Note on Sunflowers

Tolson Bell
Georgia Tech

Suchakree Chueluecha
Lehigh University

Lutz Warnke
Georgia Tech

Based on 2020 REU at Georgia Tech:
Sunflowers in Combinatorics

- Let $\mathcal{F}$ be a $k$-uniform family of subsets of $X$, i.e., $|S| = k$ and $S \subseteq X$ for all $S \in \mathcal{F}$.

- $\mathcal{F}$ is a sunflower with $p$ petals if $|\mathcal{F}| = p$ and there exists $Y \subseteq X$ with $Y = S_i \cap S_j$ for all distinct $S_i, S_j \in \mathcal{F}$.

- $Y$ is the core and $S_i \setminus Y$ are the petals.

- Note that $p$ disjoint sets forms a sunflower with $p$ petals and empty core.

Applications

Sunflowers have many uses in computer science:

- Fast algorithms for matrix multiplication
- Cryptography
- Pseudorandomness
- Lower bounds on circuitry
- Data structure efficiency
- Random approximations
Basic Results

Research Question

What is the smallest $r = r(p, k)$ such that every $k$-uniform family with at least $r^k$ sets must contain a sunflower with $p$ petals?

Erdős–Rado (1960)

(a) $r = pk$ is sufficient to guarantee a sunflower:
   every family with more than $(pk)^k > k!(p − 1)^k$ sets contains a sunflower

(b) $r > p − 1$ is necessary to guarantee a sunflower:
   there is a family of $(p − 1)^k$ sets without a sunflower

- Erdős conjectured $r = r(p)$ is sufficient (no $k$ dependency), offered $1000 reward
- Until 2018, best known upper bound on $r$ was still $k^{1−o(1)}$ with respect to $k$

“The sunflower problem] has fascinated me greatly – I really do not see why this question is so difficult.”

–Paul Erdős (1981)
Recent Exciting Developments

- Erdős conjectured $r = r(p)$ is sufficient (no $k$ dependency)
- Until 2018, best known upper bound on $r$ was still $k^{1-o(1)}$ with respect to $k$

**Alweiss–Lovett–Wu–Zhang (Breakthrough Aug 2019)**

$r = p^3 (\log k)^{1+o(1)}$ is sufficient to guarantee a sunflower

New papers built off their breakthrough ideas:

- Sep 2019: Rao used Shannon’s coding theorem for a cleaner proof and slightly better bound
- Oct 2019: Frankston–Kahn–Narayanan–Park improved a key lemma of ALWZ, enabling them to prove a conjecture of Talagrand regarding thresholds functions
- Jan 2020: Rao improved to $r = O(p \log(pk))$ by incorporating ideas from FKNP
- July 2020: Tao matched Rao’s bound with shorter proof using Shannon entropy
Rao (Jan 2020)

$r = O(p \log(pk))$ is sufficient to guarantee a sunflower

Bell–Chueluecha–Warnke (September 2020)

$r = O(p \log k)$ is sufficient to guarantee a sunflower

Further REU 2020 results:

- **Rao/Tao methods not needed for this result:**
  2019 Frankston–Kahn–Narayanan–Park result suffices with our proof variant

- **Main Technical Lemma is asymptotically sharp:**
  Bound cannot be improved further without change of proof strategy
Key Definition: \( \mathcal{F} \) is \( r \)-spread if \( |\mathcal{F}| \geq r^k \) and for every nonempty \( S \subseteq X \) the number of sets in \( \mathcal{F} \) which contain \( S \) is at most \( r^k - |S| \).

The Inductive Reduction

If every \( r \)-spread family contains \( p \) disjoint sets, then \( r^k \) sets guarantees a sunflower.

Proof. Induction on \( k \).

Question: How to find \( p \) disjoint sets in an \( r \)-spread family?

We now review the common proof framework of previous work.
Strategy: Reduction to Main Technical Lemma

**Question:** How to find $p$ disjoint sets?

### The Probabilistic Method

1. Consider a random partition of $X$ to $X_1, X_2, \ldots, X_p$ ($x \in X$ goes in random $X_i$)

2. Use **probabilistic method**
   - Show $\Pr(\nexists S_i \in \mathcal{F} \text{ such that } S_i \subseteq X_i) < \frac{1}{p}$
   - **Union bound:**
     $\Pr(\exists i \text{ where } X_i \text{ has no } S_i) < p \cdot \frac{1}{p} = 1$
   - There is partition where each $X_i$ has $S_i$
   - Then $S_1, \ldots, S_p$ are disjoint sets in $\mathcal{F}$

### Main Technical Lemma (Rao 2020)

Let $X_p$ be set where $\forall x \in X, x \in X_p$ w.p. $\frac{1}{p}$ independently. $\exists C > 1$ s.t. for $r \geq Cp \log(pk)$, $\Pr(\text{There does not exist } S_i \in \mathcal{F} \text{ such that } S_i \subseteq X_p) < \frac{1}{p}$
Main Technical Lemma (Rao 2020)

Let $X_a$ be set where $\forall x \in X, x \in X_a$ w.p. $\frac{1}{a}$ independently. $\exists C > 1$ s.t. for $r \geq Ca \log(bk)$, $\mathbb{P}(\text{There does not exist } S_i \in F \text{ such that } S_i \subseteq X_a) < \frac{1}{b}$

Bell–Chueluecha–Warnke (September 2020)

$r = O(p \log k)$ is sufficient to guarantee a sunflower

Proof Sketch (improve union bound via linearity of expectation):

- Partition $X_1, \cdots, X_{2p}$ instead of $X_1, \cdots, X_p$.
- To get $p$ disjoint sets, half of our sets need to contain a set in $F$.
- *Linearity of expectation*: if each $X_i$ has less than half chance of failure, there is some partition where at least half succeed.
- Apply main lemma with $a = 2p$, $b = 2$.
- $r = 2Cp \log(2k) = O(p \log k)$ suffices!
Summary

$\mathcal{F}$, a $k$-uniform family of subsets of $X$, is a **sunflower with $p$ petals** if $|\mathcal{F}| = p$ and there exists $Y \subseteq X$ with $Y = S_i \cap S_j$ for all distinct $S_i, S_j \in \mathcal{F}$.

**Research Question**

What is the smallest $r = r(p, k)$ such that every $k$-uniform family with at least $r^k$ sets must contain a sunflower with $p$ petals?

- **Erdős–Rado (1960):** $r = pk$ is sufficient and $r > p - 1$ is necessary
- **Erdős (1981):** Conjectured $r = r(p)$ sufficient
- **Alweiss–Lovett–Wu–Zhang (2019):** Breakthrough that $r = p^3(\log k)^{1+o(1)}$ suffices
- **Rao (2020):** By Shannon’s Coding Theorem, $r = O(p \log(pk))$ suffices

**Bell–Chueluecha–Warnke (2020)**

- $r = O(p \log k)$ suffices by minor variant of existing probabilistic arguments
- This bound cannot be improved without change of strategy
References

- Erdős (1981). *On the combinatorial problems which I would most like to see solved*. Combinatorica.
Strategy: Proving the Main Lemma

Main Technical Lemma

\[ P(\text{There does not exist } S_i \in \mathcal{F} \text{ such that } S_i \subseteq X_a) < \frac{1}{b} \]

Proof. Partition \( X_i \) to \( V_1, V_2 \) with equal size, so \( |V_1| = |V_2| = |X|/(2a) \).

- **Key Definition:** Given \( S \in \mathcal{F} \) and \( W \subseteq X \),
  \((S, W)\) is **m-good** if there exists \( S' \in \mathcal{F} \)
  such that \( S' \subseteq W \cup S \) and \( |S' \setminus W| \leq m \)

**Iteration:** \( P(\text{Less than half of sets in } \mathcal{F} \text{ are } m\text{-good with respect to } V_1) \leq \frac{1}{2b} \)

**Final Step:** \( P(\text{\( V_1 \cup V_2 \) does not contain a set in } \mathcal{F} \mid \text{successful iteration}) < \frac{1}{2b} \)
Key Definition: Given $S \in \mathcal{F}$ and $W \subseteq X$, 
$(S, W)$ is **m-good** if there exists $S' \in \mathcal{F}$ such that $S' \subseteq W \cup S$ and $|S' \setminus W| \leq m$

Iteration: $\mathbb{P}$(Less than half of sets in $\mathcal{F}$ are $m$-good with respect to $V_1) \leq \frac{1}{2b}$

- Partition $V_1$ to $W_1, W_2, \ldots, W_x$ with equal size
- Iteratively replace each good $(S, \bigcup_{1 \leq i \leq j} W_i)$ pair with the guaranteed $S'$
- Bound the number of bad pairs by a key counting lemma & Markov's inequality
- Moving from $S$ to $S'$ reduces the set sizes at each step as $\bigcup_{1 \leq i \leq j} W_i$ expands

Final Step: $\mathbb{P}(V_1 \cup V_2$ does not contain a set in $\mathcal{F} \mid$ successful iteration) $< \frac{1}{2b}$

- Construct an $m$-uniform $\mathcal{F}'$ from sets in $\mathcal{F}$ which are $m$-good with respect to $V_1$
- Apply Janson’s Inequality with $V_2$ and $\mathcal{F}'$ to bound $\mathbb{P}(\exists S \in \mathcal{F}' \text{ s.t. } S \subseteq V_2)$