# Approximate Sunflowers

Benjamin Rossman

January 26, 2019

**Abstract**

A $(p, \varepsilon)$-*approximate sunflower* is a family of sets $\mathcal{S}$ with the property that a $p$-random subset of the universe is $1-\varepsilon$ likely to contain a set of the form $A \setminus I$ where $A \in \mathcal{S}$ and $I$ is the intersection of all elements of $\mathcal{S}$. In this note, we give a proof of the Approximate Sunflower Theorem from [Ros14] (with a slightly sharper bound) showing that every $\ell$-uniform set system of size $\geq \ell!((t+\frac{1}{2})/p)^\ell$ contains a $(p, e^{-t})$-approximate sunflower. This result was originally applied to obtain monotone circuit lower bounds for the clique problem on Erdős-Rényi random graphs. The Approximate Sunflower Theorem has subsequently found applications in the sparsification of DNF formulas [GMR13] and was recently connected to questions on randomness extractors [LLZ18]. It has also been noted that improving the bound to $f(p,t)^\ell$ for any function $f(p,t)$ (which does not depend on $\ell$) would prove the notorious Sunflower Conjecture [LZ18, LSZ18].

Throughout this note, let $t > 0$ and $p \in (0,1)$ be arbitrary real numbers, let $\ell$ be a positive integer, and let $V$ be an arbitrary set. Let $\binom{V}{\ell}$ denote the set of $\ell$-element subsets of $V$, and let $\binom{V}{<\ell} = \bigcup_{j=0}^{\ell-1} \binom{V}{j}$.

We say that $\boldsymbol{X}$ is a *$p$-random subset* of $V$, written $\boldsymbol{X} \subseteq_p V$, if $\boldsymbol{X}$ contains each element of $V$ independently with probability $p$.

A *set system* over $V$ is a family $\mathcal{S}$ of subsets of $V$. For $B \subseteq V$, let $\mathcal{S}_B$ denote the set system

$$\mathcal{S}_B := \{A \setminus B : B \subseteq A \in \mathcal{S}\}.$$

Borrowing terminology from the literature on sunflowers, we define the *core* of $\mathcal{S}$ as the intersection $C = \bigcap_{A \in \mathcal{S}} A$ of all elements of $S$; elements of $\mathcal{S}_C$ are called *petals* of $\mathcal{S}$.

A set system $\mathcal{S}$ is a *sunflower* if its petals are pairwise disjoint (equivalently: if all pairs of distinct elements in $\mathcal{S}$ have the same intersection).

A set system $\mathcal{S}$ is *$\ell$-uniform* if $|A| = \ell$ for all $A \in \mathcal{S}$ (i.e., $\mathcal{S} \subseteq \binom{V}{\ell}$).

The Erdős-Rado Sunflower Theorem [ER60] establishes that every sufficiently large $\ell$-uniform set system contains a sunflower of size $k$.

**Theorem 1** (Sunflower Theorem). *Every $\ell$-uniform set system of size $> \ell!(k-1)^\ell$ contains a sunflower of size $k$.*

The following notion of *approximate sunflowers* was introduced in [Ros14]. (This was originally called *quasi-sunflower*. The much better name "approximate sunflowers" was suggested by Lovett and Zhang.)

**Definition 2.** A set system $\mathcal{S}$ over $V$ is a $(p, \varepsilon)$-*approximate sunflower* if a $p$-random subset of $V$ contains a petal of $\mathcal{S}$ with probability $> 1 - \varepsilon$.

Note that $\mathcal{S}$ contains a $(p, \varepsilon)$-approximate sunflower if, and only if, there exists a set $B \subseteq V$ such that $\mathbb{P}_{\boldsymbol{X} \subseteq_p V}[\ (\exists A \in \mathcal{S}_B)\ A \subseteq \boldsymbol{X}\ ] > 1 - \varepsilon$. In [Ros14] I showed that every $\ell$-uniform set system of size $\geq \ell!(2.5t/p)^\ell$ contains a $(p, e^{-t})$-approximate sunflower. This note proves a slightly stronger bound by a more careful analysis of the argument in [Ros14].

**Theorem 3** (Approximate Sunflower Theorem). *Every $\ell$-uniform set system of size $\geq \ell!((t+\frac{1}{2})/p)^\ell$ contains a $(p, e^{-t})$-approximate sunflower.*

The proof of Theorem 3 is by induction on $\ell$, similar to the proof of the Sunflower Theorem. A key tool in the argument is Janson's Inequality (Theorem 6). As we explain in Remark 10, the bound in Theorem 3 is essentially best possible by this method: obtaining a bound better than $\ell!(t/p)^\ell$ (or any bound of the form $f(p, t)^\ell$ without the $\ell!$ factor) appears to require a substantially different proof technique. An approach via randomness extractors was recently suggested by Li, Lovett and Zhang [LLZ18], who give an extractor-based proof of a quantitatively weaker version of Theorem 3 with the bound $2^{2\ell}((\ell + 1.5t)/p)^{c\ell}$ for a constant $c > 1$.

Before presenting the proof of Theorem 3, we remark on the relationship between sunflowers and approximate sunflowers.

**Proposition 4** (Sunflower $\Rightarrow$ Approximate Sunflower). *Every sunflower $\mathcal{S}$ of size $k$ is a $(p, e^{-kp^\ell})$-approximate sunflower where $\ell$ is the size of largest petal in $\mathcal{S}$.*

*Proof.* Let $\mathcal{S}$ be an $\ell$-uniform sunflower over $V$ with petals $A_1, \dots, A_k$. For $\boldsymbol{X} \subseteq_p V$, we have $\mathbb{P}[\ \boldsymbol{X}$ contains no petal of $\mathcal{S}\ ] \leq \prod_{i=1}^k \mathbb{P}[\ A_i \nsubseteq \boldsymbol{X}\ ] \leq (1 - p^\ell)^k \leq e^{-kp^\ell}$. $\square$

A cute relationship in the other direction was communicated to me by Jiapeng Zhang (an unpublished observation of Lovett, Solomon and Zhang [LSZ18]).

**Proposition 5** (Approximate Sunflower $\Rightarrow$ Sunflower). *Every $(\frac{1}{k}, \frac{1}{k})$-approximate sunflower contains a sunflower of size $k$.*

*Proof.* Let $\mathcal{S}$ be a $(\frac{1}{k}, \frac{1}{k})$-approximate sunflower. Let $\boldsymbol{X}_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} \boldsymbol{X}_k$ be a uniform random partition of $V$. Note that each $\boldsymbol{X}_i$ individually is a $\frac{1}{k}$-random subset of $V$. Let $\boldsymbol{I}_i \in \{0, 1\}$ be the indicator $\mathbb{1}[\ \boldsymbol{X}_i$ contains a petal of $\mathcal{S}\ ]$. Then $\mathbb{E}[\ \boldsymbol{I}_i\ ] > 1 - \frac{1}{k}$ for all $i \in \{1, \dots, k\}$. By linearity of expectations, $\mathbb{E}[\ \boldsymbol{I}_1 + \cdots + \boldsymbol{I}_k\ ] > k - 1$. Therefore, there exists a partition $X_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} X_k$ of $V$ such that each $\boldsymbol{X}_i$ contains a petal of $\mathcal{S}$. As this gives $k$ disjoint petals, we conclude that $\mathcal{S}$ contains a sunflower of size $k$. $\square$

In light of Proposition 5, if the bound $\ell!((t + \frac{1}{2})/p)^\ell$ in the Approximate Sunflower Theorem can be replaced by $f(p, t)^\ell$ for any function $f(p, t)$ (which does not depend on $\ell$), then the bound $\ell!(k - 1)^\ell$ of the Sunflower Theorem can be replaced by $f(\frac{1}{k}, \ln k)^\ell$. This would prove the notorious Sunflower Conjecture (see [ASU13, Juk11]). The hypothesis that such a function $f(p, t)$ exists was named the "Approximate Sunflower Conjecture" by Lovett and Zhang [LZ18].

The rest of this note contains the proof of Theorem 3. The key tool from probabilistic combinatorics is Janson's Inequality (a.k.a. the Extended Janson's Inequality).

**Theorem 6** (Janson's Inequality [Jan90])**.** *Let $\mathcal{S}$ be any set system over a set $V$ and let $\boldsymbol{X}$ be a random subset of $V$ such that events $\{v \in \boldsymbol{X}\}$ are independent over $v \in V$. Let*

$$\mu := \sum_{A \in \mathcal{S}} \mathbb{P}[\ A \subseteq \boldsymbol{X}\ ], \qquad \Delta := \sum_{(A_1,A_2)\in\mathcal{S}^2\,:\,A_1\cap A_2\neq\emptyset} \mathbb{P}[\ A_1 \cup A_2 \subseteq \boldsymbol{X}\ ].$$

*Then* $\mathbb{P}[\ (\forall A \in \mathcal{S})\ A \not\subseteq \boldsymbol{X}\ ] \leq \exp(-\mu^2/\Delta)$.

(In many statements of this inequality, the definition of $\Delta$ includes the condition $A_1 \neq A_2$ in the summation; in this case, one writes $\mu^2/(\mu+\Delta)$ instead of $\mu^2/\Delta$.) Rather than $\ell!((t+\frac{1}{2})/p)^\ell$, we shall prove a stronger version of Theorem 3 with the bound $c_\ell(t)/p^\ell$ for a certain sequence of polynomials $c_\ell(t)$.

**Definition 7.** Let $c_0(t), c_1(t), \ldots$ be the sequence of polynomials defined by

$$c_0(t) := 1, \qquad c_\ell(t) := t\sum_{j=0}^{\ell-1}\binom{\ell}{j}c_j(t) \text{ for } \ell \geq 1.$$

For $\ell \geq 1$, we have the explicit expression

$$c_\ell(t) = \sum_{k=1}^{\ell} t^k \sum_{0=j_0<j_1<\cdots<j_k=\ell}\ \prod_{i=1}^{k}\binom{j_i}{j_{i-1}}.$$

**Lemma 8.** *For all $t > 0$, we have*

$$\ell!t^\ell \leq c_\ell(t) \leq \ell!(t+\tfrac{1}{2})^\ell.$$

*Proof.* For the lower bound, we have

$$c_\ell(t) \geq t^\ell\binom{\ell}{\ell-1}\binom{\ell-1}{\ell-2}\cdots\binom{1}{0} = \ell!t^\ell.$$

For the upper bound, we have the following proof by induction that $c_\ell(t) \leq \ell!(1/\ln(\frac{1}{t}+1))^\ell$:

$$\begin{aligned} c_\ell(t) = t\sum_{j=0}^{\ell-1}\binom{\ell}{j}c_j(t) &\leq t\sum_{j=0}^{\ell-1}\binom{\ell}{j}j!\left(\frac{1}{\ln(\frac{1}{t}+1)}\right)^j \\ &= \ell!\left(\frac{1}{\ln(\frac{1}{t}+1)}\right)^\ell t\sum_{j=0}^{\ell-1}\frac{(\ln(\frac{1}{t}+1))^{\ell-j}}{(\ell-j)!} \\ &\leq \ell!\left(\frac{1}{\ln(\frac{1}{t}+1)}\right)^\ell t\left(-1+\sum_{k=0}^{\infty}\frac{(\ln(\frac{1}{t}+1))^k}{k!}\right) = \ell!\left(\frac{1}{\ln(\frac{1}{t}+1)}\right)^\ell. \end{aligned}$$

Finally, we use the fact that $1/\ln(\frac{1}{t}+1) < t+\frac{1}{2}$ for all $t > 0$. $\square$

In light of Lemma 8, Theorem 3 follows from the following theorem.

**Theorem 9.** *For every $\mathcal{S} \subseteq \binom{V}{\ell}$ with $|\mathcal{S}| \geq c_\ell(t)/p^\ell$, there exists $B \in \binom{V}{<\ell}$ such that*

$$\mathop{\mathbb{P}}_{\boldsymbol{X}\subseteq_p V}[\ (\forall A \in \mathcal{S}_B)\ A \not\subseteq \boldsymbol{X}\ ] < e^{-t}.$$

*Proof.* Induction on $\ell$. In the base case, let $\mathcal{S} \subseteq V$ with $|\mathcal{S}| \geq t/p$. We have

$$\mathop{\mathbb{P}}_{\boldsymbol{X} \subseteq_p V}[\ (\forall v \in \mathcal{S})\ v \notin \boldsymbol{X}\ ] = (1-p)^{|\mathcal{S}|} < e^{-p|\mathcal{S}|} = e^{-t}.$$

For the induction step, let $\ell \geq 2$ and let $S \subseteq \binom{V}{\ell}$ with $|\mathcal{S}| \geq c_\ell(t)/p^\ell$. We consider two cases.

Case 1: There exists $j \in \{1, \ldots, \ell-1\}$ and $B \in \binom{V}{\ell}$ such that $|\mathcal{S}_B| \geq c_{\ell-j}(t)/p^{\ell-j}$. By the induction hypothesis, there exists $C \in \binom{V}{<\ell-j}$ such that

$$\mathop{\mathbb{P}}_{\boldsymbol{X} \subseteq_p V}[\ (\forall A \in (\mathcal{S}_B)_C)\ A \nsubseteq \boldsymbol{X}\ ] < e^{-t}.$$

Since $(\mathcal{S}_B)_C = \mathcal{S}_{B \cup C}$, we are done.

Case 2: For all $j \in \{1, \ldots, \ell-1\}$ and $B \in \binom{V}{\ell}$, we have $|\mathcal{S}_B| < c_{\ell-j}(t)/p^{\ell-j}$. As in Theorem 6, let

$$\mu := \sum_{A \in \mathcal{S}} \mathop{\mathbb{P}}_{\boldsymbol{X} \subseteq_p V}[\ A \subseteq \boldsymbol{X}\ ], \qquad \Delta := \sum_{(A_1, A_2) \in \mathcal{S}^2 \,:\, A_1 \cap A_2 \neq \emptyset} \mathop{\mathbb{P}}_{\boldsymbol{X} \subseteq_p V}[\ A_1 \cup A_2 \subseteq \boldsymbol{X}\ ].$$

It suffices to show that $\mu^2/\Delta > t$.

First, we have the lower bound

$$\mu = p^\ell |\mathcal{S}| \geq c_\ell(t).$$

We next upper bound $\Delta$:

$$\begin{aligned}
\Delta &= \mu + \sum_{j=1}^{\ell-1} p^{2\ell-j} \sum_{B \in \binom{V}{j}} |\{(A_1, A_2) \in \mathcal{S}^2 : A_1 \cap A_2 = B\}| \\
&\leq \mu + \sum_{j=1}^{\ell-1} p^{2\ell-j} \sum_{B \in \binom{V}{j}} |\mathcal{S}_B|^2 \\
&< \mu + p^\ell \sum_{j=1}^{\ell-1} c_{\ell-j}(t) \sum_{B \in \binom{V}{j}} |\mathcal{S}_B| \\
&= \mu + p^\ell \sum_{j=1}^{\ell-1} c_{\ell-j}(t) \sum_{A \in \mathcal{S}} \binom{\ell}{j} \\
&= \mu + p^\ell |\mathcal{S}| \sum_{j=1}^{\ell-1} \binom{\ell}{j} c_{\ell-j}(t) \\
&= \mu \sum_{j=0}^{\ell-1} \binom{\ell}{j} c_j(t).
\end{aligned}$$

Therefore,

$$\frac{\mu^2}{\Delta} > \frac{\mu}{\sum_{j=0}^{\ell-1} \binom{\ell}{j} c_j(t)} \geq \frac{c_\ell(t)}{\sum_{j=0}^{\ell-1} \binom{\ell}{j} c_j(t)} = t.$$

Janson's Inequality now yields the desired bound $\mathbb{P}_{\boldsymbol{X} \subseteq_p V}[\ (\forall A \in \mathcal{S})\ A \nsubseteq \boldsymbol{X}\ ] < e^{-t}$. $\square$

**Remark 10.** This bound on $\Delta$ is essentially tight. For $i \in \{1, \ldots, \ell\}$ and $C \in \binom{V}{i}$, instead of upper bounding the number of pairs $(A_1, A_2) \in \mathcal{S}^2$ with $A_1 \cap A_2 = C$ by $|\mathcal{S}_C|^2$ (in our bound on $\Delta$), we can instead use inclusion-exclusion to get an equality:

$$|\{(A_1, A_2) \in \mathcal{S}^2 : A_1 \cap A_2 = C\}| = \sum_{j=i}^{\ell} (-1)^{j-i} \sum_{B \in \binom{V}{j} : C \subseteq B} |\mathcal{S}_B|^2.$$

This gives the following exact expression for $\Delta$:

$$\begin{aligned}
\Delta &= \sum_{i=1}^{\ell} p^{2\ell-i} \sum_{C \in \binom{V}{i}} |\{(A_1, A_2) \in \mathcal{S}^2 : A_1 \cap A_2 = C\}| \\
&= \sum_{i=1}^{\ell} p^{2\ell-i} \sum_{C \in \binom{V}{i}} \sum_{j=i}^{\ell} (-1)^{j-i} \sum_{B \in \binom{V}{j} : C \subseteq B} |\mathcal{S}_B|^2 \\
&= \sum_{j=1}^{\ell} \sum_{i=1}^{j} p^{2\ell-i} (-1)^{j-i} \binom{j}{i} \sum_{B \in \binom{V}{j}} |\mathcal{S}_B|^2 \\
&= \sum_{j=1}^{\ell} p^{2\ell-j} \sum_{i=1}^{j} (-p)^{j-i} \binom{j}{i} \sum_{B \in \binom{V}{j}} |\mathcal{S}_B|^2 \\
&= \sum_{j=1}^{\ell} p^{2\ell-j} \Big( (1-p)^j - (-p)^j \Big) \sum_{B \in \binom{V}{j}} |\mathcal{S}_B|^2.
\end{aligned}$$

For small $p$, the value of $(1-p)^j - (-p)^j$ is very close to 1. Even in the case $p = \frac{1}{2}$, we get no significant improvement; in this case we have

$$\Delta = \sum_{j \in \{1,3,5,\ldots,2\lfloor \ell/2 \rfloor - 1\}} (1/2)^{2\ell-j} \sum_{B \in \binom{V}{j}} |\mathcal{S}_B|^2.$$

This allows us to replace $c_\ell(t)$ in Theorem 3 with the polynomial

$$d_\ell(t) = \sum_{k=1}^{\ell} t^k \sum_{\substack{0 = j_0 < j_1 < \cdots < j_k = \ell : \\ j_i - j_{i-1} \text{ is odd for all } i \in \{1,\ldots,k\}}} \prod_{i=1}^{k} \binom{j_i}{j_{i-1}}.$$

However, $d_\ell(t)$ is still lower bounded by $\ell! t^\ell$ for $t > 0$. For this reason, it appear that any improvement to Theorem 3 beyond $\ell! t^\ell$ will require a substantially different proof technique.

## References


[ASU13] Noga Alon, Amir Shpilka, and Christopher Umans. On sunflowers and matrix multiplication. *computational complexity*, 22(2):219–243, 2013.

[ER60] Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, 1(1):85–90, 1960.

[GMR13] Parikshit Gopalan, Raghu Meka, and Omer Reingold. DNF sparsification and a faster deterministic counting algorithm. *Computational Complexity*, 22(2):275–310, 2013.

[Jan90] Svante Janson. Poisson approximation for large deviations. *Random Structures and Algorithms*, 1(2):221–229, 1990.

[Juk11] Stasys Jukna. *Extremal combinatorics: with applications in computer science.* Springer, 2011.

[LLZ18] Xin Li, Shachar Lovett, and Jiapeng Zhang. Sunflowers and quasi-sunflowers from randomness extractors. In *APPROX-RANDOM*, volume 116 of *LIPIcs*, pages 51:1–13, 2018.

[LSZ18] Shachar Lovett, Noam Solomon, and Jiapeng Zhang. Unpublished work, 2018.

[LZ18] Shachar Lovett and Jiapeng Zhang. DNF sparsification beyond sunflowers. *ECCC preprint TR18-190*, 2018.

[Ros14] Benjamin Rossman. The monotone complexity of $k$-clique on random graphs. *SIAM Journal on Computing*, 43(1):256–279, 2014.