $i \neq j$ and $i' \neq j'$. The term was coined by Deza and Frankl [11]. For brevity, a k-element set is called a *k-set*.

Let $f_r(k)$ be the minimum positive integer m such that every family of k-sets whose size is at least m contains r members that form an r-sunflower. Erdős and Rado [13] proved that $f_r(k) \leq k!(r-1)^k$. The Erdős–Rado "sunflower conjecture" states that there is a constant $C = C(r)$ depending only on r such that $f_r(k) \leq C^k$. Over the years, some small improvements have been made on the upper bound $k!(r-1)^k$, see [1, 18]. Very recently, a breakthrough has been achieved by Alweiss, Lovett, Wu, and Zhang [5], who proved that

$$f_r(k) \leq (cr^3 \log k \log\log k)^k,$$

Mathematics Subject Classification (2020): 05D05, 05D10, 52C10, 05D40, 03C13

Fox: Supported by a Packard Fellowship and by NSF Awards DMS-1800053 and DMS-2154169.

Pach: Supported by NKFIH grants K-131529, KKP-133864, Austrian Science Fund Z 342-N31, Ministry of Education and Science of the Russian Federation MegaGrant No. 075-15-2019-1926, ERC Advanced Grant "GeoScape."

Suk: Supported an NSF CAREER award, NSF award DMS-1952786, and an Alfred Sloan Fellowship.

where c is an absolute constant. For an alternative proof of this result, using Shannon capacities, see [22]. Some weaker versions of the conjecture are discussed in [3, 14, 20].

The aim of this note is to study the Erdős-Rado sunflower conjecture for families of bounded dimension. Apart from set systems realized in low-dimensional Euclidean spaces, we consider two additional notions of dimension: the *Vapnik-Chervonenkis dimension* (in short, VC-dimension) and the *Littlestone dimension* (LS-dimension), introduced in [29] and [19], respectively. Both are important combinatorial parameters that measure the complexity of graphs and hypergraphs, and play important roles in statistics, algebraic geometry, PAC learning, and in model theory. There is a growing body of results in extremal combinatorics and Ramsey theory which give much better bounds or stronger conclusions under the additional assumption of bounded dimension (see [15, 16]).

Given a family of sets $\mathcal{F}$ with ground set V, the *VC-dimension* of $\mathcal{F}$, denoted by VC-dim($\mathcal{F}$), is the maximum d for which there exists a d-element set $S \subset V$ such that for every subset $B \subset S$, one can find a member $A \in \mathcal{F}$ with $A \cap S = B$. In this case, we say that S is *shattered* by $\mathcal{F}$.

Let $f_r^d(k)$ denote the least positive integer m such that every family $\mathcal{F}$ of k-sets with $|\mathcal{F}| \geq m$ and VC-dim($\mathcal{F}$) $\leq d$ contains an r-sunflower. Clearly, we have $f_r^d(k) \leq f_r(k)$, and the Erdős-Rado sunflower conjecture implies the following weaker conjecture.

Conjecture 1.1. *For $d \geq 1$ and $r \geq 3$, there is a constant $C = C(d, r)$ such that $f_r^d(k) \leq C^k$.*

It is not difficult to see that, even for $d = 1$, the function $f_r^1(k)$ grows at least exponentially in k. More precisely, we have $f_r^1(k) > (r-1)^{k-1}$. Indeed, consider a rooted complete $(r-1)$-ary tree T with the root on level 0 and with $(r-1)^{k-1}$ leaves on level $k-1$. Let $\mathcal{F}$ be the family of k-sets consisting of the vertex sets of the root-to-leaf paths in T. Obviously, $\mathcal{F}$ does not contain any r-sunflower, and its VC-dimension is at most 1.

More generally, we have the recursive lower bound

$$f_r^d(k_1 + k_2) > (f_r^d(k_1) - 1)(f_r^d(k_2) - 1).$$

Indeed, for $i = 1, 2$, let $\mathcal{F}_i$ be a family of k_i-sets of size $f_r^d(k_i) - 1$ with VC-dimension at most d and without any r-sunflower. For each set S in $\mathcal{F}_1$, make a new copy of $\mathcal{F}_2$ and add S to each set in $\mathcal{F}_2$. The ground set of copies of $\mathcal{F}_2$ are pairwise disjoint for distinct sets of $\mathcal{F}_1$. The resulting set system $\mathcal{F}$ is $(k_1 + k_2)$-uniform with size $(f_r^d(k_1) - 1)(f_r^d(k_2) - 1)$, VC-dimension at most d, and has no r-sunflower. This implies that if $f_r(k') > C^{k'} + 1$ for some k' and C, then there is d depending on k' such that for all sufficiently large k, $f_r^d(k) > C^k$. Thus, any exponential lower bound for the classical sunflower problem (with unbounded VC-dimension) can be achieved by a construction with bounded (but sufficiently large) VC-dimension.

Using a result of Ding, Seymour, and Winkler [12], we settle Conjecture 1.1 for families of k-sets with VC-dimension $d = 1$.

Theorem 1.2. *For integers $r \geq 3$ and $k \geq 1$, every family of k-sets with VC-dimension $d = 1$ and cardinality at least r^{10k} has an r-sunflower. That is, we have*

$$f_r^1(k) \leq r^{10k}.$$

Let $\log^* k$ denote the *iterated logarithm* of k, *i.e.*, the minimum i for which the i times iterated logarithm of k satisfies $\log^{(i)} k \leq 2$. All logarithms used in this note are of base 2.

For $d \geq 2$, our upper bound on $f_r^d(k)$ is not far from the one stated in Conjecture 1.1.

Theorem 1.3. *For integers $d, k, r \geq 2$, every family of k-sets with VC-dimension at most d and cardinality at least $2^{10k(dr)^{2\log^* k}}$ has an r-sunflower. In notation,*

$$f_r^d(k) \leq 2^{10k(dr)^{2\log^* k}}.$$

The *Littlestone dimension* of $\mathcal{F} \subseteq 2^V$ is defined as follows. Consider a rooted complete binary tree T_d, with the root at level 0 and with 2^d leaves at the last level. Let the leaves of T_d be labeled by sets in $\mathcal{F}$, and all other vertices by elements of V. We say that T_d is *shattered* by $\mathcal{F}$ if for every root-to-leaf path with labels $v_0, v_1, \ldots, v_{d-1}, F$, we have $v_i \in F$ if and only if the $(i+1)$st vertex along the path is the left-child of v_i, for all $0 \leq i < d$. The Littlestone dimension of $\mathcal{F}$, denoted by LS-dim($\mathcal{F}$), is the largest d for which there is a labeling of T_d which is shattered by $\mathcal{F}$.

Obviously, we have VC-dim($\mathcal{F}$) $\leq$ LS-dim($\mathcal{F}$), because if the $S = \{s_0, \ldots, s_{d-1}\} \subseteq V$ is shattered by $\mathcal{F}$, then the labeling of T_d in which all vertices at level i are labeled by s_i, $0 \leq s_i < d$, and the leaves by the corresponding sets in $\mathcal{F}$ with the appropriate intersection with S, is also shattered by $\mathcal{F}$.

Let $h_r^d(k)$ denote the least positive integer m such that every family $\mathcal{F}$ of k-sets with $|\mathcal{F}| \geq m$ and LS-dim($\mathcal{F}$) $\leq d$ contains an r-sunflower. Since the Littlestone dimension of a set system is at least as large as its VC-dimension, we have

$$h_r^d(k) \leq f_r^d(k) \leq f_r(k).$$

It turns out that $h_r^d(k)$, as a function of k, grows much more slowly than $f_r^d(k)$. Its growth rate is only polynomial in k, albeit the degree of this polynomial depends on d.

Theorem 1.4. *For positive integers d, r, k, every family of k-sets with LS-dimension at most d and cardinality at least $(rk)^d$ has an r-sunflower. Using our notation, we have*

$$h_r^d(k) \leq (rk)^d.$$

On the other hand, for integers $d, r \geq 3$, and $k \geq 4d$, we have

$$h_r^d(k) \geq (rk/d)^{d-o(d)},$$

where the $o(d)$ term goes to 0 as $d \to \infty$.

For several geometrically defined set systems, one can verify the sunflower conjecture by exploring the special properties of the underlying configurations.

A collection $\mathcal{D}$ of Jordan regions in the plane is called a family of *pseudo-disks* if the boundaries of any two members in $\mathcal{D}$ intersect in at most two points. For simplicity, we will assume that $\mathcal{D}$ is in *general position*, that is, no point lies on the boundary of three regions and no two regions are touching. It is well known that the VC-dimension of the set system obtained by restricting $\mathcal{D}$ to V is at most 3 (see [8]) and, hence, Theorem 1.3 applies. However, in this case, we can verify the sunflower conjecture.

Theorem 1.5. *Let V be a planar point set and let $\mathcal{D} = \{D_1, \ldots, D_N\}$ be a family of pseudo-disks such that the size of every set $S_i = D_i \cap V$ is equal to k. If $N \geq (500 + r)^{900k}$, where $r > 2$, then there are r distinct sets $S_{i_1}, \ldots, S_{i_r}$ that form an r-sunflower.*

Our paper is organized as follows. Sections 2 and 3 contain the proofs of Theorems 1.2 and 1.3, respectively. Theorem 1.4 about set systems of bounded Littlestone dimension is established in Section 4. Section 5 is devoted to low-dimensional geometric instances of the sunflower conjecture, while the last section contains some concluding remarks.

For the clarity of presentation, throughout this paper we make no attempt to optimize the absolute constants occurring in the statements.

2 VC-dimension 1 – Proof of Theorem 1.2

Given a family $\mathcal{F}$ of subsets of a ground set V, as usual, let $\nu(\mathcal{F})$ denote the *packing number* of $\mathcal{F}$, *i.e.,* the maximum number of pairwise disjoint members of $\mathcal{F}$. Also, let $\tau(\mathcal{F})$ be the *transversal number* of $\mathcal{F}$, *i.e.*, the minimum number of elements that can be selected from V such that every member of $\mathcal{F}$ contains at least one of them. Finally, let $\lambda(\mathcal{F})$ denote the maximum integer l such that there are l sets $S_1, \ldots, S_l \in \mathcal{F}$ with the property that for any $1 \leq i < j \leq l$, there is $v = v_{ij} \in S_i \cap S_j$ such that $v \notin S_t$ for $t \in [m] \setminus \{i, j\}$. It is easy to verify that $\lambda(\mathcal{F})$ is at least as large as the VC-dimension of the set system (hypergraph) $\mathcal{F}^*$ *dual* to $\mathcal{F}$.

We need the following result of Ding, Seymour, and Winkler [12] which bounds the transversal number of $\mathcal{F}$ in terms of its packing number and $\lambda(\mathcal{F})$.

Lemma 2.1 (Ding, Seymour, Winkler)**.** *Let $\mathcal{F}$ be a set system with ground set V, and let $\nu(\mathcal{F}) = \nu, \tau(\mathcal{F}) = \tau$ and $\lambda(\mathcal{F}) = \lambda$. Then we have*

$$\tau \leq 11\lambda^2(\lambda + \nu + 3)\binom{\lambda + \nu}{\lambda}^2.$$

Notice that VC-dim$(\mathcal{F}) = 1$ implies that $\lambda(\mathcal{F}) \leq 3$. Hence, Theorem 1.2 is an immediate corollary to the following result.

Theorem 2.2. *Let $r \geq 3$ and let $\mathcal{F}$ be a family of k-sets with $\lambda(\mathcal{F}) = \lambda$ which does not contain an r-sunflower. Then we have $|\mathcal{F}| \leq (\lambda + r)^{6\lambda k}$.*

Proof. We proceed by induction on k. The base case $k = 1$ follows from the trivial bound $|\mathcal{F}| \leq r - 1$. The induction hypothesis is that the bound holds for families of $(k-1)$-sets. For the inductive step, let $\mathcal{F} \subseteq 2^V$ be a family of k-sets with no r-sunflower. In particular, $\mathcal{F}$ has no r disjoint members, so that $\nu(\mathcal{F}) < r$. By Lemma 2.1,

$$\begin{aligned}\tau(\mathcal{F}) &\leq 11\lambda^2(\lambda + r + 3)\binom{\lambda + r}{\lambda}^2 \leq 11\lambda^2(\lambda + r + 3)(\lambda + r)^{2\lambda}(\lambda!)^{-2} \\ &\leq 11(\lambda + r + 3)(\lambda + r)^{2\lambda} \leq 20(\lambda + r)^{2\lambda+1}.\end{aligned}$$

Therefore, there is $v \in V$ incident to at least $|\mathcal{F}|/\tau(\mathcal{F}) \geq |\mathcal{F}|/\left(20(\lambda + r)^{2\lambda+1}\right)$ members of $\mathcal{F}$.

Let $\mathcal{F}' = \{S \setminus \{v\} \colon S \in \mathcal{F}, v \in S\}$. Then we have $|\mathcal{F}'| \geq |\mathcal{F}|/\left(20(\lambda + r)^{2\lambda+1}\right)$, $\lambda(\mathcal{F}') \leq \lambda(\mathcal{F})$, and $\mathcal{F}'$ does not contain any r-sunflower. By the induction hypothesis, we have $|\mathcal{F}'| \leq (\lambda + r)^{6\lambda(k-1)}$. Thus, we obtain

$$|\mathcal{F}| \leq 20(\lambda + r)^{2\lambda+1}|\mathcal{F}'| \leq 20(\lambda + r)^{2\lambda+1}(\lambda + r)^{6\lambda(k-1)} \leq (\lambda + r)^{6\lambda k},$$

as required. ∎

3 Bounded VC-dimension – Proof of Theorem 1.3

In this section, we prove Theorem 1.3, which is the main result of this paper. We need the following lemma due to Sauer [23], Shelah [25], Perles, and, in a slightly weaker form, to Vapnik and Chervonenkis [29]. See also [21, 17].

Lemma 3.1 (Sauer, Shelah, Perles)**.** *Let $\mathcal{F}$ be a set system with ground set V and VC-dimension at most d. Then we have $|\mathcal{F}| \leq \sum_{i=0}^{d} \binom{|V|}{i}$.*

Before turning to the proof, we need to discuss some closely related variants of the sunflower problem.

First, we could ask the same question for *multifamilies* of sets, that is, for collections of not necessarily distinct sets. Let $g_r(k)$ be the minimum positive integer m such that every multifamily of k-sets of size m contains an r-sunflower. It is an easy exercise to prove that $g_r(k) = (r-1)f_r(k) + 1$.

Analogously, for any $d \geq 1$, let $g_r^d(k)$ be the minimum positive integer m such that every multifamily of k-sets of size m with *VC-dimension at most d* contains an r-sunflower. We similarly have $g_r^d(k) = (r-1)f_r^d(k) + 1$.

To obtain upper bounds for $g_r^d(k)$ and $f_r^d(k)$, we define the following related function. Let $\alpha_r^d(k)$ denote the maximum α such that for every nonempty multifamily $\mathcal{F}$ of k-sets with VC-dimension at most d, if we select r members uniformly at random from $\mathcal{F}$ with replacement, the probability that they have pairwise equal intersections is at least α.

Next, notice that the value of $f_r(k)$ remains the same if we change the definition from families of k-sets to families of sets with *at most* k elements.

Indeed, this can be achieved by adding distinct "dummy" vertices to each set of size smaller than k so that it will have size exactly k. The same holds for the functions $f_r^d(k)$, $g_r(k)$, $g_r^d(k)$, and $\alpha_r^d(k)$ because adding dummy vertices does not affect the VC-dimension of the family.

Considering a family of VC-dimension d which consists of $f_r^d(k)-1$ sets of size k and contains no r-sunflower, we immediately obtain the following upper bound on $\alpha_r^d(k)$ as the r-tuples of sets from the family that have pairwise equal intersections are those that consist of the same set r times.

$$\alpha_r^d(k) \le (f_r^d(k)-1)^{1-r}. \tag{1}$$

The following lemma implies that this bound on $\alpha_r^d(k)$ is tight within a factor er^{r-1}.

Lemma 3.2. *For integers $d, k, r \ge 2$ we have*

$$\alpha_r^d(k) \ge g_r^d(k)^{1-r}/e.$$

Proof. Let $S_r^d(m,k)$ denote the minimum possible number of r-sunflowers in a multifamily $\mathcal{F}$ of at most k-element sets with cardinality m and VC-dimension at most d. From the definition, if $m < g_r^d(k)$, then $S_r^d(m,k) = 0$, while if $m \ge g_r^d(k)$, then $S_r^d(m,k) \ge 1$.

Our argument is based on the proof technique used to obtain the "crossing lemma" [2], see also [27]. The idea is to use an averaging (or, equivalently, probabilistic) argument to amplify a weak bound to a better bound. By deleting one set from each r-sunflower, we get the trivial bound $S_r^d(m,k) \ge m - g_r^d(k) + 1$. For $M \ge m$, by averaging over all subfamilies of size m, we obtain

$$S_r^d(M,k) \ge S_r^d(m,k)\binom{M}{r}/\binom{m}{r}.$$

In particular, $S_r^d(m,k)/\binom{m}{r}$ is a monotone increasing function of m. Set $m_0 = (1+1/r)g_r^d(k) - 1$. Then we have $S_r^d(m_0,k) \ge m_0 - g_r^d(k) + 1 = g_r^d(k)/r$. Thus, for $m \ge m_0$, we have

$$\begin{aligned} S_r^d(m,k) &\ge S_r^d(m_0,k)\binom{m}{r}/\binom{m_0}{r} \\ &\ge \frac{1}{r}g_r^d(k)\binom{m}{r}/\binom{(1+1/r)g_r^d(k)}{r} \\ &\ge \frac{1}{er}g_r^d(k)^{1-r}m^r. \end{aligned} \tag{2}$$

Let $\alpha_r^d(m,k)$ be the maximum α with the property that for every multifamily $\mathcal{F}$ of at most k-element sets with cardinality m and VC-dimension at most d, if we uniformly at random choose r sets from $\mathcal{F}$ with replacement, the probability that they have pairwise equal intersections is at least α. Thus,

$$\alpha_r^d(m,k) \ge S_r^d(m,k)/\binom{m}{r} + m^{1-r}, \tag{3}$$

where the first term comes from possibly choosing r different sets (in terms of label, if we view the m not necessarily distinct sets as labeled from 1 to m), and the second term comes from possibly choosing the same set r times.

For $m \geq m_0$, by using (3) and then (2), we have

$$\alpha_r^d(m,k) \geq S_r^d(m,k)/\binom{m}{r} \geq r!S_r^d(m,k)m^{-r} \geq (r-1)!g_r^d(k)^{1-r}/e.$$

For $m < m_0$, using the trivial bound $S_r^d(m,k) \geq 0$, we have

$$\alpha_r^d(m,k) \geq m^{1-r} > m_0^{1-r} = \big((1+1/r)g_r^d(k)-1\big)^{1-r} \geq g_r^d(k)^{1-r}/e.$$

As $\alpha_r^d(k) = \inf_m \alpha_r^d(m,k)$, we have the desired bound $\alpha_r^d(k) \geq g_r^d(k)^{1-r}/e$. ∎

Combining the previous lemma with the Erdős–Rado bound $f_r(k) \leq k!(r-1)^k$, and the inequality $g_r^d(k) \leq g_r(k) = (k-1)f_r(k)+1$, we obtain the following corollary.

Corollary 3.3. *For any integers $d, k, r \geq 2$, we have*

$$\alpha_r^d(k) \geq \big(k!(r-1)^{k+1}+1\big)^{1-r}/e.$$

We are now in a position to prove the following result which, together with (1), immediately implies Theorem 1.3.

Theorem 3.4. *For any $d, k, r \geq 2$, we have*

$$\alpha_r^d(k) \geq 2^{-10k(dr)^{2\log^* k}}.$$

Proof. If $r = 2$, then we have $\alpha_2^d(k) = 1$ and the result follows. Therefore we can assume $r \geq 3$. We use induction on k. For the base cases $k < 8$, by Corollary 3.3, we have

$$\alpha_r^d(k) \geq \big(k!(r-1)^{k+1}+1\big)^{1-r}/e \geq 2^{-10k(dr)^{2\log^* k}}.$$

For the inductive step, let $k \geq 8$ and assume that the statement holds for all $k' < k$. Let $\mathcal{F}$ be a non-empty multifamily of at most k-element sets with VC-dimension at most d. Without loss of generality, we may assume that the ground set is $\mathbb{N}$. Let ϵ_i be the fraction of sets in $\mathcal{F}$ that contain i. By reordering the elements of the ground set, if necessary, we may also assume that $\epsilon_1 \geq \epsilon_2 \geq \ldots$, that is, the elements of the ground set are ordered in decreasing frequency.

As each member of $\mathcal{F}$ has size at most k, the expected size of the intersection of $[s] = \{1, 2, \ldots, s\}$ with a randomly selected member of $\mathcal{F}$ is at most k. On the other hand, this expectation is $\epsilon_1 + \cdots + \epsilon_s \geq s\epsilon_s$. Therefore, we have $\epsilon_s \leq k/s$.

Set $s = \lceil 4k^4/\alpha_r^d(\log k)\rceil$. Define two multifamilies, $\mathcal{F}_1$ and $\mathcal{F}_2$, as follows. Let

$$\mathcal{F}_1 = \{S\colon S \in \mathcal{F} \text{ and } |S \cap [s]| \leq \log k\}, \qquad \mathcal{F}_2 = \mathcal{F} \setminus \mathcal{F}_1.$$

Thus, we have $|\mathcal{F}| = |\mathcal{F}_1| + |\mathcal{F}_2|$. We select at random, uniformly and independently with repetition, r sets $S_1, \ldots, S_r \in \mathcal{F}$. Let X denote the event that the r sets form an r-sunflower. The proof now falls into two cases.

Case 1. *Suppose that* $|\mathcal{F}_1| \geq (1-1/r)|\mathcal{F}|$. *Let* Y *denote the event that* $S_1, \ldots, S_r \in \mathcal{F}_1$. *Let* Z *be the event that* $S_1 \cap [s], \ldots, S_r \cap [s]$ *have pairwise equal intersections, and let* W *be the event that* $S_1 \setminus [s], \ldots, S_r \setminus [s]$ *are pairwise disjoint. Hence,*

$$\mathbb{P}[X] \geq \mathbb{P}[Y \cap Z \cap W] = \mathbb{P}[Y \cap Z] - \mathbb{P}[Y \cap Z \cap \bar{W}] \geq \mathbb{P}[Y \cap Z] - \mathbb{P}[\bar{W}]. \tag{4}$$

Clearly, we have

$$\mathbb{P}[Y] \geq (1-1/r)^r \geq \frac{1}{4}, \tag{5}$$

and, by definition,

$$\mathbb{P}[Z \mid Y] \geq \alpha_r^d(\log k). \tag{6}$$

Therefore, by (5) and (6), we have

$$\mathbb{P}[Y \cap Z] = \mathbb{P}[Y]\mathbb{P}[Z \mid Y] \geq \frac{1}{4}\alpha_r^d(\log k). \tag{7}$$

Fixing $S_i \setminus [s]$, *which has size at most* k, *the probability that* $S_j \setminus [s]$ *contains at least one of the elements of* $S_i \setminus [s]$ *is at most* $k\epsilon_{s+1} \leq k^2/(s+1)$. *Hence, by the probability union bound, we have*

$$\mathbb{P}[\bar{W}] \leq \frac{\binom{k}{2}k^2}{s+1} < \frac{k^4}{2s} \leq \frac{\alpha_r^d(\log k)}{8}. \tag{8}$$

Combining (4), (7), and (8), we obtain

$$\mathbb{P}[X] = \mathbb{P}[Y \cap Z] - \mathbb{P}[\bar{W}] \geq \frac{\alpha_r^d(\log k)}{4} - \frac{\alpha_r^d(\log k)}{8} = \frac{\alpha_r^d(\log k)}{8}.$$

Hence, by the induction hypothesis, we have

$$\alpha_r^d(k) \geq \mathbb{P}[X] \geq \frac{1}{8}\alpha_r^d(\log k) \geq \frac{1}{8}2^{-10(\log k)(dr)^{2\log^* k - 2}} \geq 2^{-10k(dr)^{2\log^* k}}.$$

Case 2. *Suppose that* $|\mathcal{F}_2| \geq |\mathcal{F}|/r$. *Since* $\mathcal{F}$ *has VC-dimension at most* d, *by the Sauer-Shelah-Perles lemma, Lemma 3.1, the number of distinct sets in* $\{S \cap [s] \colon S \in \mathcal{F}\}$ *is at most* s^d. *By the pigeonhole principle, there is a subset* $A \subset [s]$ *with* $|A| \geq \log k$ *such that the family*

$$\mathcal{F}' = \{S \in \mathcal{F} \colon S \cap [s] = A\}$$

has at least $|\mathcal{F}_2|/s^d \geq |\mathcal{F}|/(rs^d)$ *members.*

Select r sets $S_1, \ldots, S_r$ from $\mathcal{F}$ uniformly at random with repetition. Let Y' denote the event that $S_1, \ldots, S_r \in \mathcal{F}'$ and let Z' denote the event that

$S_1 \setminus [s], \ldots, S_r \setminus [s]$ form an r-sunflower. Hence,

$$\begin{aligned}\mathbb{P}[X] &\geq \mathbb{P}[Y' \cap Z'] \\ &= \mathbb{P}[Y'] \cdot \mathbb{P}[Z' \mid Y'] \\ &\geq \left(\frac{1}{rs^d}\right)^r \alpha_r^d(k - \log k) \\ &\geq \frac{1}{r^r(5k^4)^{dr}} \left(\alpha_r^d(\log k)\right)^{dr} \alpha_r^d(k - \log k).\end{aligned}$$

By the induction hypothesis, we obtain

$$\mathbb{P}[X] \geq \frac{1}{r^r(5k^4)^{dr}} \left(2^{-10(\log k)(dr)^{2\log^* k - 2}}\right)^{dr} \left(2^{-10(k-\log k)(dr)^{2\log^* k}}\right).$$

Since $dr \geq 6$ and $k \geq 8$, we have

$$\mathbb{P}[X] \geq \frac{1}{r^r(5k^4)^{dr}} 2^{-10k(dr)^{2\log^* k} + 8\log k(dr)^{2\log^* k}} \geq 2^{-10k(dr)^{2\log^* k}}.$$

This completes the proof. ∎

4 Littlestone dimension – Proof of Theorem 1.4

Originally, the Littlestone dimension was introduced for the characterization of regret bounds in online learning, see [4, 19, 7]. As Chase and Freitag [9] pointed out, the notion is equivalent to Shelah's model theoretic rank. The definition can also be reformulated as follows.

For a finite family $\mathcal{F}$ of sets with ground set V, define LS-dim($\mathcal{F}$), the *Littlestone dimension* of $\mathcal{F}$, recursively. If $|\mathcal{F}| \leq 1$, then let LS-dim($\mathcal{F}$) $= 0$. For an element x of the ground set, let $\mathcal{F}_x = \{S \setminus \{x\} \colon x \in S \text{ and } S \in \mathcal{F}\}$ and $\mathcal{F}'_x = \{S \colon x \notin S \text{ and } S \in \mathcal{F}\}$. If $|\mathcal{F}| > 1$, then let

$$\text{LS-dim}(\mathcal{F}) = 1 + \max_{x \in V} \min\left(\text{LS-dim}(\mathcal{F}_x), \text{LS-dim}(\mathcal{F}'_x)\right).$$

For $d \geq 1$, let $h_r^d(k)$ be the minimum positive integer m such that every family of k-sets with size at least m and Littlestone dimension at most d contains an r-sunflower.

Lemma 4.1. *For positive integers k and r, we have $h_r^1(k) = k + r - 1$.*

Proof. We have $h_r^1(k) > k + r - 2$ by considering the following family $\mathcal{F}_{r,k}$ of k-sets. For $k = 1$, let the family consist of $r - 1$ singleton sets. For $k > 1$, we obtain $\mathcal{F}_{r,k}$ from $\mathcal{F}_{r,k-1}$ by adding one new ground element to all sets in $\mathcal{F}_{r,k-1}$, and then including one additional k-set with entirely new ground elements. It is straightforward to check that this family of k-sets has $k + r - 2$ members, its Littlestone dimension is 1, and it does not contain any r-sunflower.

We prove the upper bound inductively on k, with the base case $k = 1$ being trivial. Let $k \geq 2$ and let $\mathcal{F}$ be a family of k-sets with size $h_r^1(k) - 1$ which has Littlestone dimension at most 1 and does not contain an r-sunflower. A family of sets has Littlestone dimension at most 1 if and only if every element x of the ground set belongs to *at most one* or to *all but at most one* set in the family, that is, if $|\mathcal{F}_x| \leq 1$ or $|\mathcal{F}'_x| \leq 1$ for all x. If there is an element x for which $|\mathcal{F}'_x| \leq 1$, then $|\mathcal{F}_x| = |\mathcal{F}| - 1 = h_r^1(k) - 2$ and $\mathcal{F}_x$ is a family of $(k-1)$-sets of Littlestone dimension at most 1 which does not contain an r-sunflower, from which we obtain $h_r^1(k-1) \leq h_r^1(k-1) + 1$. If there is no ground element x in more than one set in $\mathcal{F}$, then all members of $\mathcal{F}$ are disjoint. Therefore, $|\mathcal{F}| < r$ and $h_r^1(k) \leq r$. ∎

Lemma 4.2. *For any family $\mathcal{F}$ of sets of size at most k with no $(r+1)$-sunflower, there is an element of the ground set which belongs to at least a $\frac{1}{kr}$-fraction of the sets.*

Proof. Consider a maximum family $\{S_1, \ldots, S_s\}$ of sets in $\mathcal{F}$ which are pairwise disjoint. Such a family forms a sunflower and hence $s \leq r$. In particular, any set in $\mathcal{F}$ contains at least one element from $\bigcup_{i=1}^{s} S_i$, which has a total of $ks \leq kr$ elements. By the pigeonhole principle, there is an element of the ground set which belongs to at least a fraction $\frac{1}{kr}$ of the sets in $\mathcal{F}$. ∎

Lemma 4.3. *For integers $k, r \geq 1$ and $d \geq 2$, we have*

$$h_r^d(k) \leq \max\left(k(r-1)\left(h_r^{d-1}(k-1) - 1\right) + 1, h_r^d(k-1) + h_r^{d-1}(k) - 1\right).$$

Proof. Let $\mathcal{F}$ be a family of k-sets with size $h_r^d(k) - 1$ which has Littlestone dimension at most d and does not contain an r-sunflower. By Lemma 4.2, there is an element x of the ground set in at least a fraction $\frac{1}{k(r-1)}$ of the sets in $\mathcal{F}$. As $\mathcal{F}$ has Littlestone dimension d, at least one of $\mathcal{F}_x$ or $\mathcal{F}'_x$ has Littlestone dimension at most $d-1$.

If $\mathcal{F}_x$ has Littlestone dimension at most $d-1$, then $\mathcal{F}_x$ is a family of $(k-1)$-sets which has no r-sunflower, and hence

$$\frac{1}{k(r-1)}\left(h_r^d(k) - 1\right) = \frac{1}{k(r-1)}|\mathcal{F}| \leq |\mathcal{F}_x| \leq h_r^{d-1}(k-1) - 1,$$

from which it follows that $h_r^d(k) \leq k(r-1)\left(h_r^{d-1}(k-1) - 1\right) + 1$.

If $\mathcal{F}'_x$ has Littlestone dimension at most $d-1$, then we have $|\mathcal{F}'_x| \leq h_r^{d-1}(k) - 1$ and $|\mathcal{F}'_x| \leq h_r^d(k-1) - 1$, from which it follows that

$$h_r^d(k) - 1 = |\mathcal{F}| = |\mathcal{F}_x| + |\mathcal{F}'_x| \leq h_r^{d-1}(k) - 1 + h_r^d(k-1) - 1,$$

and, hence, $h_r^d(k) \leq h_r^{d-1}(k) + h_r^d(k-1) - 1$. ∎

We can now prove Theorem 1.4.

Proof of Theorem 1.4. For the upper bound, the proof is by induction on the Littlestone dimension d. In the base case $d = 1$, we have $h_r^1(k) = k + r - 1 \leq kr$. Suppose $d \geq 2$. Consider the recursive upper bound on $h_r^d(k)$ from Lemma 4.3. We split the proof into two cases depending on the maximum of the two functions in the upper bound on $h_r^d(k)$. In each case, we use the induction hypothesis.

In the first case, we have

$$h_r^d(k) \leq k(r-1)\left(h_r^{d-1}(k-1) - 1\right) + 1 \leq kr(kr)^{d-1} = (kr)^d.$$

In the latter case, we have

$$\begin{aligned} h_r^d(k) &\leq h_r^{d-1}(k) + h_r^d(k-1) - 1 < (kr)^{d-1} + ((k-1)r)^d \\ &\leq (kr)^{d-1} + \left(1 - \frac{1}{k}\right)(kr)^d < (kr)^d. \end{aligned}$$

In either case, we obtained the desired bound.

For the lower bound, let $d \geq 6$, $r \geq 3$, k be sufficiently large with $k \geq 4d$, $n = k^2 r/(500d \log k)$, $t = \lceil \log d \rceil$ and $m = n^{-1}(n/k)^{d-t}$. We use the probabilistic method to show that there exists a family $\mathcal{F}$ of k-element subsets of $[n] := \{1, \ldots, n\}$ with $|\mathcal{F}| \geq m/2$, $\mathcal{F}$ does not contain any r-sunflower, and the Littlestone dimension of $\mathcal{F}$ is at most d. This implies the desired lower bound on $h_r^d(k)$.

We will show that $\mathcal{F}$ satisfies four properties each with high probability. This means that the probability is of the form $1 - o(1)$ with the $o(1)$ term tending to 0 as k tends to infinity. Hence, all four properties hold with high probability. These four properties guarantee that $\mathcal{F}$ has the desired properties and hence there is a choice of $\mathcal{F}$ with the desired properties.

Pick m subsets $S_1, \ldots, S_m \subset \binom{[n]}{k}$ uniformly and independently at random. Let $\mathcal{F}$ be the family of distinct S_i. Since $m \ll \binom{n}{k}$, it is easy to see that, with high probability, we have $|\mathcal{F}| \geq m/2$.

We next show that, with high probability, $\mathcal{F}$ does not contain any r-sunflower. Consider a subsequence of r of these random sets, say $S_1, \ldots, S_r$. The number of sequences of r sets in $\binom{[n]}{k}$ which have pairwise equal intersection, and this intersection has size s, is

$$\binom{n}{s} \prod_{i=1}^{r} \binom{n - s - (i-1)(k-s)}{k-s} = s!^{-1}(k-s)!^{-r} n! / \left(n - s - r(k-s)\right)!$$

This is because there are $\binom{n}{s}$ ways of choosing the common intersection of size s, and given the S_j with $j < i$, the remaining $k - s$ elements from S_i not in the common intersection must be chosen from the $n - s - (i-1)(k-s)$ elements not in any of the S_j with $j < i$. As there are $\binom{n}{k}$ ways to pick each S_i, the probability that the r random sets $S_1, \ldots, S_r$ have pairwise equal intersection

of size s is

$$\begin{aligned}(9)\qquad &\binom{n}{k}^{-r} s!^{-1}(k-s)!^{-r}\frac{n!}{(n-s-r(k-s))!}\\ &=\binom{k}{s}\cdot\left(k!/(k-s)!\right)^{r-1}\cdot\frac{n!/(n-s-r(k-s))!}{\left(k!\binom{n}{k}\right)^{r}}.\end{aligned}$$

Note that the expression in the right hand side of (9) is the product of three factors. The middle factor is at most $k^{s(r-1)}$. In the third factor in the right hand side of (9), the numerator can be expressed as the product of factors $(n-j)$ for $j=0,\ldots,s+r(k-s)-1$, which is a total of $s+r(k-s)$ factors, while the denominator can be expressed as the product of rk factors which are of the form $(n-h)$ with $h\leq k$. It follows that the third factor in the right hand side of (9) is at most

$$(n-k)^{-s(r-1)}\prod_{j=1}^{(r-1)(k-s)-1}\left(1-\frac{j}{n-k}\right)\leq(n-k)^{-s(r-1)}e^{-(r-1)^2(k-s)^2/(4n)},$$

where we used the inequality $1-x\leq e^{-x}$ for $x\geq 0$ to bound each factor in the product.

It follows that the expression on the right hand side of (9) is at most

$$(10)\qquad \left(k\left(k/(n-k)\right)^{r-1}\right)^{s}\cdot e^{-(r-1)^2(k-s)^2/(4n)}.$$

Thus, the probability that $S_1,\ldots,S_r$ form an r-sunflower is at most

$$\sum_{s=0}^{k}\left(k\left(k/(n-k)\right)^{r-1}\right)^{s}e^{-(r-1)^2(k-s)^2/(4n)}.$$

We bound the probability that there is an r-sunflower in $\mathcal{F}$ by taking the union bound over all the $\binom{m}{r}$ choices of r sets from $S_1,\ldots,S_m$. Note that (10) is the product of two factors which are each at most 1. We bound (10) for $s\geq 2d$ by the first factor, and for $s<2d$ by the second factor. We also use the inequality $\binom{m}{r}\leq(em/r)^r$. Substituting in the chosen values for n and m, we get an $o(1)$ probability that there is an r-sunflower in $\mathcal{F}$.

Finally, we bound the probability that $\mathcal{F}$ has Littlestone dimension greater than d. If $\mathcal{F}$ has Littlestone dimension greater than d, in the rooted complete binary tree realizing the Littlestone dimension, going down from the root by taking the left-child each time for $d-t$ levels, we see that there are at least $2^t\geq d$ sets that each contain the same $d-t$ vertices. So the probability that $\mathcal{F}$ has Littlestone dimension greater than d is at most the probability that there are d sets in $\mathcal{F}$ that each contain the same $d-t$ elements from $[n]$. This probability in turn is at most

$$\binom{m}{d}\binom{n}{d-t}\left(\frac{k}{n}\right)^{(d-t)d}<\left(\frac{em}{d}\right)^{d}\left(\frac{en}{d-t}\right)^{d-t}\left(\frac{k}{n}\right)^{(d-t)d}=o(1).$$

Here we used the union bound over all $\binom{m}{d}$ choices of d indices from $[m]$ and over all $\binom{n}{d-t}$ choices of $d-t$ distinct integers in $[n]$. We also used that the

probability that a given set of $d-t$ elements in $[n]$ is in a random k-set is at most $(k/n)^{d-t}$. The last inequality is by substituting in the chosen values of n and m. ∎

5 Geometric versions of the sunflower conjecture

We start with the proof of Theorem 1.5. We need the following lemma due to Sharir [24].

Lemma 5.1. *Let $\mathcal{D} = \{D_1, \ldots, D_t\}$ be a family of pseudo-disks in the plane, and let P denote the set of all intersection points of the boundaries of D_i. Then the number of points in P covered by the interior of at most k other regions D_i is at most $26kt$.*

Proof of Theorem 1.5. Given $r > 2$, let $N = (500+r)^{900k}$. We proceed by induction on k. The base case $k = 1$ is trivial. Now assume the statement holds for $k' < k$.

Let V be a planar point set and let $\mathcal{D} = \{D_1, \ldots, D_N\}$ be a family of pseudo-disks in the plane such that $|D_i \cap V| = k$ for all i. By slightly perturbing each region D_i, we can assume that no point in V lies on the boundary of D_i for all i. Set $S_i = D_i \cap V$ and $\mathcal{F} = \{S_1, \ldots, S_N\}$.

Let $t = \lambda(\mathcal{F})$ and suppose the sets $S_1, \ldots, S_t \in \mathcal{F}$ have the property that for any $1 \le i < j \le t$, there is a vertex $v \in S_i \cap S_j$ from V such that $v \notin S_\ell$ for $\ell \in [t] \setminus \{i, j\}$. Then, by letting C_i denote the boundary of D_i, there are at least $\binom{t}{2}$ connected components in $\mathbb{R}^2 \setminus \bigcup_i C_i$ that are covered by at most two regions D_i. On the other hand, by Lemma 5.1, there are at most $4(52)t$ such regions, since every point in the arrangement $\bigcup_i C_i$ is incident to at most four such connected components. Therefore, we have

$$\binom{t}{2} \le 208t,$$

which implies that $t \le 417$.

Further, we can assume that $\nu(\mathcal{F}) \le r-1$, since otherwise we would be done. By Lemma 2.1, we have

$$\tau(\mathcal{F}) \le 11(417)^2(419+r)\binom{416+r}{417}^2 \le (500+r)^{900}.$$

There is a vertex $v \in V$ which is incident to at least $N/\tau(\mathcal{F})$ members in $\mathcal{F}$. Let $\mathcal{D}' = \{D_i \in \mathcal{D} \colon v \in D_i\}$, $V' = V \setminus \{v\}$, and $S_i' = V' \cap D_i'$. Hence, $|\mathcal{D}'| \ge N/\tau(\mathcal{F}) \ge (500+r)^{900(k-1)}$. By the induction hypothesis, there are r sets $S_{i_1}', \ldots, S_{i_r}'$ in $\mathcal{D}'$ that form an r-sunflower. Together with vertex v, we obtain an r-sunflower in $\mathcal{F}$. ∎

Replacing Lemma 5.1 with Clarkson's theorem on levels in arrangement of hyperplanes [10], the argument above gives the following.

Theorem 5.2. *Given $r > 2$, there is a constant $C = C(r)$ for which the following statement is true. If V is a point set in $\mathbb{R}^3$ and $\mathcal{H} = \{H_1, \ldots, H_N\}$ is a family of $N \geq C^k$ half-spaces such that the size of the set $S_i = H_i \cap V$ is k for all i, then there are r distinct sets $S_{i_1}, \ldots, S_{i_r}$ that form an r-sunflower.*

6 Concluding remarks

The Erdős–Rado sunflower conjecture remains an outstanding open problem. Although we made progress in this paper, it still remains open for families of bounded VC-dimension. It is unclear if the methods used in this paper can be used towards the original sunflower conjecture.

We were able to prove the conjecture in a geometric setting in the plane (Theorem 1.5). We think it would be interesting to prove the conjecture in other geometric settings, such as for families of sets that are the intersection of the ground set with semi-algebraic sets of bounded description complexity. Such families are of bounded VC-dimension. The following conjecture is a natural special case to consider.

Conjecture 6.1. *For each integer $r \geq 3$, there is a constant $C = C(r)$ such that the following holds. If $V \subset \mathbb{R}^3$ and $\mathcal{F}$ is a family of subsets of V each of size k with $|\mathcal{F}| \geq C^k$ such that every set in $\mathcal{F}$ is the intersection of V with a unit ball in $\mathbb{R}^3$, then $\mathcal{F}$ contains an r-sunflower.*

Acknowledgement. We would like to thank Amir Yehudayoff for suggesting working with the Littlestone dimension, and the SoCG 2021 referees for helpful comments.

References

[1] H. L. Abbott, D. Hanson and N. Sauer: Intersection theorems for systems of sets, *J. Combin. Theory Ser. A* **12** (1972), 381–389.

[2] M. Ajtai, V. Chvátal, M. Newborn and E. Szemerédi: Crossing-free subgraphs, in: *Theory and Practice of Combinatorics* **60**, *Ann. Discrete Math., North-Holland Math. Stud.* **12**, North-Holland, Amsterdam, 1982, 9–12.

[3] N. Alon and R. Holzman: Near-sunflowers and focal families, *to appear in Israel Journal of Mathematics* arXiv:2010.05992.

[4] N. Alon, R. Livni, M. Malliaris and S. Moran: Private PAC learning implies finite Littlestone dimension, *Proc. STOC* 2019, 852–860.

[5] R. Alweiss, S. Lovett, K. Wu and J. Zhang: Improved bounds for the sunflower lemma, *Ann. Math.* **194** (2021), 795–815.

[6] M. Axenovich, D. Fon-Der-Flaass and A. Kostochka: On set systems without weak 3-Δ-subsystems, *Discrete Math.* **138** (1995), 57–62.

[7] S. Ben-David, D. Pál and S. Shalev-Shwartz: Agnostic online learning, in *COLT*, 2009.

[8] S. Buzaglo, R. Holzman and R. Pinchasi: On s-intersecting curves and related problems, *Symposium on Computational Geometry* (2008), 79–84.

[9] H. Chase and J. Freitag: Model theory and machine learning, *Bull. Symb. Log.* **25** (2019), 319–332.

[10] K. L. Clarkson: Applications of random sampling in computational geometry, II, *Symposium on Computational Geometry* (1988), 1–11.

[11] M. Deza and P. Frankl: Every large set of equidistant $(0, +1, -1)$-vectors forms a sunflower, *Combinatorica* **1** (1981), 225–231.

[12] G. Ding, P. Seymour and P. Winkler: Bounding the vertex cover number of a hypergraph, *Combinatorica* **14** (1994), 23–34.

[13] P. Erdős and R. Rado: Intersection theorems for systems of sets, *J. London Math. Soc.* **35** (1960), 85–90.

[14] P. Erdős and E. Szemerédi: Combinatorial properties of systems of sets, *J. Combin. Theory Ser. A* **24** (1978), 308–313.

[15] J. Fox, J. Pach and A. Suk: Erdős-Hajnal conjecture for graphs with bounded VC-dimension, *Discrete Comput. Geom.* **61** (2019), 809–829.

[16] J. Fox, J. Pach and A. Suk: Bounded VC-dimension implies the Schur-Erdős conjecture, *Symposium on Computational Geometry* (2020), 46:1–46:8.

[17] P. Frankl: On the trace of finite sets, *J. Combin. Theory Ser. A* **34** (1983), 41–45.

[18] A. V. Kostochka: An intersection theorem for systems of sets, *Random Structures Algorithms* **9** (1996), 213–221.

[19] N. Littlestone: Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm, *Mach. Learn.* **2** (1987), 285–318.

[20] E. Naslund and W. Sawin: Upper bounds for sunflower-free sets, *Forum Math. Sigma* **5** (2017), Paper No. e15.

[21] J. Pach and P. Agarwal: *Combinatorial Geometry*, Wiley-Interscience, New York, 1995.

[22] A. Rao: Coding for sunflowers, *Discrete Anal.* (2020) Paper No. 2.

[23] N. Sauer: On the density of families of sets, *J. Combinat. Theory Ser. A* **13** (1972), 145–147.

[24] M. Sharir: On k-sets in arrangements of curves and surfaces, *Discrete Comput. Geom.* **6** (1991), 593–617.

[25] S. Shelah: A combinatorial problem, stability and order for models and theories in infinitary languages, *Pacific J. Math.* **41** (1972), 247–261.

[26] S. Shelah: *Classification Theory and the Number of Non-Isomorphic Models*, North-Holland, Amsterdam, 1978.

[27] S. Smorodinsky and M. Sharir: Selecting points that are heavily covered by pseudo-circles, spheres or rectangles, *Combin. Probab. Comput.* **13** (2004), 389–411.

[28] J. Spencer: Intersection theorems for systems of sets, *Canad. Math. Bull.* **20** (1977), 249–254.

[29] V. Vapnik and A. Chervonenkis: On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* **16** (1971), 264–280.

Jacob Fox

Department of Mathematics
Stanford University
Stanford, CA, USA
jacobfox@stanford.edu

János Pach

Rényi Institute
Budapest, Hungary
pach@cims.nyu.edu

Andrew Suk

Department of Mathematics
University of California at San Diego
La Jolla, CA, 92093 USA
asuk@ucsd.edu