Electronic Colloquium on Computational Complexity, Report No. 110 (2019)

Improved bounds for the sunflower lemma

Ryan Alweiss
Department of Mathematics
Princeton University
alweiss@math.princeton.edu

Shachar Lovett*
Computer Science Department
University of California, San Diego
shachar.lovett@gmail.com

Kewen Wu
School of Electronics Engineering and Computer Science
Peking University, Beijing
shlw_kevin@pku.edu.cn

Jiapeng Zhang†
Computer Science Department
University of California, San Diego
jpeng.zhang@gmail.com

August 23, 2019

Abstract

A sunflower with r petals is a collection of r sets so that the intersection of each pair is equal to the intersection of all. Erdős and Rado proved the sunflower lemma: for any fixed r, any family of sets of size w, with at least about w^w sets, must contain a sunflower. The famous sunflower conjecture is that the bound on the number of sets can be improved to c^w for some constant c. In this paper, we improve the bound to about $(\log w)^w$. In fact, we prove the result for a robust notion of sunflowers, for which the bound we obtain is tight up to lower order terms.

1 Introduction

Let X be a finite set. A set system $\mathcal{F}$ on X is a collection of subsets of X. We call $\mathcal{F}$ a w-set system if each set in $\mathcal{F}$ has size at most w.

*Research supported by NSF award 1614023.
†Research supported by NSF award 1614023.

ISSN 1433-8092

Definition 1.1 (Sunflower)**.** *A collection of sets $S_1, \dots, S_r$ is an r-*sunflower *if*

$$S_i \cap S_j = S_1 \cap \cdots \cap S_r, \qquad \forall i \neq j.$$

We call $K = S_1 \cap \cdots \cap S_r$ the kernel *and $S_1 \setminus K, \dots, S_r \setminus K$ the* petals *of the sunflower.*

Erdős and Rado [2] proved that large enough set systems must contain a sunflower.

Lemma 1.2 (Sunflower lemma [2])**.** *Let $r \geq 3$ and $\mathcal{F}$ be a w-set system of size $|\mathcal{F}| \geq w! \cdot (r-1)^w$. Then $\mathcal{F}$ contains an r-sunflower.*

Erdős and Rado conjectured in the same paper that the bound in Lemma 1.2 can be improved. This is the famous sunflower conjecture.

Conjecture 1.3 (Sunflower conjecture [2])**.** *Let $r \geq 3$. There exists $c(r)$ such that any w-set system $\mathcal{F}$ of size $|\mathcal{F}| \geq c(r)^w$ contains an r-sunflower.*

Consider a fixed r. The bound in Lemma 1.2 is of the order of $(\Theta(w))^w$. Despite nearly 60 years of research, the best known bounds towards the sunflower conjecture are still of the order of $w^{w(1-o(1))}$, even for $r = 3$. More precisely, Kostochka [5] proved that any w-set system of size $|\mathcal{F}| \geq cw! \cdot (\log\log\log w / \log\log w)^w$ must contain a 3-sunflower for some absolute constant c.

Our main result improves upon this, getting closer towards the sunflower conjecture. We prove that any w-set system of size $(\log w)^{w(1+o(1))}$ must contain a sunflower. More precisely, we obtain the following bounds.

Theorem 1.4 (Main theorem, sunflowers)**.** *Let $r \geq 3$. Any w-set system $\mathcal{F}$ of size $|\mathcal{F}| \geq (\log w)^w (r \cdot \log\log w)^{O(w)}$ contains an r-sunflower.*

2 Robust sunflowers

Our approach to find sunflowers is to find a more general type of structure. This was called *quasi-sunflower* in [11] and *approximate sunflower* in [7]. However, as its existence implies the existence of sunflowers, a better name is *robust sunflower*, which we adopt in this paper.

First, we define the notion of a satisfying set system. Given a finite set X, we denote by $\mathcal{U}(X, p)$ the distribution of sets $Y \subset X$, where each element $x \in X$ is included in Y independently with probability p. We note that throughout the paper, we interpret "$\subset$" as "subset or equal to".

Definition 2.1 (Satisfying set system)**.** *Let $0 < \alpha, \beta < 1$. A set system $\mathcal{F}$ on X is (α, β)-satisfying if*

$$\Pr_{Y \sim \mathcal{U}(X,\alpha)}[\exists S \in \mathcal{F}, S \subset Y] > 1 - \beta.$$

Given a set system $\mathcal{F}$ on X and a set $T \subset X$, the *link* of $\mathcal{F}$ at T is

$$\mathcal{F}_T = \{S \setminus T : S \in \mathcal{F}, T \subset S\}.$$

Definition 2.2 (Robust sunflower)**.** *Let $0 < \alpha, \beta < 1$, $\mathcal{F}$ be a set system, and let $K = \bigcap_{S \in \mathcal{F}} S$ be the common intersection of all sets in $\mathcal{F}$. $\mathcal{F}$ is an (α, β)-robust sunflower if (i) $K \notin \mathcal{F}$; and (ii) $\mathcal{F}_K$ is (α, β)-satisfying. We call K the* kernel*.*

The connection between robust sunflowers and sunflowers was made in [7].

Lemma 2.3 ([7])**.** *If $\mathcal{F}$ is a $(1/r, 1/r)$-satisfying set system and $\emptyset \notin \mathcal{F}$, then $\mathcal{F}$ contains r pairwise disjoint sets.*

For completeness, we include the proof.

Proof. Let $\mathcal{F}$ be a set system on X. Consider a random r-coloring of X, where each element obtains any of the r colors with equal probability. Let $Y_1, \dots, Y_r$ denote the color classes, which are a random partition of X. For $i = 1, \dots, r$, let $\mathcal{E}_i$ denote the event that $\mathcal{F}$ contains an i-monochromatic set, namely,

$$\mathcal{E}_i = [\exists S \in \mathcal{F}, S \subset Y_i]\,.$$

Note that $Y_i \sim \mathcal{U}(X, 1/r)$, and since we assume $\mathcal{F}$ is $(1/r, 1/r)$-satisfying, we have

$$\Pr[\mathcal{E}_i] > 1 - 1/r.$$

By the union bound, with positive probability all $\mathcal{E}_1, \dots, \mathcal{E}_r$ hold. In this case, $\mathcal{F}$ contains a set which is i-monochromatic for each $i = 1, \dots, r$. Such sets must be pairwise disjoint. $\square$

Lemma 2.4 ([7])**.** *Any $(1/r, 1/r)$-robust sunflower contains an r-sunflower.*

Proof. Let $\mathcal{F}$ be a $(1/r, 1/r)$-robust sunflower, and let $K = \bigcap_{S \in \mathcal{F}} S$ be the common intersection of the sets in $\mathcal{F}$. Note that by assumption, $\mathcal{F}_K$ does not contain the empty set as an element. Lemma 2.3 gives that $\mathcal{F}_K$ contains r pairwise disjoint sets $S_1, \dots, S_r$. Thus $S_1 \cup K, \dots, S_r \cup K$ is an r-sunflower in $\mathcal{F}$. $\square$

The proof of Theorem 1.4 follows from the following stronger theorem, by setting $\alpha = \beta = 1/r$ and applying Lemma 2.4. The theorem verifies a conjecture raised in [7], and answers a question of [11].

Theorem 2.5 (Main theorem, robust sunflowers)**.** *Let $0 < \alpha, \beta < 1$. Any w-set system $\mathcal{F}$ of size $|\mathcal{F}| \geq (\log w)^w \cdot (\log \log w \cdot \log(1/\beta)/\alpha)^{O(w)}$ contains an (α, β)-robust sunflower.*

We make a couple of notes. The bound in Theorem 2.5 for large w can be improved to $(\log w)^w \cdot (\log \log w \cdot \log(1/\beta)/\alpha)^{w(1+o(1))}$. Moreover, for robust sunflowers the bound of $(\log w)^{w(1+o(1))}$ is sharp; it cannot be improved beyond $(\log w)^{w(1-o(1))}$. We give an example demonstrating this in Lemma 4.1.

Below, we briefly describe a couple of applications of our techniques beyond the improved bound for the sunflower lemma.

2.1 Intersecting set systems

The study of how "spread out" an intersecting set system can be was investigated in [7], motivated by its connection to the sunflower conjecture. Applying Lemma 2.3 for $r = 2$ shows that a $(1/2, 1/2)$-satisfying set system cannot be intersecting. Applying Theorem 5.5 for $\alpha = \beta = 1/2$ gives the following bound, which proves a conjecture raised in [6, 7], and may be of independent interest.

Theorem 2.6. *Let $\mathcal{F}$ be an intersecting set system on n elements. Then there exists a non-empty set $T \subset [n]$ such that $|\mathcal{F}_T| \geq |\mathcal{F}|/(\log n)^{O(|T|)}$.*

We note that the intersecting condition cannot be replaced by the weaker condition that most pairs of sets intersect. For example, if $\mathcal{F}$ is the family of all sets of size $10\sqrt{n}$ in $[n]$, then over 99% of the pairs of sets in $\mathcal{F}$ intersect. However, for any $T \subset [n]$ it holds that $|\mathcal{F}_T| \leq |\mathcal{F}|/(0.1\sqrt{n})^{|T|}$.

2.2 Improved bounds for Erdős-Szemerédi sunflowers

Erdős and Szemerédi [3] defined a weaker notion of sunflowers, where instead of bounding the size of the sets in the family, they bound the size of the base set X. The following conjecture follows from Conjecture 1.3.

Conjecture 2.7. *For any $r \geq 3$ there exists $\varepsilon = \varepsilon(r) > 0$ such that the following holds. Let $\mathcal{F}$ be a set system on X, with $|X| = n$ and $|\mathcal{F}| \geq 2^{(1-\varepsilon)n}$. Then $\mathcal{F}$ contains an r-sunflower.*

Erdős and Szemerédi showed that the sunflower lemma (Lemma 1.2) implies a weaker version of Conjecture 2.7, where the bound needed on $\mathcal{F}$ is $|\mathcal{F}| \geq 2^{n\left(1-cn^{-1/2}\right)}$ for some $c = c(r) > 0$. These are the best known bounds for $r \geq 4$. For $r = 3$ Conjecture 2.7 is known to be true - it follows from the resolution of the cap-set problem (see [9] for details).

Plugging in our improved bounds for the sunflower lemma to Erdős and Szemerédi framework yields the following improved bounds for $r \geq 4$.

Theorem 2.8. *For any $r \geq 3$ there exists $c = c(r)$ such that the following holds. Let $\mathcal{F}$ be a set system on X, with $|X| = n$ and $|\mathcal{F}| \geq 2^{n\left(1-c(\log n)^{-(1+o(1))}\right)}$. Then $\mathcal{F}$ contains an r-sunflower.*

3 Proof overview

In this section, we explain the high level ideas underlying the proof of Theorem 2.5. Our framework builds upon the work of Lovett, Solomon and Zhang [7]. Their main idea was to apply a structure vs. pseudo-randomness approach. However, the proof relied on a certain conjecture on the level of pseudo-randomness needed for the argument to go through. Our main technical result is a resolution of this conjecture.

To be concrete, let's consider the problem of finding a 3-sunflower, which corresponds in our framework to finding a $(1/3, 1/3)$-robust sunflower (see Lemma 2.4). Given $w \geq 2$, our goal is to find a parameter $\kappa = \kappa(w)$ such that any w-set system of size κ^w must contain a $(1/3, 1/3)$-robust sunflower, and hence also a 3-sunflower.

Recall the definition of links: $\mathcal{F}_T = \{S \setminus T : S \in \mathcal{F}, T \subset S\}$. We say that a w-set system is κ-bounded if (i) $|\mathcal{F}| \geq \kappa^w$; and (ii) $|\mathcal{F}_T| \leq \kappa^{w-|T|}$ for all non-empty T (The actual definition needed in the proof is bit more delicate, see Definition 5.1 for details).

Let $\mathcal{F}$ be a w-set system of size $|\mathcal{F}| \geq \kappa^w$. Then either $\mathcal{F}$ is κ-bounded, or otherwise there is a link $\mathcal{F}_T$ of size $|\mathcal{F}_T| \geq \kappa^{w-|T|}$. In the latter case, we can focus on this link and repeat the argument (this is the structured case).

So, from now on we consider only w-set systems which are κ-bounded (this is the pseudo-random case). The main question is: how large should κ be to ensure that $\mathcal{F}$ is $(1/3, 1/3)$-satisfying? The answer was conjectured to be $(\log w)^{O(1)}$ in [6, 7]. If true, then it completes the proof of Theorem 2.5. This is our main technical contribution. We show that in fact $\kappa = (\log w)^{1+o(1)}$ is sufficient (see Theorem 5.5). This is essentially tight, as in [7] it was observed that $\kappa \geq (\log w)^{1-o(1)}$ is necessary.

We next explain how we obtain the bound on κ. Let $\mathcal{F}$ be a w-set system which is κ-bounded. In [7] it was conjectured that there exists a $(w/2)$-set system $\mathcal{F}'$ that "covers" $\mathcal{F}$: for any set $S \in \mathcal{F}$, there exists $S' \in \mathcal{F}'$ such that $S' \subset S$. Also, $\mathcal{F}'$ is κ'-bounded for $\kappa' \approx \kappa$. If this conjecture is true, then it is sufficient to prove that $\mathcal{F}'$ is $(1/3, 1/3)$-satisfying, as this would imply that $\mathcal{F}$ is also $(1/3, 1/3)$-satisfying (in the language of [7], this corresponds to "upper bound compression for DNFs". For more details on the connection to DNF compression see [7, 8]).

What we show is that this conjecture is true with two modifications: we are allowed to remove a small fraction of the sets in $\mathcal{F}$, and also remove a small random fraction of the elements in the base set X. To be more precise, sample $W \sim \mathcal{U}(X, p)$ for $p = O(1/\log w)$. We show that with high probability over W, for most sets $S \in \mathcal{F}$, there exists a set $S' \in \mathcal{F}$ such that: (i) $S' \setminus W \subset S \setminus W$; and (ii) $|S' \setminus W| \leq w/2$. Thus we can move to study the set system $\mathcal{F}'$ comprised of the $S' \setminus W$ above, which "approximately covers" $\mathcal{F}$. Note that $\mathcal{F}'$ is a $(w/2)$-set system which is κ'-bounded for $\kappa' \approx \kappa$. In the actual proof, we replace $w/2$ with $(1-\varepsilon)w$ for a small ε to optimize the parameters. For details see Lemma 5.6.

Applying this "reduction step" iteratively $t = \log w$ times reduces the size w to constant, where we can apply standard tools (Janson's inequality, see Lemma 5.9). We get that if we sample $W_1, \ldots, W_t \sim \mathcal{U}(X, p)$ (formally, they are disjoint, but we suppress this detail here), then with high probability there exists $S \in \mathcal{F}$ such that $S \subset W_1 \cup \cdots \cup W_t$. Setting $p \cdot t = 1/3$ and the high probability to be $2/3$ gives that $\mathcal{F}$ is $(1/3, 1/3)$-satisfying, as desired.

To conclude, let us comment on how we prove the reduction step (Lemma 5.6). The main idea is to use an encoding lemma, inspired by Razborov's proof of Håstad's switching lemma [4, 10]. Concretely, for $W \subset X$ and $S \in \mathcal{F}$, we say that the pair (W, S) is *bad* if there is no $S' \in \mathcal{F}$ such that (i) $S' \setminus W \subset S \setminus W$; and (ii) $|S' \setminus W| \leq w/2$. We show that bad pairs can be efficiently encoded, crucially relying on the κ-boundedness condition. This allows to bound the probability that for a random W there are many bad sets. The $(w/2)$-set system

$\mathcal{F}'$ is then taken to be all those $S' \setminus W$ of size at most $w/2$.

4 Lower bound for robust sunflowers

In this section, we construct an example of a w-set system which does not contain a robust sunflower, even though it has size $(\log w)^{w(1-o(1))}$. For concreteness we fix $\alpha = \beta = 1/2$, but the construction can be easily modified for any other constant values of α, β. We assume that w is large enough.

Lemma 4.1. *There exists a w-set system of size $((\log w)/8)^{w-\sqrt{w}} = (\log w)^{w(1-o(1))}$ which does not contain a $(1/2, 1/2)$-robust sunflower.*

Let $c \geq 1$ be determined later. Let $X_1, \ldots, X_w$ be pairwise disjoint sets of size $m = \log(w/c)$, and let X be their union. Let $\widehat{\mathcal{F}} = X_1 \times \cdots \times X_w$ be the w-set system containing all sets which contain exactly one element from each of the X_i. We first argue that $\widehat{\mathcal{F}}$ is not satisfying.

Claim 4.2. *For $c \geq 1$, $\widehat{\mathcal{F}}$ is not $(1/2, 1/2)$-satisfying.*

Proof. Let $Y \sim \mathcal{U}(X, 1/2)$. We analyze the probability that some X_i is disjoint from Y, which implies that no set in $\widehat{\mathcal{F}}$ is contained in Y. The probability is $1-(1-2^{-m})^w = 1-(1-c/w)^w$, which is more than $1/2$ for $c \geq 1$. $\square$

Unfortunately, $\widehat{\mathcal{F}}$ does contain a $(1/2, 1/2)$-robust sunflower with a large kernel. For example, if T contains exactly one element from each of $X_1, \ldots, X_{w-1}$, then $\widehat{\mathcal{F}}_T$ is isomorphic to X_w, and in particular is $(1/2, 1/2)$-satisfying.

To overcome this, let $\varepsilon > 0$ be determined later, and choose $\mathcal{F} \subset \widehat{\mathcal{F}}$ to be a sub-set system that satisfies:

$$|S \cap S'| \leq (1 - \varepsilon)w, \qquad \forall S, S' \in \mathcal{F}, S \neq S'.$$

For example, we can obtain $\mathcal{F}$ by a greedy procedure, each time choosing an element S in $\widehat{\mathcal{F}}$ and deleting all S' such that $|S \cap S'| > (1 - \varepsilon)w$. The number of such S' is at most $\binom{w}{\varepsilon w} m^{\varepsilon w} \leq 2^w m^{\varepsilon w}$. Hence we can obtain $\mathcal{F}$ of size $|\mathcal{F}| \geq 2^{-w} m^{(1-\varepsilon)w}$.

Claim 4.3. *For $c \geq 1/\varepsilon$, $\mathcal{F}$ does not contain a $(1/2, 1/2)$-robust sunflower.*

Proof. Consider any set $K \subset X$. We need to prove that $\mathcal{F}$ does not contain a $(1/2, 1/2)$-robust sunflower with kernel K. In particular, $\mathcal{F}_K$ must contain at least two sets, which implies that $|K \cap X_i| \leq 1$ for all i, and that in addition $|K| \leq (1 - \varepsilon)w$. However, in this case we claim even $\widehat{\mathcal{F}}_K$ is not $(1/2, 1/2)$-satisfying.

To prove this, let $I = \{i : |K \cap X_i| = 0\}$ where $|I| \geq \varepsilon w$. Let $Y \sim \mathcal{U}(X, 1/2)$. The probability that there exists $i \in I$ such that Y is disjoint from X_i is $1 - (1 - 2^{-m})^{|I|} \geq 1 - (1 - c/w)^{\varepsilon w}$ which is more than $1/2$ for $c \geq 1/\varepsilon$. $\square$

To conclude the proof of Lemma 4.1 we optimize the parameters. Set $c = 1/\varepsilon$. We have $|\mathcal{F}| \geq 2^{-w} (\log(\varepsilon w))^{(1-\varepsilon)w}$. Setting $\varepsilon = 1/\sqrt{w}$ gives $|\mathcal{F}| \geq ((\log w)/8)^{w-\sqrt{w}} = (\log w)^{(1-o(1))w}$.

5 Proof of Theorem 2.5

We proceed to prove Theorem 2.5. The main idea is to apply a structure vs. pseudo-randomness paradigm, following the approach outlined in [7]. Let $\mathcal{F}$ be a set system, and let $\sigma : \mathcal{F} \mapsto \mathbb{R}_{\geq 0}$ be a weight function, assigning non-negative weights to sets in $\mathcal{F}$. We consider the pair $(\mathcal{F}, \sigma)$ as a *weighted set system*. For a subset $\mathcal{F}' \subset \mathcal{F}$ we shorthand $\sigma(\mathcal{F}') = \sum_{S \in \mathcal{F}'} \sigma(S)$ the sum of the weights of the sets in $\mathcal{F}'$.

A *weight profile* is a vector $\mathbf{s} = (s_0; s_1, \ldots, s_k)$ where $1 \geq s_0 \geq s_1 \geq \cdots \geq s_k \geq 0$ are rational numbers. We assume implicitly that $s_i = 0$ for all $i > k$.

Definition 5.1 (Bounded weighted set system)**.** *Let* $\mathbf{s} = (s_0; s_1, \ldots, s_w)$ *be a weight profile. A weighted set system* $(\mathcal{F}, \sigma)$ *is* $\mathbf{s}$*-bounded if*

(i) $\sigma(\mathcal{F}) \geq s_0$*.*

(ii) $\sigma(\mathcal{F}_T) \leq s_{|T|}$ *for any link* $\mathcal{F}_T$ *with non-empty* T*.*

In particular, $\mathcal{F}$ *is a* w*-set system.*

Definition 5.2 (Bounded set system)**.** *Let* $\mathbf{s}$ *be a weight profile. A set system* $\mathcal{F}$ *is* $\mathbf{s}$*-bounded if there exists a weight function* $\sigma : \mathcal{F} \mapsto \mathbb{R}_+$ *such that* $(\mathcal{F}, \sigma)$ *is* $\mathbf{s}$*-bounded.*

We note that one may always normalize a weight profile to have $s_0 = 1$. However, keeping s_0 as a free parameter helps to simplify some of the arguments later.

The main idea is to show that set systems which are $\mathbf{s}$-bounded, for $\mathbf{s}$ appropriately chosen, are "random looking" and in particular must be (α, β)-satisfying. This motivates the following definition.

Definition 5.3 (Satisfying weight profile)**.** *Let* $0 < \alpha, \beta < 1$*. A weight profile* $\mathbf{s}$ *is* (α, β)*-satisfying if any* $\mathbf{s}$*-bounded set system is* (α, β)*-satisfying.*

The following lemma underlies our proof of Theorem 2.5.

Lemma 5.4. *Let* $0 < \alpha, \beta < 1$ *and* $w \geq 2$*. Let* $\kappa > 1$ *be an integer such that the weight profile* $(1; \kappa^{-1}, \ldots, \kappa^{-\ell})$ *is* (α, β)*-satisfying for all* $\ell = 1, \ldots, w$*. Then any* w*-set system* $\mathcal{F}$ *of size* $|\mathcal{F}| > \kappa^w$ *must contain an* (α, β)*-robust sunflower.*

Proof. Let $\mathcal{F}$ be a w-set system on X of size $|\mathcal{F}| > \kappa^w$. Let $K \subset X$ be maximal so that $|\mathcal{F}_K| > \kappa^{w-|K|}$. Note that we cannot have $|K| = w$, as in this case $|\mathcal{F}_K| = 1 = \kappa^0$, and so $|K| \leq w - 1$. Let $\mathcal{F}' = \mathcal{F}_K \setminus \{\emptyset\}$. Note that $|\mathcal{F}'| \geq \kappa^{w-|K|}$, where for any non-empty set R disjoint from K, $|\mathcal{F}'_R| = |\mathcal{F}_{K \cup R}| \leq \kappa^{w-|K|-|R|}$. Let $\sigma(S) = 1$ for $S \in \mathcal{F}'$. Then $(\mathcal{F}', \sigma)$ is $(1; \kappa^{-1}, \ldots, \kappa^{-\ell})$-bounded for $\ell = w - |K|$. Hence by our assumption, $\mathcal{F}'$ is (α, β)-satisfying, and hence $\{S \cup K : S \in \mathcal{F}'\}$ is an (α, β)-robust sunflower contained in $\mathcal{F}$. $\square$

Lemma 5.4 motivates the following definition. For $0 < \alpha, \beta < 1$ and $w \geq 2$, let $\kappa(w, \alpha, \beta)$ be the least κ such that $(1; \kappa^{-1}, \ldots, \kappa^{-w})$ is (α, β)-satisfying. Theorem 2.5 follows by combining Lemma 5.4 with the following theorem, which bounds $\kappa(w, \alpha, \beta)$. We note that the theorem proves a conjecture raised in [7].

Theorem 5.5. $\kappa(w, \alpha, \beta) \leq \log w \cdot (\log \log w \cdot \log(1/\beta)/\alpha)^{O(1)}$.

We prove Theorem 5.5 in the remainder of this section.

5.1 A reduction step

Let $\mathcal{F}$ be a w-set system on X, and fix $w' \leq w$. The main goal in this section is to reduce $\mathcal{F}$ to a w'-set system $\mathcal{F}'$. We prove the following lemma in this section.

Lemma 5.6. *Let* $\mathbf{s} = (s_0; s_1, \ldots, s_w)$ *be a weight profile,* $w' \leq w$*,* $\delta > 0$ *and define* $\mathbf{s}' = ((1-\delta)s_0; s_1, \ldots, s_{w'})$*. Assume* $\mathbf{s}'$ *is* (α', β')*-satisfying. Then for any* $p > 0$*,* $\mathbf{s}$ *is* (α, β)*-satisfying for*

$$\alpha = p + (1-p)\alpha', \quad \beta = \beta' + \frac{(4/p)^w s_{w'}}{\delta s_0}.$$

Let $W \subset X$. Given a set $S \in \mathcal{F}$, the pair (W, S) is said to be *good* if there exists a set $S' \in \mathcal{F}$ (possibly with $S' = S$) such that

(i) $S' \setminus W \subset S \setminus W$.

(ii) $|S' \setminus W| \leq w'$.

If no such S' exists, we say that (W, S) is *bad*. Note that if it happens that W contains a set in $\mathcal{F}$ (namely, $S' \subset W$ for some $S' \in \mathcal{F}$) then all pairs (W, S) are good.

Lemma 5.7. *Let* $(\mathcal{F}, \sigma)$ *be an* $\mathbf{s} = (s_0; s_1, \ldots, s_w)$*-bounded weighted set system on* X*. Let* $W \subset X$ *be a uniform subset of size* $|W| = p|X|$ *and* $\mathcal{B}(W) = \{S \in \mathcal{F} : (W, S) \text{ is bad}\}$*. Then* $\mathbb{E}_W[\sigma(\mathcal{B}(W))] \leq (4/p)^w s_{w'}$*.*

Proof. First, we simplify the setting a bit. We may assume by scaling σ and $\mathbf{s}$ by the same factor that $\sigma(S) = N_S, S \in \mathcal{F}$ are all integers. Let $N = \sum N_S \geq \mathbf{s}_0$. We can identify $(\mathcal{F}, \sigma)$ with the multi-set system $\mathcal{F}' = \{S_1, \ldots, S_N\}$, where every set $S \in \mathcal{F}$ is repeated N_S times. Observe that $|\mathcal{F}'_T| = \sigma(\mathcal{F}_T)$ and that (W, S) is bad in $\mathcal{F}$ iff (W, S_i) is bad in $\mathcal{F}'$ where $S_i = S$ is any copy of S. Thus

$$\sigma(\mathcal{B}(W)) = |\{i : S_i \in \mathcal{F}' \text{ and } (W, S_i) \text{ is bad}\}|.$$

Assume that (W, S_i) is bad in $\mathcal{F}'$. In particular, this means that W does not contain any set in $\mathcal{F}$. We describe (W, S_i) with a small amount of information. Let $|X| = n$ and $|W| = pn$. We encode (W, S_i) as follows:

1. The first piece of information is $W \cup S_i$. The number of options for this is

$$\sum_{i=0}^{w} \binom{n}{pn+i} \leq \binom{n+w}{pn+w} \leq p^{-w} \binom{n}{pn}.$$

2. Given $W \cup S_i$, let j be minimal such that $S_j \subset W \cup S_i$; in particular, this is equivalent to $S_j \setminus W \subset S_i \setminus W$. There are fewer than 2^w possibilities for $A = S_i \cap S_j$ given that we know S_j. As such, we will let A be the second piece of information.

3. Note that as (W, S_i) is bad, $|A| = |S_j \setminus W| > w'$. So we know a subset A of S_i of size larger than w'. The number of the sets in $\mathcal{F}'$ which contain A is $|\mathcal{F}'_A| \le s_{w'}$. The third piece of information will be which one of these is S_i.

4. Finally, once we have specified S_i, we will specify $S_i \cap W$, which is of course one of 2^w possible subsets of S_i.

From these four pieces of information one can uniquely reconstruct (W, S_i). Thus the total number of bad pairs (W, S_i) is bounded by

$$p^{-w}\binom{n}{pn} \cdot 2^w \cdot s_{w'} \cdot 2^w = (4/p)^w s_{w'}\binom{n}{pn}.$$

The number of sets $W \subset X$ of size $|W| = p|X|$ is $\binom{n}{pn}$. The lemma follows by taking expectation over W. $\square$

The following is a corollary of Lemma 5.7, where we replace sampling $W \subset X$ of size $|W| = p|X|$ with sampling $W \sim \mathcal{U}(X, p)$.

Corollary 5.8. *Let $(\mathcal{F}, \sigma)$ be an $\mathbf{s} = (s_0; s_1, \ldots, s_w)$-bounded weighted set system on X. Let $W \sim \mathcal{U}(X, p)$ and $\mathcal{B}(W) = \{S \in \mathcal{F} : (W, S) \text{ is bad}\}$. Then $\mathbb{E}_W[\sigma(\mathcal{B}(W))] \le (4/p)^w s_{w'}$.*

Proof. The proof is by a reduction to Lemma 5.7. Replace the base set X with a much larger set X' (without changing $\mathcal{F}$, so the new elements do not belong to any set in $\mathcal{F}$). Let $W' \subset X'$ be a uniform set of size $|W'| = p|X'|$, and let $W = W' \cap X$. Then as X' gets bigger, the distribution of W' approaches $\mathcal{U}(X, p)$, while the conclusion of the lemma depends only on W. $\square$

Proof of Lemma 5.6. Let $(\mathcal{F}, \sigma)$ be an $\mathbf{s} = (s_0; s_1, \ldots, s_w)$-bounded weighted set system on X. Let $W \sim \mathcal{U}(X, p)$. Say that W is δ-bad if $\sigma(\mathcal{B}(W)) \ge \delta s_0$. By applying Corollary 5.8 and Markov's inequality, we obtain that

$$\Pr[W \text{ is } \delta\text{-bad}] \le \frac{\mathbb{E}[\sigma(\mathcal{B}(W))]}{\delta s_0} \le \frac{(4/p)^w s_{w'}}{\delta s_0}.$$

Fix W which is not δ-bad. By assumption, if (W, S) is good for $S \in \mathcal{F}$, then there exists $\pi(S) = S' \in \mathcal{F}$ (possibly with $S' = S$) such that (i) $S' \setminus W \subset S \setminus W$ and (ii) $|S' \setminus W| \le w'$. Choose such π with the smallest possible image so that if S', S'' in the image of π are distinct then $S' \setminus W \ne S'' \setminus W$.

Define a new weighted set system $(\mathcal{F}', \sigma')$ on $X' = X \setminus W$ as follows:

$$\mathcal{F}' = \{\pi(S) \setminus W : S \in \mathcal{F} \setminus \mathcal{B}(W)\}, \quad \sigma'(S' \setminus W) = \sigma(\pi^{-1}(S')).$$

We claim that $\mathcal{F}'$ is $\sigma' = ((1-\delta)s_0; s_1, \ldots, s_{w'})$-bounded. To see that, note that $\sigma'(\mathcal{F}') = \sigma(\mathcal{F} \setminus \mathcal{B}(W)) \ge (1-\delta)s_0$ and that for any set $T \subset X'$,

$$\sigma'(\mathcal{F}'_T) = \sum_{S' \supset T} \sigma'(S') = \sum_{S : \pi(S) \supset T} \sigma(S) \le \sum_{S \supset T} \sigma(S) = \sigma(\mathcal{F}_T) \le s_{|T|}.$$

Finally, all sets in $\mathcal{F}'$ have size at most w'. Thus, if we choose $W' \sim \mathcal{U}(X', \alpha')$ then we obtain that with probability more than $1-\beta'$, there exist $S^* \in \mathcal{F}'$ such that $S^* \subset W'$. Recall that $S^* = S \setminus W$ for some $S \in \mathcal{F}$. Thus $S \subset W \cup W'$, which is distributed according to $\mathcal{U}(p+(1-p)\alpha')$. □

5.2 A final step

In this section, we directly show that bounded set systems (with very good bounds) are satisfying. A similar argument appears in [11].

Lemma 5.9. *Let $0 < \alpha, \beta < 1$, $w \geq 2$, and set $\kappa = 4\ln 2 \cdot w \cdot \log(1/\beta)/\alpha$. Let $(\mathcal{F}, \sigma)$ be an $\mathbf{s} = (s_0; s_1, \ldots, s_w)$-bounded weighted set system where $s_i \leq \kappa^{-i} s_0$. Then $\mathcal{F}$ is (α, β)-satisfying.*

Proof. We may assume without loss of generality that all sets in $\mathcal{F}$ have size exactly w, by adding dummy elements to each set of size below w. Note that this new set system $\mathcal{F}'$ satisfies the assumption of the lemma, and that for any set $W \subset X$, if W contains a set of $\mathcal{F}'$ then it also contains a set of $\mathcal{F}$. We can also assume by scaling that $N_S = \sigma(S)$ for $S \in \mathcal{F}$ are all integers. Let $\mathcal{F}'$ be the multi-set system, where each $S \in \mathcal{F}$ is repeated N_S times. Let $N = \sum N_S \geq s_0$ and denote $\mathcal{F}' = \{S_1, \ldots, S_N\}$.

The proof is by Janson's inequality (see for example [1, Theorem 8.1.2]). Let $W \sim \mathcal{U}(X, \alpha)$ and $\mathcal{Z}_i$ be the indicator variable for $S_i \subset W$. Denote $i \sim j$ if S_i, S_j intersect. Define

$$\mu = \sum_i \mathbb{E}[\mathcal{Z}_i], \qquad \Delta = \sum_{i \sim j} \mathbb{E}[\mathcal{Z}_i \mathcal{Z}_j].$$

We have $\mu = N\alpha^w$. To compute Δ, let p_ℓ denote the fraction of pairs (i,j) such that $|S_i \cap S_j| = \ell$. Then

$$\Delta = \sum_{\ell=1}^{w} p_\ell N^2 \alpha^{2w-\ell}.$$

To bound p_ℓ, note that for each $S_i \in \mathcal{F}$, and any $R \subset S_i$ of size $|R| = \ell$, the number of $S_j \in \mathcal{F}$ such that $R \subset S_j$ is $|\mathcal{F}_R| \leq N/\kappa^{|R|}$. Thus we can bound

$$\Delta \leq \sum_{\ell=1}^{w} \binom{w}{\ell} \kappa^{-\ell} N^2 \alpha^{2w-\ell} \leq \sum_{\ell=1}^{w} \left(\frac{w}{\alpha\kappa}\right)^\ell \mu^2.$$

Let $\kappa = qw/\alpha$ for $q \geq 2$. Then $\Delta \leq 2\mu^2/q$. Note that in addition $\Delta \geq \mu$, as we include in particular the pairs (i,i) in Δ. Thus by Janson's inequality,

$$\Pr[\forall i, \mathcal{Z}_i = 0] \leq \exp\left\{-\frac{\mu^2}{2\Delta}\right\} \leq \exp\left\{-\frac{q}{4}\right\}.$$

The lemma follows by setting $q = 4\ln 2 \cdot \log(1/\beta)$. □

5.3 Putting everything together

We prove Theorem 5.5 in this subsection, where our goal is to bound $\kappa(w, \alpha, \beta)$. We will apply Lemma 5.6 iteratively, until we decrease w enough to apply Lemma 5.9.

Let $w \geq 2$ to be fixed throughout, and $\kappa > 1$ to be optimized later. We first introduce some notation. For $0 < \Delta < 1, \ell \geq 1$, let $\mathbf{s}(\Delta, \ell) = (1 - \Delta; \kappa^{-1}, \ldots, \kappa^{-\ell})$ be a weight profile. Let $A(\Delta, \ell), B(\Delta, \ell)$ be bounds such that any $\mathbf{s}(\Delta, \ell)$-bounded set system is $(A(\Delta, \ell), B(\Delta, \ell))$-satisfying.

Lemma 5.6 applied to $w' \geq w''$ and p, δ gives the bound

$$\begin{aligned} A(\Delta, w') &\leq A(\Delta + \delta, w'') + p, \\ B(\Delta, w') &\leq B(\Delta + \delta, w'') + \frac{(4/p)^{w'}}{\delta(1 - \Delta)\kappa^{w''}}. \end{aligned}$$

We apply this iteratively for some widths $w_0, \ldots, w_r$. Set $w_0 = w$ and $w_{i+1} = \lceil (1 - \varepsilon) w_i \rceil$ for some ε as long as $w_i > w^*$ for some w^*. In particular, we need $w^* \geq 1/\varepsilon$ to ensure $w_{i+1} < w_i$ and we will optimize ε, w^* later. The number of steps is thus $r \leq (u \log w)/\varepsilon$ for some constant $u > 0$. Let $p_1, \ldots, p_r$ and $\delta_1, \ldots, \delta_r$ be the values we use for p, δ at each iteration. To simplify the notation, let $\Delta_i = \delta_1 + \cdots + \delta_i$ and $\Delta_0 = 0$. Furthermore, define

$$\gamma_i = \frac{(4/p_i)^{w_{i-1}}}{\kappa^{w_i}}.$$

Then for $i = 1, \ldots, r$, we have

$$\begin{aligned} A(\Delta_{i-1}, w_{i-1}) &\leq A(\Delta_i, w_i) + p_i, \\ B(\Delta_{i-1}, w_{i-1}) &\leq B(\Delta_i, w_i) + \frac{\gamma_i}{\delta_i(1 - \Delta_{i-1})}. \end{aligned}$$

Set $p_i = \alpha/(2r)$ and $\delta_i = \sqrt{\gamma_i}$. We will select the parameters so that $\Delta_i \leq 1/2$ for all i. Thus

$$\begin{aligned} A(0, w) &\leq A(\Delta_r, w_r) + \alpha/2 \leq A(1/2, w^*) + \alpha/2, \\ B(0, w) &\leq B(\Delta_r, w_r) + 2\Delta_r \leq B(1/2, w^*) + 2\Delta_r. \end{aligned}$$

Plugging in the values for δ_i, we compute the sum

$$\Delta_r = \sum_{i=1}^{r} \delta_i \leq \sum_{i=1}^{r} \sqrt{\frac{(4/p)^{w_{i-1}}}{\kappa^{(1-\varepsilon)w_{i-1}}}} \leq \sum_{k \geq w^*} \left(\frac{u \log w}{\varepsilon \alpha \kappa^{1-\varepsilon}} \right)^{k/2} \leq 2 \left(\frac{u \log w}{\varepsilon \alpha \kappa^{1-\varepsilon}} \right)^{w^*/2},$$

assuming $\kappa^{1-\varepsilon} = \Omega\left((\log w)/(\varepsilon\alpha)\right)$. More precisely, if we take κ so that

$$\kappa^{1-\varepsilon} = \frac{K \cdot u \log w}{\varepsilon \alpha}, \qquad K \geq 4,$$

then $\Delta_r \leq 2K^{-w^*/2}$.

Next, we apply Lemma 5.9 to bound $A(1/2, w^*) \leq \alpha/2$ and $B(1/2, w^*) \leq \beta/2$. We use the simple observation that $(1/2; \kappa^{-1}, \ldots, \kappa^{-w^*})$-bounded set systems are also $(1; (\kappa/2)^{-1}, \ldots, (\kappa/2)^{-w^*})$-bounded, in which case we can apply Lemma 5.9 and obtain that we need

$$\kappa \geq \Omega(w^* \cdot \log(1/\beta)/\alpha).$$

Let us now put the bounds together. We still have the freedom to choose $\varepsilon > 0$ and $w^* \geq 1/\varepsilon$. To obtain $A(0, w) \leq \alpha, B(0, w) \leq \beta$, we also need $\Delta_r \leq \beta/2 < 1/2$. Thus all the constraints are:

1. $w^* \geq 1/\varepsilon$;
2. $\kappa^{1-\varepsilon} = (K \cdot u \log w)/(\varepsilon\alpha)$ for some constant $K \geq 4$;
3. $\kappa \geq \Omega(w^* \log(1/\beta)/\alpha)$;
4. $2K^{-w^*/2} \leq \beta/2 \Longleftarrow w^* \geq \Omega(\log(1/\beta)/\log K)$.

Set $\varepsilon = 1/\log\log w$ and $w^* = c \cdot \max\{\log\log w, \log(1/\beta)\}$ for some $c \geq 1$. Then we obtain that the result holds whenever

$$\kappa \geq c \log w \cdot (\log\log w \cdot \log(1/\beta)/\alpha)^{c'}.$$

For large enough w, the exponent can be taken to be $c' = 1 + o(1)$.

Acknowledgements

We thank Noga Alon for helpful suggestions on an earlier version of this paper.

References

[1] N. Alon and J. H. Spencer. *The probabilistic method.* John Wiley & Sons, 2004.

[2] P. Erdős and R. Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, 35(1):85–90, 1960.

[3] P. Erdös and E. Szemerédi. Combinatorial properties of systems of sets. *Journal of Combinatorial Theory, Series A*, 24(3):308–313, 1978.

[4] J. Håstad. *Computational limitations of small-depth circuits.* MIT Press, Cambridge, MA, USA, 1987.

[5] A. Kostochka. A bound of the cardinality of families not containing Δ-systems. In *The Mathematics of Paul Erdös II*, pages 229–235. Springer, 1997.

[6] X. Li, S. Lovett, and J. Zhang. Sunflowers and quasi-sunflowers from randomness extractors. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[7] S. Lovett, N. Solomon, and J. Zhang. From DNF compression to sunflower theorems via regularity. In *34th Computational Complexity Conference, CCC 2019, July 18-20, 2019, New Brunswick, NJ, USA.*, pages 5:1–5:14, 2019.

[8] S. Lovett and J. Zhang. DNF sparsification beyond sunflowers. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 454–460. ACM, 2019.

[9] E. Naslund and W. Sawin. Upper bounds for sunflower-free sets. In *Forum of Mathematics, Sigma*, volume 5. Cambridge University Press, 2017.

[10] A. A. Razborov. Bounded arithmetic and lower bounds in boolean complexity. In *Feasible Mathematics II*, pages 344–386. Springer, 1995.

[11] B. Rossman. The monotone complexity of k-clique on random graphs. *SIAM J. Comput.*, 43(1):256–279, 2014.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il