Computer Science 294 Lecture 25 Notes

Daniel Raban

April 18, 2023

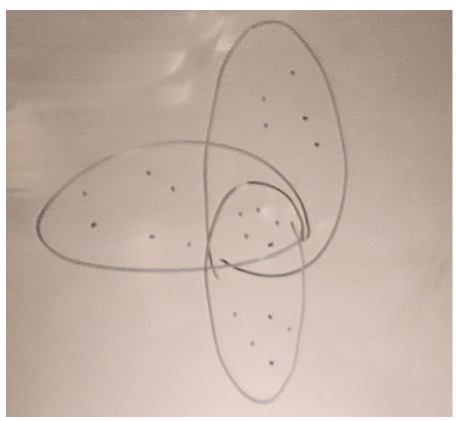
1 The Sunflower Lemma

1.1 Introduction to the sunflower lemma

Today, we will be discussing the work of Alweiss, Lovett, Wu, and Zhang (with some simplifications in presentation by Rao, Tao, BCW). The sunflower lemma has been long open and was one of Erdős' favorite unsolved problems.[1] The setting of the lemma comes from extremal combinatorics.

Let $U = [n]$, and let $S_1, S_2, \dots, S_m \subseteq U$ of size w.

Definition 1.1. An r**-sunflower** is a a set system where for any $i \neq j$, $S_i \cap S_j = \bigcap_k S_k$.



In other words, all the points sit in the core of the sunflower or in one of the petals. The question we are interested in is: How large does m need to be before the must exist an r-sunflower in $S_1, \dots, S_m$?

Theorem 1.1 (Erdős-Rado, 1960)**.** *If* $m > (r-1)^w \cdot w!$*, then there exists an* r*-sunflower.*

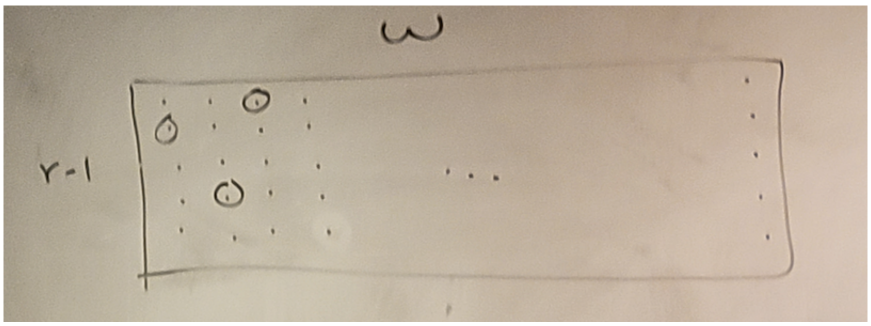
[1] He even offered a \$1000 prize for solving the problem.

Conjecture 1.1 (Sunflower conjecture)**.** *There exists a constant C_r such that if $m > (C_r)^w$, then there exists an r-sunflower.*

For many years, the only improvements were polynomial in w (many of which are from the last decade). Here is a lower bound.

Proposition 1.1. *There exists a family of $(r-1)^w$ sets with no r-sunflower.*

Proof. Consider a square lattice of points of size $(r-1) \times w$. Define the sets S_i by selecting exactly 1 element in every column.



□

Here is an application: Razborov showed that if you try to compute the clique problem with monotone circuits, the size of the circuit cannot be polynomial. The clique problem is monotone in the input, which is why we consider monotone circuits. Razborov used the sunflower lemma in his proof; the idea is to replace the sunflower by the core, which allows you to make the circuit smaller and smaller, yielding a contradiction. There are also applications for lower bounds for data structures.

Theorem 1.2 (ALWZ, 2020)**.** *If $m \geq (C \cdot r^3 \cdot \log w \cdot \log\log w)^w$, then there exists an r-sunflower.*

Quick followups by Rao, Tao, and BCW cleaned up the argument to give the bound

$$m \geq (Cr \log w)^w.$$

1.2 Proof sketch of ALWZ

Let's provide a proof of the result by Erdős and Rado.

Proof. We take a greedy approach. Try to find disjoint sets among $S_1, \ldots, S_m$. If we are successful, then we are done. Otherwise, there exists a set T (the union of up to $r-1$ sets from $S_1, \ldots, S_m$) with $|T| \le (r-1)w$ and T intersects each S_i. Therefore, there exists an element $i \in T$ that appears in $\ge \frac{m}{|T|}$ of the sets. We write

$$\frac{m}{|T|} > \frac{(r-1)^w w!}{(r-1)w} = (r-1)^{w-1}(w-1)!.$$

Look at all the sets containing i, and remove i from them. By induction, we find an $r-1$-sunflower among them. Then add i back in to get an r-sunflower. $\square$

The first case of these theorem is the typical case when picking random sets, while the second case is more structured. This is an example of the structure vs randomness paradigm, where we show that every object can be decomposed into a structured object and a random object. For example, a boolean function can be decomposed into the influential coordinates (structured part) and the remaining coordinates (random part).

Definition 1.2. A family $\mathcal{F} = \{S_1, \ldots, S_n\}$ is k-**spread** if for every subset $Z \subseteq U$ of size $\le w$,

$$|\{S_i : S_i \supseteq Z\}| \le k^{-|Z|} \cdot |\mathcal{F}|.$$

Notice that a k-spread family must be of size $\ge k^w$ because if we take $Z = S_1$, we get $1 \le k^{-w}|\mathcal{F}|$.

Here is the main lemma from the proof by ALWZ:

Lemma 1.1. *If $\mathcal{F} = \{S_1, \ldots, S_m\}$ has $> k^w$ sets and $\mathcal{F}$ is k-spread (for $k = k(r,w) = Cr\log w$), then $\mathcal{F}$ contains r disjoint sets.*

Assuming this main lemma, here is the proof of the theorem. It is similar to the proof of Erdős and Rado.

Proof of ALWZ theorem. If $\mathcal{F}$ is k-spread, we are done by the lemma. Otherwise, there exists a Z such that

$$|\{S_i : S_i \supseteq Z\}| > k^{-|Z|}k^w.$$

Look at all the sets containing Z, and remove Z from them. By induction, we find an r-sunflower among them. Then add Z back in to get an r-sunflower in the original family. $\square$

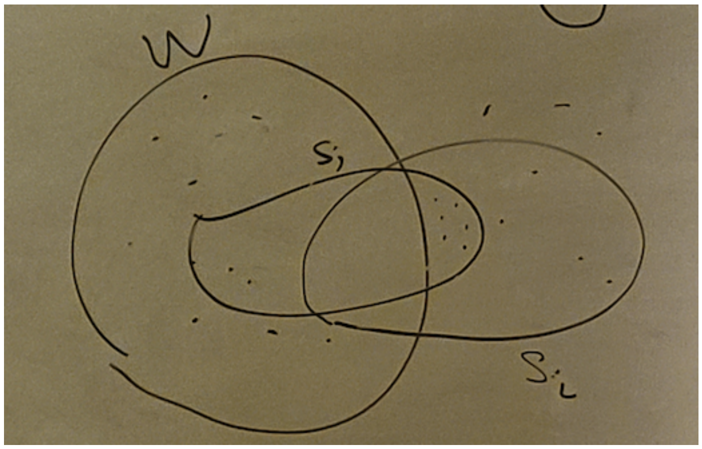
How do we prove the lemma? One naive approach is to partition the universe U into r sets and take an S_i from each of the pieces of the partition. It turns out that this approach works.

Proof idea of spread lemma. Partition the universe U to uniformly at random chosen $2r$ subsets. Show that each part (of size $\frac{n}{2r}$) contains a set from $\mathcal{F}$ with probability $\ge 1/2$. By the linearity of expectation, the average number of partition sets containing an S_i is $\ge r$. So there exists a choice of partition such that r of the parts contain a set S_i from $\mathcal{F}$. Thus, we have found r disjoint sets.

Here is how we show that each part of the partition contains a set from $\mathcal{F}$ with probability $\geq 1/2$. Pick $p = \frac{1}{2r}$. We take W to be a uniform random subset of U of size pn. We want to show that

$$\mathbb{P}_W(\exists j \text{ s.t. } S_j \subseteq W) \geq \frac{1}{2}.$$

Say that W is bad if for all j, $|S_j \setminus W| \geq w/2$. We want to show that $\mathbb{P}(W \text{ is bad})$ is small. Call a set S_i **compressible** with respect to W if there exists a set $S_j \in \mathcal{F}$ such that $S_j \subseteq W \cup S_i$ and $|S_j \setminus W| \leq w/2$.



If W is bad, then no set is compressible. We will show that most sets are compressible.

We use the following lemma, which is similar to the argument in the Håstad switching lemma.

Lemma 1.2. $\mathbb{P}_{W, S_i \in \mathcal{F}}(S_i$ *is compressible*$) = 1 - o(1)$

Proof. Call (W, S_i) bad if S_i is incompressible. Encode bad pairs (W, S_i) as follows:

- Encode $W \cup S_i$ ($\binom{n}{pn} + \binom{pn+1}{+} \cdots + \binom{n}{pn+w} \leq \binom{n}{pn} \cdot \frac{1}{p^w}$ options).
- Let S_j be the lexicographically first set in $\mathcal{F}$ that is contained in $W \cup S_i$. (By definition of "bad," $|S_j \setminus W| > w/2$.)
- Encode $S_i \cap S_j$ (2^w options).
- Encode the index of S_i among all sets containing $S_i \cap S_j$ ($|\mathcal{F}|k^{-|S_i \cap S_j|} \leq |\mathcal{F}|k^{-w/2}$ options).
- Encode $S_i \cap W$ (2^w options).

We get that

$$\frac{\#\text{Bad } (W, S_i)}{\#(W, S_i)} \leq \frac{\binom{n}{pn} \frac{1}{p^w} 4^w |\mathcal{F}| \frac{1}{k^{w/2}}}{\binom{n}{pn} |\mathcal{F}|} \leq \left(\frac{4}{p\sqrt{k}} \right)^w .$$

So picking appropriately large k gives the result. $\square$

We then compress $\log w$ times, each time decreasing the number of exceptional points by half. The argument does this by induction, and we omit the details.