Paper Reading – Sunflowers: from soil to oil

Ziyang Jin[1]

Department of Computer Science
University of Toronto
ziyang@cs.toronto.edu

Abstract. This note is prepared for presenting the paper *Sunflowers: from soil to oil*[1] in the theory reading group at University of Toronto in summer 2023. A *sunflower* is a collection of sets whose pairwise intersections are identical. We will introduce two related concepts: sunflower and the threshold of monotone functions. The paper introduces a main theorem, which can be used to prove 1. the newest result in the size of set that contains a sunflower, and 2. the Kahn-Kalai conjecture of threshold vs. expectation threshold for monotone functions.

Keywords: Complexity · Erdős-Rado sunflower conjecture · Kahn-Kalai conjecture · probabilistic combinatorics.

Disclaimer: This note is essentially a selected copy-paste from Anup Rao's paper *Sunflowers: from soil to oil* and his YouTube video *The Sunflower Lemma and Monotone Thresholds* with some of the comments and explanations made by me to help people in the reading group better understand the original paper.

1 Introduction to Sunflowers

Definition 1 (Sunflower). *A sunflower is a collection of sets whose pairwise intersections are identical.*

A sunflower with w petals is a collection of w sets whose pairwise intersections are identical. The common intersection is called the *core*. Note that the core can be an empty set, i.e., a collection of pairwise disjoint sets is also a sunflower. On the other hand, a collection of sets each containing exactly the same elements is also trivially a sunflower.

Sunflower was originally called Δ-systems in the paper by Erdős and Rado[4] in 1960. The name *sunflower* was given by Deza and Frankl [5] and is now widely accepted. In Erdős and Rado [4]'s paper, they proved that every collection of more than $k!(w-1)^k$ sets of size at most k must contain a sunflower with w petals. In the same paper, they conjectured that there is a constant c such that every family of $(cw)^k$ sets of size k contains a sunflower with w petals.

In 2019, Alweiss, Lovett, Wu, and Zhang [3] proved for $w \geq 3$, there exists some constant c, such that any k-set system $\mathcal{S}$ of size $|\mathcal{S}| \geq (cw^3 \log k \log\log k)^k$ contains a w-sunflower. Subsequently, Rao [8], Frankston, Kahn, Narayanan and

Park [6] and Bell, Chueluecha, and Warnke [7] further improved it to $(cw \log k)^k$ for some constant c. This is the best known result so far for the sunflower conjecture, which is a $\log k$ term off the conjectured size.

Below is a picture taken from Rao's YouTube video that shows a sunflower with 4 petals exists in the given set.

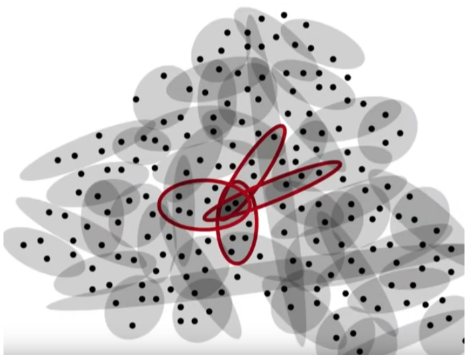

Fig. 1. Figure taken from Anup Rao's YouTube video [2]

2 Threshold of Monotone Functions

Definition 2 (Monotone function). *Function* $f : 2^{\{1,\ldots,n\}} \to \{0,1\}$ *is monotone if* $S \subseteq T$ *implies* $f(S) \leq f(T)$.

For example, a monotone function can take a graph as the input and output 1 if the graph contains a K_5 (complete graph of 5 vertices).

Definition 3 (Family of minimal sets). *Let* $f : 2^{\{1,\ldots,n\}} \to \{0,1\}$ *be a monotone function, define the family of minimal sets* $\mathcal{F}$ *to be a collection of minimal sets* X*, where* $f(X) = 1$.

See the figure below, the set $\mathcal{F}$ containing all bold black dots are a collection of minimal sets in f.

Let P be a random set where every element in $\{1, ..., n\}$ is independently drawn to set P with probability p; let Q be a random set where every element in $\{1, ..., n\}$ is independently drawn to set Q with probability q.

Definition 4 (Threshold of monotone function). *The threshold of* f *is the minimal probability* p *such that* $\mathbb{E}[f(P)] = 1/2$.

For any non-trivial (i.e. f is not always 0 or not always 1) monotone function f, when $p = 0$, then $\mathbb{E}[f(P)] = 0$ since P will be an empty set. When $p = 1$, $\mathbb{E}[f(P)] = 1$ since every element will be selected. Therefore, as p is increasing

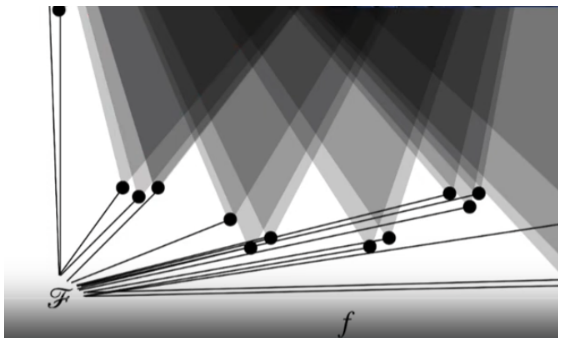

Fig. 2. Figure taken from Anup Rao's YouTube video [2]

from 0 to 1, there must be a value of p that makes the expectation equal to $\frac{1}{2}$ exactly. It is interesting to understand the threshold p for different monotone functions, as the threshold captures something about the structure of f.

Definition 5 (Shadow). *Given a family of sets $\mathcal{F}$ and a set X, define the shadow $\mathcal{F}_X = \{F \in \mathcal{F} : F \subseteq X\}$.*

I would like to think about $\mathcal{F}_X$ as X's shadow projected on $\mathcal{F}$. It is easy to see that every monotone function f is associated with a minimal collection of sets $\mathcal{F}$ such that $f(X) = 1$ for every $X \in \mathcal{F}$. Moreover, if $\mathcal{F}$ is the family of minimal sets of f, then $f(X) = 1$ if and only if $|\mathcal{F}_X| \geq 1$.

Suppose we have a monotone function f, and $\mathcal{F}$ is the minimal family of f. Let $\mathbf{X} \in 2^{[n]}$ be a random set, where is element in $\{1, ..., n\}$ is drawn to $\mathbf{X}$ with probability ϵ. So we have

$$\mathbb{E}[f(\mathbf{X})] = \Pr[\bigcup_{Y \in \mathcal{F}} (Y \subseteq \mathbf{X})] \leq \sum_{Y \in \mathcal{F}} \Pr[Y \subseteq \mathbf{X}] = \mathbb{E}[|\mathcal{F}_{\mathbf{X}}|]$$

The middle inequality is due to union bound.

Thus, if $\mathbf{X}$ is a random set, then the expectation of $f(\mathbf{X})$ is less than or equal to the expected size of the shadow of $\mathbf{X}$ in $\mathcal{F}$.

More generally, for every monotone function g where $f(X) \leq g(X)$ for every set X, and suppose $\mathbf{X} \in 2^{[n]}$ is a random set, and $\mathcal{G}$ is the family of minimal sets of g, then we have the bound

$$\mathbb{E}[f(\mathbf{X})] \leq \mathbb{E}[g(\mathbf{X})] \leq \mathbb{E}[|\mathcal{G}_{\mathbf{X}}|]$$

Why do we want to bound $\mathbb{E}[f(\mathbf{X})]$ by $\mathbb{E}[|\mathcal{G}_{\mathbf{X}}|]$? Suppose f is a complicated function, where it contains a lot of small sets in the family of minimal set $\mathcal{F}$. Thus, it makes finding the threshold of f hard. But we can find a much nicer/simpler function g, that "covers" f, and it is much easier to compute the expected size of shadow of $\mathbf{X}$ in the family of minimal sets $\mathcal{G}$. We hope $\mathbb{E}[|\mathcal{G}_{\mathbf{X}}|]$ would be a good/close estimate of $\mathbb{E}[f(\mathbf{X})]$.

Definition 6 (Expectation Threshold). *The expectation threshold of f is the largest value of q such that $\mathbb{E}[|\mathcal{G}_Q|] = \frac{1}{2}$ for some monotone function g with $f \leq g$.*

For example, if f is a boolean function that computes whether a graph has a perfect matching, the threshold $p \approx \frac{\log n}{n}$, while the expectation threshold $q \approx \frac{1}{n}$. In 2006, Kahn and Kalai conjectured that the threshold is always at most $O(\log n)$ times greater than the expectation-threshold[10].

Theorem 1 (Kahn-Kalai Conjecture (Resolved)). *For any monotone boolean function $f : \{0,1\}^n \to \{0,1\}$, the threshold p is at most $O(\log n)$ times larger than the expectation threshold q.*

The conjecture is proven by Park and Pham [9] in 2022, following a similar idea to 2019 paper by Alweiss, Lovett, Wu, and Zhang in finding sunflowers.

3 Relationship Between Threshold and Sunflower

Suppose you have a family of minimal set $\mathcal{F}$, and you can define function f on $\mathcal{F}$. Let $\mathbf{W}$ be a uniform random set of size $\frac{n}{2w}$ drawn from $\{1, ..., n\}$, and $\mathbb{E}[f(\mathbf{W})] = \frac{1}{2}$, so the threshold is like $p = \frac{1}{2w}$.

Let $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_{2w}$ be a uniform random partition of $\{1, ..., n\}$, each has the same size $\frac{n}{2w}$. Then we have

$$\sum_{i=1}^{2w} \mathbb{E}[f(\mathbf{W}_i)] = 2w \times \frac{1}{2} = w$$

which means there exists a fixed partition $W_1, ..., W_{2w}$ where we can find $\geq w$ disjoint sets $W'_1, ..., W'_w$ and $f(W'_1), ..., f(W'_w)$ are evaluates to 1, which means we can find w disjoint minimal sets from $\mathcal{F}$, a trivial sunflower with an empty core.

4 Main Theorem and Proof

4.1 Statement of the main theorem

Definition 7 (r-spread). *Given a collection of sets $\mathcal{S}$, let $\mathbf{U} \in \mathcal{S}$ be uniformly random. We shall say that $\mathbf{U}$ is r-spread if for every set Z, $\Pr[Z \subseteq \mathbf{U}] \leq r^{-|Z|}$.*

Here Z is any set since if Z contains other unrelated elements, Z will not be a subset of $\mathbf{U}$ thus the probability will be 0. So it does not matter what Z is. We think r-spread is more of a property of $\mathcal{S}$ rather than a property of $\mathbf{U}$, but we will be consistent with the notation in the paper.

Here is how the concept of r-spread related to sunflowers. If $\mathbf{U}$ is r-spread, then we are likely to find a collection of pairwise disjoint sets, which is a trivial sunflower. If U is not r-spread, then that means there exists a set Z (that acts like the core of the sunflower), such that $\Pr(Z \subseteq \mathbf{U}) \geq r^{-|Z|}$. Suppose $|\mathcal{S}| \geq r^k$. That means there are at least $r^k \times r^{-|Z|} = r^{k-|Z|}$ sets in $\mathcal{S}$ such that Z is a subset of. We call this family of sets $\mathcal{S}' = \{S \in \mathcal{S} : Z \subseteq S\}$ and so $|\mathcal{S}'| \geq r^{k-|Z|}$. And we obtain a new family of sets $\mathcal{S}'^{-} = \{S \setminus Z : S \in \mathcal{S}'\}$ by deleting Z from each element S, and we will inductively find sunflower in $\mathcal{S}'^{-}$, and then adding Z back will give a sunflower in the original set $\mathcal{S}$.

Theorem 2 (Main Theorem). *Let $\mathcal{S} \subseteq 2^{[n]}$ be a family of sets of size at most k. Then there is a distribution on pairs $(\mathbf{W}, \mathcal{G})$, where $\mathbf{W} \in 2^{[n]}$ is a uniformly random set of size ϵn and $\mathcal{G} \subseteq 2^{[n]}$ is a family of sets, then we have the following two guarantees:*

1. either $\mathcal{S}_{\mathbf{W}} \neq \emptyset$, or for every $S \in \mathcal{S}, \mathcal{G}_S \neq \emptyset$ and

2. for any r-spread $\mathbf{U}$ that is independent of $(\mathbf{W}, \mathcal{G})$ with $r = \frac{64 \log k}{\epsilon}$, we have $\mathbb{E}[|\mathcal{G}_{\mathbf{U}}|] < \frac{1}{8}$.

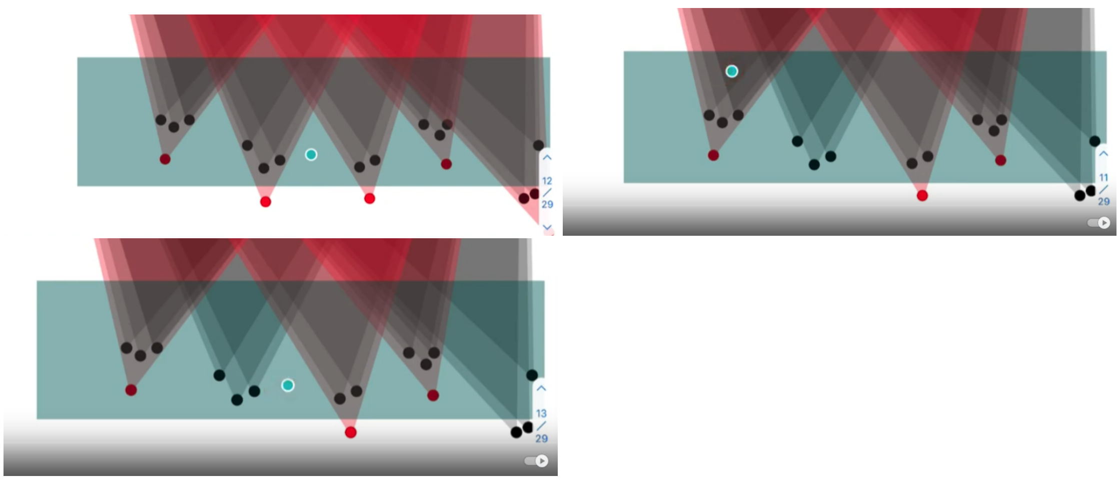

Fig. 3. Figure taken from Anup Rao's YouTube video [2]

Basically, we can think $\mathcal{S}$ as the family of minimal sets that defines the monotone function f. Then for every probability ϵ, we can draw a uniform random set $\mathbf{W}$ of size ϵn from $\{1, ..., n\}$, and we are able to find a family of minimal sets $\mathcal{G}$ that defines the monotone function g.

In the figure above, the grey area represents the function f, and the red area represents the function g. The first condition of the main theorem says that, it is either the case where g covers f (the top left picture), or the $\mathbf{W}$ we draw lies in the grey area of f (the top right picture). It cannot be the case that the $\mathbf{W}$ we draw is outside of the grey area and g does not cover f (the bottom picture).

Such that it is either $\mathcal{S}_{\mathbf{W}} \neq \emptyset$, which means $f(\mathbf{W}) = 1$, or for every minimal set $S \in \mathcal{S}$, we have $\mathcal{G}_S \neq \emptyset$, which means $g \geq f$. The second condition basically says that $\mathcal{G}$ cannot have too many small sets.

4.2 Proof of the main theorem

Proof. Let $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_{\log k}$ be uniformly random disjoint sets of size $m = \frac{\epsilon n}{\log k}$. Here all logs are base 2. Our goal is to use $\mathbf{W}_1, ..., \mathbf{W}_{\log k}$ to define a sequence of sets $\mathcal{G}_1, ..., \mathcal{G}_{\log k}$. Eventually, we will set $\mathbf{W} = \mathbf{W}_1 \cup ... \cup \mathbf{W}_{\log k}$ (since $\mathbf{W}_i$'s are disjoint, then the final set $\mathbf{W}$ will have size ϵn) and $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup ... \cup \mathcal{G}_{\log k}$. (i.e. we are gradually constructing the family of minimal sets that defines monotone function g which is supposed to "cover" f, and we do not want to include too many small sets in $\mathcal{G}$).

Let $\mathbf{W}^i := \mathbf{W}_1 \cup ... \cup \mathbf{W}_i$, and let $\mathcal{G}^i := \mathcal{G}_1 \cup ... \cup \mathcal{G}_i$. Define $\mathcal{G}_1, ..., \mathcal{G}_{\log k}$ iteratively as follows. For each i, and for each $S \in \mathcal{S}$, include $T := S - W^i$ in $\mathcal{G}_i$ if and only if

(i). $|T| \geq \frac{k}{2^i}$, and

(ii). T is a minimal set of $\{S - W^i : S \in \mathcal{S}, \mathcal{G}_S^{i-1} = \emptyset\}$

We could similarly define g^i to be the function defined by $\mathcal{G}^i$. The second condition says $T := S - W^i$ is a minimal set of $\{S - W^i : S \in \mathcal{S}, f(S) > g^{i-1}(S)\}$, so in the ith round, we want to add the sets that have not been covered by g^{i-1}.

Intuitively, the above process attempts to *cover* all the sets of $\mathcal{S}$. In each step, we discard the elements of $\mathcal{S}$ that have already been covered (either by $\mathbf{W}^i$ or by $\mathcal{G}^i$), and proceed to cover more elements by including sets of size at least $\frac{k}{2^i}$ in $\mathcal{G}_i$. By the time $i = \log k$, then the first condition becomes $|S - \mathbf{W}^i| \geq k/2^{\log k} = 1$, which means in the final round, we will cover all remaining sets that are not included in $\mathbf{W}_1 \cup ... \cup \mathbf{W}_{\log k}$. So, a set of $\mathcal{S}$ is left uncovered in the process only if it is contained in $\mathbf{W} = \mathbf{W}^{\log k}$. This proves the first guarantee of the theorem, that either $\mathcal{S}_{\mathbf{W}} \neq \emptyset$ or $g \geq f$ (i.e., for every $S \in \mathcal{S}, \mathcal{G}_S \neq \emptyset$).

Now we need to prove the bound $\mathbb{E}[|\mathcal{G}_{\mathbf{U}}|] < \frac{1}{8}$. The idea is that we do not want to include too many small sets, as small sets in $\mathcal{G}$ will increase the expectation of $|\mathcal{G}_{\mathbf{U}}|$. We start by giving an upper bound on the expected number of sets $T \in \mathcal{G}_i$ of size a. Then eventually we do a summary of size a from $k/2^i$ to ∞ to include every potential set to be added, and we will bound that sum.

Claim: expected number of sets T in $\mathcal{G}_i$ of size a is at most $(\frac{\log k}{\epsilon})^a \cdot 4^{k/2^i}$.

Fix $\mathbf{W}_1, ..., \mathbf{W}_{i-1}$. First we bound the number of choices of $\mathbf{W}_i \cup T$.

(i). Let n_i denote the size of the universe after deleting $\mathbf{W}^{i-1}$. Note that each $\mathbf{W}_i$ is of size m and we consider T with size a. So there are at most $\binom{n_i}{m+a}$ choices for the set $T \cup \mathbf{W}_i$. We have

$$\binom{n_i}{m+a} = \frac{n_i!}{(m+a)!(n_i - m - a)!} = \frac{n_i!(n_i - m - 1)...(n_i - m - a)}{m!(m+1)...(m+a)(n_i - m)!} = \binom{n_i}{m}$$

$$\binom{n_i}{m+a} = \binom{n_i}{m} \prod_{j=1}^{a} \frac{n_i - m - j}{m + j} \leq \binom{n_i}{m} \left(\frac{n_i}{m}\right)^a$$

and notice that $m = \frac{\epsilon n}{\log k}$, so $\frac{n_i}{m} = \frac{\log k n_i}{\epsilon n} \leq \frac{\log k}{\epsilon}$. Therefore,

$$\binom{n_i}{m+a} \leq \binom{n_i}{m} \left(\frac{\log k}{\epsilon}\right)^a$$

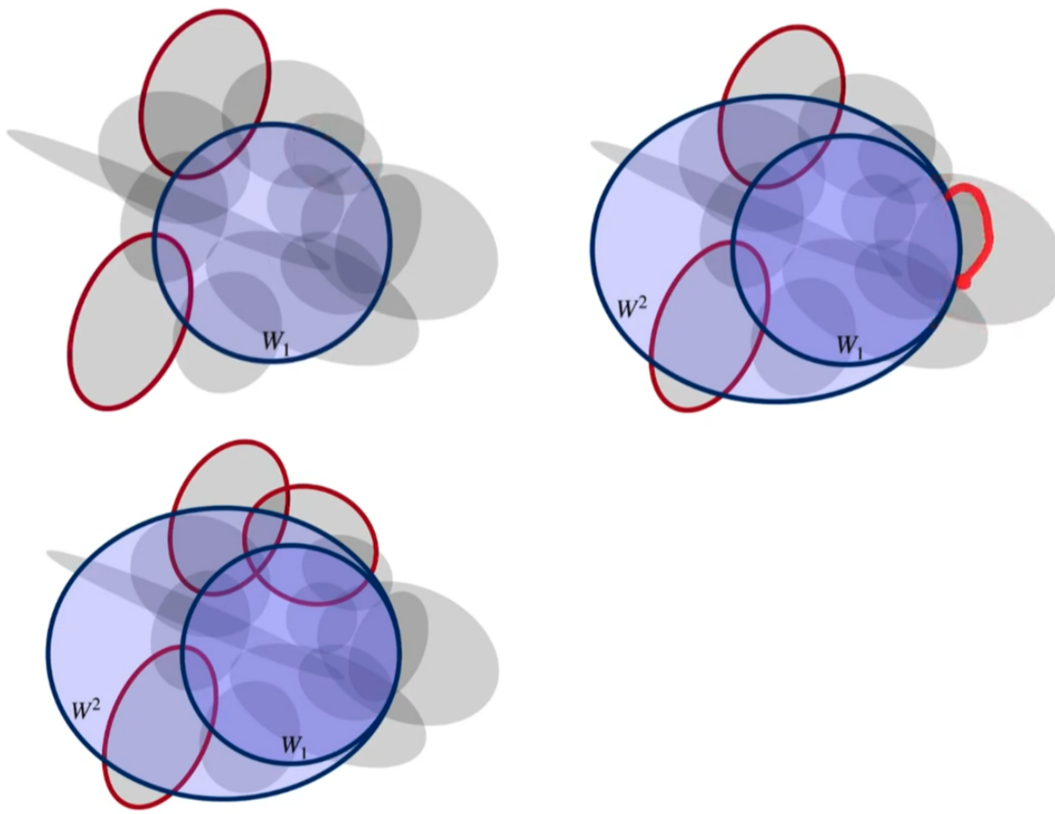

Fig. 4. Figure taken from Anup Rao's YouTube video [2]

Then we want to say, once the union $\mathbf{W}_i \cup T$ is specified, there is relatively few choices for T.

(ii). Given a fixed $T \cup \mathbf{W}_i$, let $T' := S' - W^{i-1}$ be the smallest set of $\{S - \mathbf{W}^{i-1} : S \in \mathcal{S}, \mathcal{G}_S^{i-1} = \emptyset\}$ that is contained in $T \cup \mathbf{W}_i$; break ties by picking the lexicographically first set. So T' is a candidate for generating T, and T must be a subset of T', otherwise $S' - W^i$ would be a strict subset of T, and T would not be included in $\mathcal{G}_i$. Secondly, it must be that $|T'| \leq k/2^{i-1}$, otherwise T' would have been included in the previous round $\mathcal{G}_{i-1}$. Since $|T'| \leq k/2^{i-1}$ and T must be a subset of T', then there can be at most $2^{k/2^{i-1}} = 4^{k/2^i}$ choices of T consistent with $T \cup \mathbf{W}_i$.

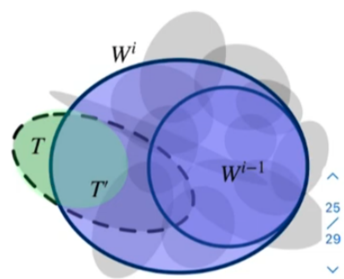

Fig. 5. Figure taken from Anup Rao's YouTube video [2]

The above count shows that the expected number of sets T of size a in $\mathcal{G}_i$ is at most $4^{k/2^i}\left(\frac{\log k}{\epsilon}\right)^a$. Thus we can bound

$$\begin{aligned}
\mathbb{E}[|\mathcal{G}_U|] &\leq \mathbb{E}[\sum_{Y \in \mathcal{G}} (\frac{\epsilon}{64 \log k})^{|Y|}] \\
&= \sum_{i=1}^{\log k} \mathbb{E}[\sum_{Y \in \mathcal{G}_i} \left(\frac{\epsilon}{64 \log k}\right)^{|Y|}] \\
&\leq \sum_{i=1}^{\log k} \sum_{a=k/2^i}^{\infty} \left(\frac{\epsilon}{64 \log k}\right)^a \cdot 4^{k/2^i} \left(\frac{\log k}{\epsilon}\right)^a \\
&= \sum_{i=1}^{\log k} \frac{(1/16)^{k/2^i}}{1 - 1/64} \\
&< \sum_{j=1}^{\infty} \frac{64}{63} (\frac{1}{16})^j \\
&< \frac{1}{8}
\end{aligned}$$

The second line is by linearity of expectation where we separate the contribution from each round $\mathcal{G}^i$. The third line is using the bound we just proved, which we can see that the numbers are being set-up so they cancel each other, leaving a sum unrelated to ϵ. The last two lines are just sum of geometric sequence. □

5 Using the Main Theorem to prove sunflower and Kahn-Kalai conjecture

5.1 Proof of sunflower lemma

Lemma 1 (Sunflower Lemma). *If* $|\mathcal{S}| \geq q^{-k} = (128w \log k)^k$*, then we can find a sunflower with* w *petals in* $\mathcal{S}$*.*

Proof. Let f be the monotone function defined by $\mathcal{S}$. Let $\mathbf{U}$ be a uniform random set drawn from $\mathcal{S}$.

Let $\epsilon = \frac{1}{2w}$, then $r = \frac{64 \log k}{\epsilon} = 128w \log k$, and $q = \frac{1}{128w \log k}$. Let $\mathbf{W}$ be a random set where every element in $\{1, ..., n\}$ is independently drawn to set $\mathbf{W}$ with probability ϵ. Let $\mathbf{Q}$ be a random set where every element in $\{1, ..., n\}$ is independently drawn to set $\mathbf{Q}$ with probability q. There are 2 cases.

Case 1: For every set Z, $\Pr(Z \subseteq \mathbf{U}) \leq \Pr(Z \subseteq \mathbf{Q})$. In this case,

$$\Pr(g \geq f) = \Pr(\forall U \in \mathcal{S}, \mathcal{G}_U \neq \emptyset) = \Pr(\forall U \in \mathcal{S}, |\mathcal{G}_U| \geq 1) \leq \mathbb{E}[|\mathcal{G}_\mathbf{U}|]$$

and

$$\mathbb{E}[|\mathcal{G}_\mathbf{U}|] = \sum_{G \in \mathcal{G}} \Pr(G \subseteq \mathbf{U})$$

Since $\Pr(Z \subseteq \mathbf{U}) \leq \Pr(Z \subseteq \mathbf{Q})$ for any set Z, and G such a set, so

$$\mathbb{E}[|\mathcal{G}_\mathbf{U}|] = \sum_{G \in \mathcal{G}} \Pr(G \subseteq \mathbf{U}) \leq \sum_{G \in \mathcal{G}} \Pr(G \subseteq \mathbf{Q}) = \mathbb{E}[|\mathcal{G}_\mathbf{Q}|] < \frac{1}{8}$$

Thus we get

$$\Pr(g \geq f) < \frac{1}{8}$$

and applying the first "either-or" property of the main theorem, we know that the $\Pr[f(\mathbf{W}) = 1] > \frac{7}{8}$.

Let $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_{2w}$ be a random partition of $\{1, ..., n\}$, so each set has size $\frac{n}{2w}$. Let $\frac{n}{2w} = \epsilon n$, we get $\epsilon = \frac{1}{2w}$, and $r = 128w \log k$.

$$\sum_{i=1}^{2w} \mathbb{E}[f(\mathbf{W}_i)] > \frac{7}{8} \times 2w = \frac{7}{4} w$$

So there must exist an instance of $W_1, W_2, ..., W_{2w}$ such that at least $\frac{7}{4}w$ of them will have f evaluate to 1. Consider $\mathcal{S}_{W_i}$, that means we can find at least $\frac{7}{4}w$ disjoint sets (since the W_i's are disjoint) in $\mathcal{S}$, which is a sunflower of at least $\frac{7}{4}w$ petals in $\mathcal{S}$.

Case 2: There exists some set Z such that $\Pr(Z \subseteq \mathbf{U}) > \Pr(Z \subseteq \mathbf{Q})$. Let $\mathcal{S}'^- = \{S \setminus Z : Z \subseteq S, S \in \mathcal{S}\}$. Since $|\mathcal{S}| \geq q^{-k}$, then we have $|\mathcal{S}'^-| \geq q^{-(k-|Z|)}$, and the sets in $\mathcal{S}'^-$ have size at most $k - |Z|$, so we can inductively find sunflower in $\mathcal{S}'^-$. Once we find a sunflower in $\mathcal{S}'^-$, we can put Z back to the sunflower and obtain a sunflower in $\mathcal{S}$. □

5.2 Proof of Kahn-Kalai conjecture

Let us reiterate the statement: For any monotone boolean function $f : \{0,1\}^n \to \{0,1\}$, the threshold p is at most $O(\log n)$ times larger than the expectation threshold q.

Proof. Let $\mathcal{S}$ be the family of minimal sets that defines monotone function f, and let $p = \epsilon$ be the threshold of f. Let $\mathbf{P}$ be a uniformly random set of size ϵn from $\{1, ..., n\}$, so $\mathbb{E}[f(\mathbf{P})] = \frac{1}{2}$.

By standard concentration bound, there must exist some number w close to p, and let $\mathbf{W}$ be a uniformly random set of size wn from $\{1, ..., n\}$, then $\mathbb{E}[f(\mathbf{W})] \leq \frac{3}{4}$.

Since $\mathbb{E}[f(\mathbf{W})] \leq \frac{3}{4}$, by definition of expectation, we have $\Pr[f(\mathbf{W}) = 1] \leq \frac{3}{4}$. Applying the main theorem, the "either-or" property, we know that

$$\Pr(f \leq g) \geq \frac{1}{4}$$

Let $q = \frac{w}{64 \log k}$, and let $\mathbf{Q}$ be a random set of size qn selected from $\{1, ..., n\}$. Now we use the other condition provided by the theorem, which is $\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|] < \frac{1}{8}$. By Markov's inequality,

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

$$\Pr_{\mathcal{G}}[\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|] \geq \frac{1}{2}] \leq \frac{\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|]}{1/2} < \frac{1/8}{1/2} = \frac{1}{4}$$

in short

$$\Pr_{\mathcal{G}}[\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|] \geq \frac{1}{2}] < \frac{1}{4}$$

which means

$$\Pr_{\mathcal{G}}[\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|] > \frac{1}{2}] > \frac{3}{4}$$

Therefore, there exists some choice of g (and associated $\mathcal{G}$), such that $g \geq f$, and $\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|] < \frac{1}{2}$. Note that q here is not the true expectation threshold, because we want $\mathbb{E}[|\mathcal{G}_{\mathbf{Q}}|] = \frac{1}{2}$ if q is the true expectation threshold. Let $\hat{q}$ be the true expectation threshold, so $\hat{q} > q$.

This proves the true expectation threshold $\hat{q}$ must be at least $q = \frac{w}{64 \log k}$, and the threshold p is very close to w ($p \approx w$). So $\hat{q} \geq \frac{p}{64 \log k} \Rightarrow \frac{p}{\hat{q}} \leq 64 \log k$, and $k \leq n$, so $\frac{p}{\hat{q}} = O(\log n)$. □

References

1. Anup Rao. Sunflowers: from soil to oil. Electronic Colloquium on Computational Complexity, Revision 3 of Report No. 5 (2022)
2. Anup Rao. The Sunflower Lemma and Monotone Thresholds. YouTube video: https://www.youtube.com/watch?v=Rw6FEZnuE5Y (2022)

3. Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. In *Prococedings of the 52nd Annual Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 624-630. ACM, 2020.
4. Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *Journal of London Mathematical Society*, 35:85-90, 1960.
5. Michel Deza and Peter Frankl. Every large set of equidistant $(0, +1, -1)$-vectors forms a sunflower. *Combinatorica*, 1(3):225–231, 1981.
6. Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. Thresholds versus fractional expectation-thresholds. *CoRR*, abs/1910.13433, 2019.
7. Tolson Bell, Suchakree Chueluecha, and Lutz Warnke. Notes on sunflowers. *Discret. Math.*, 344(7):112367, 2021.
8. Anup Rao. Coding for sunflowers. *Discrete Analysis*, 2020.
9. Jinyoung Park and Huy Tuan Pham. A proof of the Kahn-Kalai conjecture. *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), Denver, CO, USA, 2022*, pp. 636-639. (2022)
10. Jeff Kahn and Gil Kalai. Thresholds and expectation thresholds. *Comb. Probab. Comput.*, 16(3):495-502, 2007.