THEORY OF COMPUTING, Volume 18 (2), 2022, pp. 1–18
www.theoryofcomputing.org

SPECIAL ISSUE: RANDOM 2018

Sunflowers and Robust Sunflowers from Randomness Extractors

Xin Li* Shachar Lovett† Jiapeng Zhang

Received September 19, 2018; Revised May 15, 2021; Published January 31, 2022

Abstract. The Erdős–Rado Sunflower Theorem (J. London Math. Soc. 1960) is a fundamental result in combinatorics, and the corresponding Sunflower Conjecture is a central open problem. Motivated by applications in complexity theory, Rossman (FOCS 2010) extended the result to robust sunflowers, where similar conjectures emerge about the optimal parameters for which it is expected to hold.

We exhibit a surprising connection between the existence of sunflowers and robust sunflowers in sufficiently large families of sets and the problem of constructing certain randomness extractors. This allows us to rederive the known results in a systematic manner. Our techniques have also motivated significant subsequent progress on the Sunflower Conjecture by Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang (STOC'20 and Annals of Math. 2021).

A preliminary version of this paper appeared in the Proceedings of the 22nd International Conference on Randomization and Computation (RANDOM'18) [19].

*Research supported by NSF award CCF-1617713

†Research supported by NSF CCF-1614023

ACM Classification: F.1.3, G.2.1

AMS Classification: 05D10, 68Q17, 68Q25

Key words and phrases: CNF-DNF formulas, extractors, combinatorics, sunflowers

DOI: 10.4086/toc.2022.v018a002

1 Introduction

Let $\mathcal{F}$ be a family of sets from some universe X. A common theme and extensively studied phenomenon in combinatorics is the following: if the cardinality of $\mathcal{F}$ (when $\mathcal{F}$ is finite) or the density of $\mathcal{F}$ (when $\mathcal{F}$ is infinite) is large enough, then some nice patterns will occur in $\mathcal{F}$. Well-known examples of this kind include (1) Szemerédi's theorem [28], which asserts that all subsets of the natural numbers of positive density contain arbitrarily long arithmetic progressions; (2) Ramsey's Theorem [12], which asserts that if one colors the edges of a large enough complete graph with a finite number of colors, then there must exist a monochromatic clique of a certain size; and (3) the Erdős–Rado Sunflower Theorem [11], which asserts that a large enough family of sets of bounded size must contain a large sunflower.[1]

The study of these problems has resulted in many important tools (e.g., Szemerédi's Regularity Lemma [29] and the probabilistic method), which have found applications not only in combinatorics, or mathematics more broadly, but also in theoretical computer science (TCS). Conversely, ideas from TCS have influenced related research in combinatorics quite often. For example, the first two problems we mentioned above, Szemerédi's Theorem and Ramsey's Theorem, are intimately connected to the area of pseudorandomness in TCS. Indeed, by constructing a certain sparse pseudorandom subset of natural numbers and proving an appropriate Szemerédi-type theorem with respect to that subset, a celebrated result of Green and Tao [15] shows that prime numbers contain arbitrarily long arithmetic progressions. As for Ramsey's Theorem, a recent line of work on randomness extractors [3–5, 7, 17, 18] gives strongly explicit[2] constructions of Ramsey graphs that get close to the probabilistic bound [10].

In this paper we study the Sunflower Theorem and its related variants. We show that again there is an intimate connection to randomness extractors. In fact, using techniques from randomness extractors, we build a general proof framework that can unify the Sunflower Theorem and its variant known as the Robust Sunflower Lemma [27].

Sunflowers. A *family* is a set of sets. (Repeated sets are not permitted in counting the size of a family.) A family $\mathcal{F}$ is *w-uniform* if every element of $\mathcal{F}$ has size w. An *r-sunflower* is defined to be a family of r sets such that the intersection of any two sets in the family is the same set (which can be the empty set). Choose any w-uniform family $\mathcal{F}$, the main question of interest is, how large $\mathcal{F}$ needs to be in order to ensure that there is an r-sunflower in $\mathcal{F}$. Erdős and Rado proved the following theorem.

Theorem 1.1 (Sunflower Theorm, Erdős–Rado [11])**.** *Let $\mathcal{F}$ be a w-uniform family. If $|\mathcal{F}| > w!(r-1)^w$ then $\mathcal{F}$ contains an r-sunflower.*

They also conjectured that the bound on $|\mathcal{F}|$ can be replaced by c_r^w where c_r is a constant that only depends on r for every $r > 0$. This conjecture is one of the most well-known open problems in combinatorics, which remains open today despite extensive research.

[1] A sunflower is a family of sets whose pairwise intersection is constant, which we will formally define shortly.

[2] A class of graphs is *strongly explicit* if the vertices of each graph $G = (V, E)$ in the class are encoded by $(0,1)$-strings of length $O(\log|V|)$ and a polynomial-time algorithm $P(|V|, i, j)$ decides adjacency of any pair (i, j) of vertices of G in polynomial (i. e., polylog$(|V|)$) time.

The Sunflower Theorem has applications in TCS, such as in the proof of strong lower bounds for monotone circuits [26]. In addition, Alon, Shpilka and Umans [1] related the Sunflower Conjecture and its variants to possible approaches to fast matrix multiplication. Recently, following the breakthrough proof of the Cap-set Conjecture [8, 9], a weaker version of the Sunflower Conjecture was also proved using the same method [24] (concretely, the Erdős–Szemerédi Sunflower Conjecture for $w = 3$). However, the general conjecture remains open.

Robust sunflowers. Motivated by the applications of sunflowers to proving monotone circuit lower bounds, *robust sunflowers* were introduced by Rossman [27] under the name *"quasi-sunflowers"* to prove monotone circuit lower bounds for the k-clique problem on random graphs. We denote by $\mathcal{P}(X)$ the family of all subsets of a finite set X.

Definition 1.2 (Robust sunflower [27])**.** Let X be a finite set and $\mathcal{S} \subseteq \mathcal{P}(X)$ a family of size $|\mathcal{S}| \geq 2$. Denote $Y = \bigcap_{T \in \mathcal{S}} T$. Following Rossman [27], for $p, \gamma \in [0,1]$, we say that $\mathcal{S}$ is a (p,γ)-*robust sunflower* if for a random set $W \subseteq X$, with each element of X present in W independently with probability p, it holds that

$$\Pr[\exists T \in \mathcal{S}, (T \setminus Y) \subseteq W] \geq 1 - \gamma.$$

As Rossman explains, the robust sunflower concept is a relaxation of the sunflower concept where petals may overlap in a limited way. Indeed, if $\mathcal{F}$ is a w-uniform sunflower then for every $p \in [0,1]$, $\mathcal{F}$ is a (p,γ)-robust sunflower with $\gamma = \exp(-|\mathcal{F}|p^w)$ [27, Remark 13].

In the same paper, Rossman also proved the following lemma, which says there is always a robust sunflower in a large family.

Lemma 1.3 (Robust Sunflower Lemma, Rossman [27])**.** *Let $\mathcal{F}$ be a w-uniform family. If*

$$|\mathcal{F}| \geq w! \cdot (1.71 \log(1/\gamma)/p)^w,$$

then[3] $\mathcal{F}$ contains a (p,γ)-robust sunflower.

Besides the original application, Rossman's Robust Sunflower Lemma was also used by Gopalan, Meka and Reingold [14] to study the problem of DNF sparsification. Given a DNF formula f on n variables, there are two natural ways to measure the complexity of f: the number of clauses (also called size) $s(f)$, and the maximum width of a clause $w(f)$. It is easy to show that any DNF of small size can be approximated well by another DNF of small width, by truncating clauses of large width. Gopalan et al. [14] used Rossman's Robust Sunflower Lemma to show the reverse direction, that any DNF with small width can also be approximated well by another DNF with small size. In particular, they showed that any width-w DNF formula can be ε-approximated by another DNF formula with size at most $(w \log(1/\varepsilon))^{O(w)}$. This kind of sparsification has applications in constructing pseudorandom generators and approximately counting the number of satisfying assignments for DNF formulas.

[3] Throughout this paper, "log" refers to base-2 logarithms.

Similarly to the Sunflower Conjecture, one can also ask whether the bound on $|\mathcal{F}|$ in the Robust Sunflower Lemma can be improved. This question was further studied by Alweiss, Lovett, Wu and Zhang [2] and Rao [25]. We discuss further details in Section 1.4.

1.1 Our contribution

We provide a general framework to prove both the Sunflower Theorem and the Robust Sunflower Lemma. In fact, we reduce both of these problems to the construction of a certain type of randomness extractors. To state our results, we first formally define the notions that are going to be used in our extractors.

A random variable X, taking values in a finite set Σ, defines a probability measure on Σ, namely, for every subset $\Delta \subseteq \Sigma$ we have the measure $\mu(\Delta) = \Pr(X \in \Delta)$. Risking some confusion, we shall also refer to X as a "*distribution over* Σ," reducing the meaning of X to the probability measure it defines on Σ. We refer to Σ as the *sample space*. The *support* of X, denoted $\mathsf{Supp}(X)$, is the smallest subset $\Delta \subseteq \Sigma$ such that $\Pr(X \in \Delta) = 1$.

Definition 1.4. Let X be a distribution over a finite sample space Σ. The *min-entropy* of X is defined as

$$\mathcal{H}_\infty(X) = \min_{x \in \Sigma} \left\{ \log \left(\frac{1}{\Pr[X = x]} \right) \right\}.$$

A distribution over $\{0,1\}^n$ is said to be an (n,k) source if the distribution has min-entropy at least k.

Definition 1.5 (Block min-entropy source)**.** Let $X_1, \ldots, X_m$ be random variables with sample space $\{0,1\}^n$. Consider the random variable $X = (X_1, \ldots, X_m)$ (with sample space $\{0,1\}^{nm}$). The distribution X is an (m,n,k) *block min-entropy* source if for every non-empty subset $S \subseteq [m]$, the joint distribution of $(X_i : i \in S)$ has min-entropy at least $k|S|$.

We note that the definition of block min-entropy sources was initiated in [13] as a tool to prove lifting theorems in communication complexity.

Definition 1.6 (Block min-entropy extractor)**.** A function $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^d$ is a (k,ε,d,s) *block min-entropy* extractor if for any $m,n \in \mathbb{N}$ and any (m,n,k) block min-entropy source $X = (X_1, \ldots, X_m)$, it holds that

$$(E(X_1,R_1), \ldots, E(X_m,R_m)) \approx_\varepsilon U_{dm}.$$

Here, each $R_i \in \{0,1\}^s$ is an independent uniform random string, U_{dm} is the uniform distribution over (dm)-bit strings, and $\approx_\varepsilon$ means ε-close in statistical distance. If in addition it holds that

$$(E(X_1,R_1), R_1, \ldots, E(X_m,R_m), R_m) \approx_\varepsilon U_{(d+s)m},$$

then we say that the function E is a *strong* (k,ε,d,s) block min-entropy extractor.

We also define a weaker object called a disperser.

Definition 1.7 (Block min-entropy disperser)**.** A function

$$E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^d$$

is a (k,ε,d,s) *block min-entropy disperser* if for any $m,n \in \mathbb{N}$ and any (m,n,k) block min-entropy source $X = (X_1,\ldots,X_m)$, it holds that

$$|\mathsf{Supp}(E(X_1,R_1),\ldots,E(X_m,R_m))| \geq (1-\varepsilon)2^{dm}.$$

Here, each $R_i \in \{0,1\}^s$ is an independent uniform random string. If in addition there exists at least one way to fix $R_1 = r_1,\ldots,R_m = r_m$ such that

$$|\mathsf{Supp}(E(X_1,r_1),\ldots,E(X_m,r_m))| \geq (1-\varepsilon)2^{dm},$$

then we say that the function E is a *strong* (k,ε,d,s) block min-entropy disperser.

In this paper, we make connections between the block min-entropy disperser and sunflower and robust sunflower structures. Formally, we prove the following result.

Theorem 1.8. *Suppose that there exists a strong $(k,0,d,s)$ block min-entropy disperser, $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^d$ for any (w,n,k) block min-entropy source. Then the following holds.*

Let $\mathcal{F}$ be a w-uniform family. Assume that $|\mathcal{F}| \geq 2^{(k+2)w}$. Then:

(i) $\mathcal{F}$ contains a 2^d-sunflower.

(ii) $\mathcal{F}$ contains a $\left(p, w(1-p)^{2^d}\right)$-robust sunflower.

Observe that the seed length s of the extractor does not play a part in the conclusion of Theorem 1.8. We then show that we can construct strong zero-error block min-entropy extractors. As a corollary, we are able to construct strong block min-entropy dispersers. Specifically, we have the following result.

Theorem 1.9. *There is a constant $c > 1$ such that for any $m,n,k \in \mathbb{N}$ with $k \geq c\log m$, we have:*

- *There is an explicit[4] strong (k,ε,d,s) block min-entropy extractor $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^d$ for (m,n,k) block min-entropy sources, where $s = n$, $d = k/c$ and $\varepsilon = 2^{-\Omega(k)}$.*
- *There is an explicit strong $(k,0,d,s)$ block min-entropy disperser $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^d$ for (m,n,k) block min-entropy sources, where $s = n$, $d = k/c$.*

Combined with Theorem 1.8, this gives the Sunflower Theorem and the Robust Sunflower Lemma. In Theorem 1.9, we give explicit constructions. We also note that the existence of block min-entropy dispersers is also enough to prove sunflower lemmas.

[4] A class of finite functions is *explicit* if a polynomial-time algorithm evaluates each function in the class on each of its inputs. The class of bipartite graphs corresponding to an explicit class of functions is strongly explicit in the sense of Footnote 2.

Corollary 1.10 (Sunflower Theorem, this paper)**.** *There is a constant c such that for any w-uniform family and any $r > 1$, if $|\mathcal{F}| \geq (wr)^{cw}$, then $\mathcal{F}$ contains an r-sunflower.*

The reader will note that this result is weaker than Theorem 1.1, the original result by Erdős and Rado. The point we are making here is that our result is comparable, and it is obtained using a different approach that may lead to improved bounds. Indeed, a variant of our method did lead to improved bounds [2, 25].

Corollary 1.11 (Robust Sunflower Lemma, this paper)**.** *There is a constant c such that for any w-uniform family $\mathcal{F}$, if $|\mathcal{F}| \geq \left(\frac{w+\log(1/\gamma)}{p}\right)^{cw}$, then $\mathcal{F}$ contains a (p,γ)-robust sunflower.*

1.2 Overview of the techniques

Our reduction from sunflower/robust sunflower problems to block min-entropy dispersers is as follows. Suppose the family $\mathcal{F} \subseteq \mathcal{P}(X)$ for some set X, where each set in $\mathcal{F}$ has size w. We first show that without loss of generality we can assume $\mathcal{F}$ is a w-partite family.

Definition 1.12 (w-partite family)**.** Let X be a finite set and let $\mathcal{F} = \{U_i\}_{i\in I}$ be a family of subsets of X. We say that $\mathcal{F}$ is a w-partite family if

- $\mathcal{F}$ is w-uniform;
- there is a partition $X_1, \ldots, X_w$ of X such that for every $U \in \mathcal{F}$, we have $|X_j \cap U| = 1$ for each $j \in [w]$.

Consider the uniform distribution over a w-partite family $\mathcal{F}$. There are two possible cases:

- **Case 1:** there is a subset S which is a subset of many elements of $\mathcal{F}$, specifically,

$$\left|\{U \in \mathcal{F} : S \subseteq U\}\right| \geq \frac{|\mathcal{F}|}{r^{|S|}},$$

  where r is a parameter to be determined.

- **Case 2:** every set S does not appear in too many sets of $\mathcal{F}$.

In Case 1, S is already like a core in a sunflower or robust sunflower, thus we can apply induction on the subfamily $\mathcal{F}_S := \{U \setminus S : S \subseteq U \in \mathcal{F}\}$. In Case 2, the condition basically implies that the distribution is relatively flat, which equivalently translates into a block min-entropy source as we defined above. One can naturally imagine that the worst-case situation here is that the distribution is actually the uniform distribution over $X_1 \times \cdots \times X_w$, and we show that indeed this is the case by using our zero-error block min-entropy disperser (Theorem 1.9). It is then easy to see that in the worst case, the empty set is a robust sunflower, or one can choose a sunflower with size 2^d (the support size in the output of the disperser) whose core is the empty set.

1.3 The role of extractors in our reduction

One can view the block min-entropy extractor/disperser used in our reduction as a gadget, which reduces the sunflower/robust sunflower problem in the general case to the much easier case of a uniform distribution (or full support) on $X_1 \times \cdots \times X_w$. This is similar to the role of extractors in recent work that showed lifting theorems from query complexity to communication complexity [13], and linear programming lower bounds for constraint satisfaction problems [16].

In fact, the extractors used in these articles are essentially the same as the extractors used in the present paper (although here we need to show that the extractor/disperser is strong, while in [13] and [16] this is not necessary), and the barriers to further improvement are also similar. Specifically, in all such constructions one needs the min-entropy $k \geq c \log m$ for some constant $c > 1$, where m is equal to the size of the sets (i.e., w) in our applications. It is interesting to ask whether $k = \Omega(\log m)$ is necessary. Any improvement of such extractors leads to improvements of both sunflower theorems and lifting theorems. Subsequent to the conference version of this paper [19], Meka [23] gave a counterexample showing that such extractors do not exist.

1.4 Subsequent work

In this section we report work done after the publication of the conference version of this paper [19].

Connection to DNF sparsification: The connection of sunflowers and DNF sparsification was first discovered by Gopalan, Meka and Reingold [14]. Building on the Robust Sunflower Lemma, Gopalan et al. [14] proved any width-w DNF (with arbitrary size) can be ε-approximated by a DNF of size at most $(w \cdot \log(1/\epsilon))^{O(w)}$. Similarly to the Sunflower Conjecture, Gopalan et al. also believed the DNF compression bound can be improved to $(\log(1/\epsilon))^{O(w)}$. This conjecture was confirmed by subsequent work by Lovett, Wu and Zhang [21, 22].

In another piece of subsequent work Lovett, Solomon and Zhang [20] also proved that the reverse direction is also true. That is, an improved upper bound on DNF sparsification implies an improved sunflower bound.

Progress on the Sunflower Conjecture. In this paper, we show that any w-uniform family $\mathcal{F}$ of size $|\mathcal{F}| \geq w^{cw}$ for some constant $c > 1$ contains a 3-sunflower. Furthermore, we show that any family $\mathcal{F}$ with the following spread condition must contain 3 pairwise disjoint sets.

Definition 1.13. Given a family $\mathcal{F}$ and $r > 0$, we say that $\mathcal{F}$ is *r-spread*, if for any set S,

$$\left|\{U \in \mathcal{F} : S \subseteq U\}\right| \leq \frac{\left|\mathcal{F}\right|}{r^{|S|}}.$$

The following corollary follows directly from the proof of Theorem 1.8.

Corollary 1.14. *There is a constant $c > 0$, such that for any w-partite family $\mathcal{F}$, if $\mathcal{F}$ is w^c-spread then it contains w pairwise disjoint sets.*

This w^c-spread condition actually comes from the requirement of our disperser that $k \geq c \log w$. As discussed above, it is interesting to ask whether this is necessary. In an article subsequent to the conference version of this paper [19], Alweiss, Lovett, Wu and Zhang [2] improved the w^c bound on the spread to $(\log w)^{O(1)}$ and Rao [25] further improved it to $O(\log w)$.

Lemma 1.15 ([2, 25]). *There is a constant $c > 0$, such that for any w-partite family $\mathcal{F}$, if $\mathcal{F}$ is $(cr \cdot \log(wr))$-spread then it contains r pairwise disjoint sets.*

This lemma finally leads to an improved sunflower lemma.

Corollary 1.16 ([2, 25]). *There is a constant $c > 0$ such that, for each w-uniform family $\mathcal{F}$, if $|\mathcal{F}| > (cr \cdot \log(wr))^w$ then $\mathcal{F}$ contains an r-sunflower.*

Discussion of robust sunflowers. In this paper, we also study robust sunflower structures. In particular, we have the following corollary. Below, we use the notation $O_{p,\gamma}(\cdot)$ to hide the specific dependency on the parameters p, γ, which is of less interest to us here.

Corollary 1.17. *There is a constant $c > 0$ such that, for any w-partite family $\mathcal{F}$, if $\mathcal{F}$ is r-spread where $r = (O_{p,\gamma}(w))^c$, then $\mathcal{F}$ is a (p, γ)-robust sunflower with empty kernel.*

The subsequent work of Alweiss, Lovett, Wu and Zhang [2], and its improvement by Rao [25], gives an improved robust sunflower lemma.

Lemma 1.18 ([2], [25]). *There is a constant $c > 0$ such that, for any w-partite family $\mathcal{F}$, if $\mathcal{F}$ is $(c \log(w/\gamma)/p)$-spread then $\mathcal{F}$ is a (p, γ)-robust sunflower with empty kernel.*

Furthermore, the $\log w$ term is tight by the following example. Fix $p = \gamma = 1/2$ for convenience. Let $X_1, \ldots, X_w$ be w disjoint sets each of size $c \log w$ for some large enough $c > 0$. Define the family $\mathcal{F} := X_1 \times \cdots \times X_w$. By this we mean, with some abuse of notation, the complete w-partite hypergraph on $\{X_1, \ldots, X_w\}$. Then $\mathcal{F}$ does not contain a (p, γ)-robust sunflower. This example also sets a barrier to proving the Sunflower Conjecture via robust sunflowers.

Discussion of the block min-entropy extractor and the disperser construction. In this paper, we construct explicit strong block min-entropy extractors and dispersers. However, in our constructions, we require the min-entropy k to be as large as $\Omega(\log m)$. In the conference version of this paper we asked whether the min-entropy condition could be improved to be $o(\log m)$. The following counterexample due to Raghu Meka [23] shows that this is impossible. Therefore it sets a barrier to improving (robust) sunflowers via extractors.

Theorem 1.19. *Let m, n be parameters and set $k := \min(\log(2^n - 1), \log m)$. Then for any function $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}$, there is an (m, n, k) block min-entropy source $X = (X_1, \ldots, X_m)$ such that, for every $r_1, \ldots r_m \in \{0,1\}^s$, the support of*

$$(E(X_1, r_1), \ldots, E(X_m, r_m))$$

is not full.

Proof. Let $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}$ be the given function. Fix an arbitrary string $x^* \in \{0,1\}^n$. Define the distribution $X = (X_1, \ldots, X_m)$ as follows. First, sample an index $i^* \in [m]$ uniformly and set $X_{i^*} = x^*$. For every $i \neq i^*$, sample $X_i \in \{0,1\}^n \setminus \{x^*\}$ uniformly and independently. Observe that X is a (m, n, k) block min-entropy source for $k = \min(\log(2^n - 1), \log m)$.

We will show that for every $r_1, \ldots, r_m \in \{0,1\}^s$, the random variable $(E(X_1, r_1), \ldots, E(X_m, r_m))$ does not have full support on $\{0,1\}^m$. This is because if we let $z = (E(x^*, r_1), E(x^*, r_2), \ldots, E(x^*, r_m))$, then the output of $(E(X_1, r_1), \ldots, E(X_m, r_m))$ always agrees with z in at least one position. Thus the string $z \oplus 1$ is not in the support of $(E(X_1, r_1), \ldots, E(X_m, r_m))$.

□

Acknowledgements

We thank Raghu Meka and Ben Rossman for helpful discussions. We also thank anonymous reviewers for helpful suggestions on an earlier version of this paper.

2 Preliminaries

We first review some basic definitions in probability.

Definition 2.1. Let D be a distribution over a finite sample space Σ. Its entropy is

$$\mathcal{H}(D) = \sum_{x \in \Sigma} \Pr[D = x] \cdot \log\left(\frac{1}{\Pr[D = x]}\right).$$

Its min-entropy is

$$\mathcal{H}_\infty(D) = \min_{x \in \Sigma} \left\{\log\left(\frac{1}{\Pr[D = x]}\right)\right\}.$$

Its max-entropy is

$$\mathcal{H}_0(D) = \log |\mathsf{Supp}(D)|.$$

Definition 2.2 (Statistical distance)**.** Let D_0 and D_1 be distributions over a finite sample space Σ. The statistical distance between D_0 and D_1 is defined as

$$\mathrm{dist}(D_0, D_1) = \frac{1}{2} \sum_{x \in \Sigma} \big|\Pr[D_0 = x] - \Pr[D_1 = x]\big|.$$

3 A construction of a block min-entropy extractor

We use the following well-known extractor based on the inner product function [6]. We denote by $\mathbb{F}_q$ the finite field on q elements. When $q = 2^\ell$ we identify $\mathbb{F}_q$ with $\{0,1\}^\ell$ and $\mathbb{F}_q^t$ with $\{0,1\}^{t\ell}$.

Theorem 3.1. *Let $t, \ell \geq 1$ and take $q = 2^\ell, n = t\ell$. Let X, Y be independent sources on $\mathbb{F}_q^t \cong \{0,1\}^n$ with min-entropies k_1, k_2, respectively. Let* IP *be the inner product function over the field $\mathbb{F}_q$. Then:*

$$\operatorname{dist}((\mathrm{IP}(X,Y),X),(U_\ell,X)) \leq \varepsilon \quad \text{and} \quad \operatorname{dist}((\mathrm{IP}(X,Y),Y),(U_\ell,Y)) \leq \varepsilon$$

where $\varepsilon = 2^{-(k_1+k_2-n-\ell)/2}$.

Now we can construct a block min-entropy extractor as follows. Given parameters n, k, choose a field $\mathbb{F}_q$ such that $q = 2^\ell$ with $\ell = \alpha k$ for some constant $0 < \alpha < 1$ to be determined later. Without loss of generality we assume that $n = t\ell$ for some integer t. We view $X \in \{0,1\}^n$ as a vector in $\mathbb{F}_q^t$ and choose a uniform independent seed $R \in \{0,1\}^n \cong \mathbb{F}_q^t$.

A block min-entropy extractor

1. Given parameters m, n, k let q, t be as described above.
2. Sample $(X_1, \ldots, X_m) \in (\mathbb{F}_q^t)^m$ from the block min-entropy distribution.
3. Sample $(R_1, \ldots, R_m) \in (\mathbb{F}_q^t)^m$ uniformly and independently.
4. Output $Z := (\mathrm{IP}(X_1, R_1), \ldots, \mathrm{IP}(X_m, R_m))$.

We are now ready to prove the following theorem.

Theorem 3.2 (Theorem 1.9 restated)**.** *Let $X = (X_1, \ldots, X_m)$ be an (m, n, k) block min-entropy source. Let $Z \in \{0,1\}^{\ell m}$ be the output of the above block min-entropy extractor applied to X. There exists a constant $c > 1$ such that if $k \geq c \log m$, then the following holds for error $\varepsilon = 2^{-\Omega(k)}$:*

- *With probability $1 - \varepsilon$ over the fixing of the seed $(R_1, \ldots, R_m)$,*

$$\left|\Pr[Z = z] - 2^{-\ell m}\right| \leq \varepsilon \cdot 2^{-\ell m} \qquad \forall z \in \{0,1\}^{\ell m}.$$

  In particular, in such cases $\mathcal{H}_0(Z) = \ell m$

- $\operatorname{dist}((Z, R_1, \ldots, R_m), (U, R_1, \ldots, R_m)) \leq 2\varepsilon$.

Proof. We have a joint distribution $(X_1, \ldots, X_m)$ that has block min-entropy k. The output of the local extractor applied to $(X_1, \ldots, X_m)$, using m independent uniform seeds $(R_1, \ldots, R_m)$, is a distribution $(Z_1, \ldots, Z_m)$ over $\{0,1\}^{\ell m} = \mathbb{F}_q^m$ where $Z_i = \mathrm{IP}(X_i, R_i)$ for each i.

For any fixing of the seed $(R_1 = r_1, \ldots, R_m = r_m)$, the distribution $(Z_1, \ldots, Z_m)$ is a deterministic function of $(X_1, \ldots, X_m)$, and we will view this distribution as a function $\mathcal{D} : \{0,1\}^{\ell m} \to [0,1]$ where the image of each input is its associated probability in the distribution. We now write this function in its Fourier basis:

$$\mathcal{D}(z) = \sum_{S \subseteq [\ell m]} \hat{\mathcal{D}}(S) \chi_S(z),$$

where $z = (z_1, \ldots, z_m) \in \{0,1\}^{\ell m}$, $\chi_S(z) = (-1)^{\sum_{i \in S} z(i)} \in \{+1, -1\}$, and

$$\hat{\mathcal{D}}(S) = 2^{-\ell m} \cdot \sum_z \mathcal{D}(z)\chi_S(z) = 2^{-\ell m} \cdot \mathbb{E}_{z \sim \mathcal{D}}[\chi_S(z)].$$

Here we use $z(i)$ to stand for the i-th *bit* of the string z. This is to distinguish between the notation z_i, which refers to the i-th *block* of the string z, that contains ℓ bits.

Note that $\hat{\mathcal{D}}(\emptyset) = 2^{-\ell m}$ since $\mathcal{D}$ is a probability distribution. Thus we have that $\forall z \in \{0,1\}^{\ell m}$,

$$\left|\mathcal{D}(z) - 2^{-\ell m}\right| = \left|\sum_{S \subseteq [\ell m], S \neq \emptyset} \hat{\mathcal{D}}(S)\chi_S(z)\right| \leq \sum_{S \subseteq [\ell m], S \neq \emptyset} \left|\hat{\mathcal{D}}(S)\right|. \tag{3.1}$$

Note that for any $S \subseteq [\ell m]$, $\chi_S(Z)$ corresponds to the parity of a subset of the bits in Z. For each $Z_j, j \in [m]$, this parity may or may not involve any bits in Z_j. We will be interested in the number of indices j such that $\chi_S(Z)$ involves at least one bit from Z_j, and we call this number $\Delta(S)$. Note that $\Delta(\emptyset) = 0$ and $1 \leq \Delta(S) \leq m$ for any $S \neq \emptyset$.

We now have the following lemma.

Lemma 3.3. *If $\Delta(S) = h$, then with probability $1 - 2^{-h(1-\alpha)k/4}$ over the fixing of the seed $(R_1 = r_1, \ldots, R_m = r_m)$, we have that $|\hat{\mathcal{D}}(S)| \leq 2 \cdot 2^{-\ell m} 2^{-h(1-\alpha)k/4}$.*

Proof. Without loss of generality assume that the Z_j from which $\chi_S(Z)$ involves at least one bit are $(Z_1, \ldots, Z_h)$. Note that for any $Z_i \in \{0,1\}^\ell = \mathbb{F}_q$, any parity of the bits of Z_i corresponds exactly to the first bit of $a \cdot Z_i$ viewed as a vector in $\{0,1\}^\ell$, for some $a \in \mathbb{F}_q$ and where the operation $\cdot$ is multiplication in the field $\mathbb{F}_q$. Moreover this correspondence is a bijection in the sense that different parities correspond to different elements $a \in \mathbb{F}_q$. The special case of parity over the empty set corresponds to the case of $a = 0$. Thus, $\sum_{i \in S} Z(i)$ corresponds to the first bit of $\sum_{j \in [h]} a_j Z_j$ viewed as a vector in $\{0,1\}^\ell$, for some non-zero $\{a_j \in \mathbb{F}_q : j \in [h]\}$. Note that

$$\sum_{j \in [h]} a_j Z_j = \sum_{j \in [h]} a_j \mathrm{IP}(X_j, R_j) = \sum_{j \in [h]} \mathrm{IP}(a_j X_j, R_j) = \mathrm{IP}((a_1 X_1, \ldots, a_h X_h), (R_1, \ldots, R_h)). \tag{3.2}$$

Since each $a_j \neq 0$ the transformation from $(x_1, \ldots, x_h)$ to $(a_1 x_1, \ldots, a_h x_h)$ is a bijection. Thus we know the distribution $(a_1 X_1, \ldots, a_h X_h)$ has min-entropy kh, while $(R_1, \ldots, R_h)$ has min-entropy nh. Thus by Theorem 3.1 applied over the field $\mathbb{F}_{2^{\ell h}}$ we have that

$$\mathrm{dist}\left((\sum_{j \in [h]} a_j Z_j, R_1, \ldots, R_m), (U_{\ell h}, R_1, \ldots, R_m)\right) \leq 2^{\frac{-(h(k-\ell))}{2}} = 2^{\frac{-h(1-\alpha)k}{2}}. \tag{3.3}$$

In particular, as $\chi_S(Z)$ is the first bit of $\sum_{j \in [h]} a_j Z_j$, we have

$$\mathrm{dist}\,(\chi_S(Z), R_1, \ldots, R_m), (U_1, R_1, \ldots, R_m)) \leq 2^{\frac{-h(1-\alpha)k}{2}}. \tag{3.4}$$

By Markov's inequality this means that with probability $1 - 2^{-h(1-\alpha)k/4}$ over the fixing of the seed $R = (R_1, \ldots, R_m)$, we have $|\hat{\mathcal{D}}(S)| = |2^{-\ell m} \cdot \mathbb{E}_{z \sim \mathcal{D}}[\chi_S(z)]| \leq 2 \cdot 2^{-\ell m} 2^{-h(1-\alpha)k/4}$. □

Next, note that the number of S with $\Delta(S) = h$ is $\binom{m}{h}(2^\ell - 1)^h \le 2^{(\ell+\log m)h}$. Recall that $\ell = \alpha k$ and $k \ge c \log m$. We can choose the constants α, c such that $2^{\ell+\log m} 2^{-(1-\alpha)k/4} \le 2^{-k/8}$. Now we have as long as $k \ge 8$,

$$\sum_{h=1}^{m} \binom{m}{h} (2^\ell - 1)^h 2^{\frac{-h(1-\alpha)k}{4}} \le \sum_{h=1}^{m} 2^{-\frac{hk}{8}} \le 2^{-\frac{k}{8}+1}. \tag{3.5}$$

Set $\varepsilon = 2^{-k/8+2} = 2^{-\Omega(k)}$.

By the union bound we have that with probability at least $1-\varepsilon$ over the fixing of the seed $(R_1 = r_1, \ldots, R_m = r_m)$, for every $S \neq \emptyset$ with $\Delta(S) = h$, $|\hat{\mathcal{D}}(S)| \le 2 \cdot 2^{-\ell m} 2^{\frac{-h(1-\alpha)k}{4}}$. Thus for any such seed we have that

$$\left|\mathcal{D}(z) - 2^{-\ell m}\right| \le \sum_{S \subseteq [\ell m], S \neq \emptyset} \left|\hat{\mathcal{D}}(S)\right| \le \varepsilon \cdot 2^{-\ell m}. \tag{3.6}$$

This concludes the proof of the first part of Theorem 1.9. For the second part, notice that conditioned on the fixing of any seed $R_1, \ldots, R_m$, with probability $1-\varepsilon$ the statistical distance is at most ε, and otherwise it is trivially bounded by 1. So overall the statistical distance between $(Z, R_1, \ldots, R_m)$ and $(U_{\ell m}, R_1, \ldots, R_m)$ is at most 2ε. □

4 Compressing set systems by the block min-entropy extractor

In this section, we focus on the set systems that satisfy the spread condition, and show a compression operator for such set systems. Our compression is based on the block min-entropy extractor. We first show that it suffices to consider w-partite families (see Definition 1.12).

Lemma 4.1. *Let $\mathcal{F}$ be a w-uniform family. Then $\mathcal{F}$ has a w-partite subfamily $\mathcal{F}'$ of size $|\mathcal{F}'| \ge |\mathcal{F}|/2^{2w}$.*

Proof. Let $U \in \mathcal{F}$ be a set, and let $X_1, \ldots, X_w$ be a random partition of X. Then

$$\Pr_{X_1,\ldots,X_w}\left[\forall j \in [w], |U \cap X_j| = 1\right] = \frac{w!}{w^w}.$$

Then, by averaging, there is a partition $(X_1, \ldots, X_w)$ such that

$$\left|\left\{U \in \mathcal{F} : \forall j \in [w], |U \cap X_j| = 1\right\}\right| \ge |\mathcal{F}| \cdot \frac{w!}{w^w}$$

The claim then follows since $\frac{w!}{w^w} \ge 2^{-2w}$. □

Now we can focus on w-partite families. Given a finite set X, we denote by X_p the distribution over subsets $W \subset X$, where each $x \in X$ appears in W independently with probability p.

Lemma 4.2. *Let $u \ge w$. Let c be the constant from Theorem 1.9. Then for every w-partite family which is u^c-spread (recall Definition 1.13), it holds that*

$$\Pr_{W \sim X_p}[\exists U \in \mathcal{F}, U \subseteq W] \ge 1 - w(1-p)^u.$$

To prove this lemma, we first define a "worst case" instance, and then show that all other instances behave better than this case. Let $X_1^*, \ldots, X_w^*$ be w disjoint sets each of size u. Define the family $\mathcal{U}^*$ as

$$\mathcal{U}^* = \left\{ \{x_1, \ldots, x_w\} : \forall j \in [w], x_j \in X_j^* \right\}.$$

Claim 4.3. *Let $\mathcal{U}^*$ as defined above. Then*

$$\Pr_{W \sim X_p}[\exists U \in \mathcal{U}^*, U \subseteq W] \geq 1 - w(1-p)^u.$$

Proof. By the definition of $\mathcal{U}^*$, we have that

$$\begin{aligned} \Pr_W[\forall U \in \mathcal{U}^*, U \not\subseteq W] &= \Pr_W[\exists j \in [w], X_j \cap W = \emptyset] \\ &\leq \sum_{j \in [w]} \Pr_W[X_j \cap W = \emptyset] \\ &= w(1-p)^u. \end{aligned}$$

□

Let X, Y be finite sets, $h : X \to Y$ a map. Given a set $U \subset X$ define $h(U) = \{h(x) : x \in U\} \subset Y$. Given a family $\mathcal{F} \subseteq \mathcal{P}(X)$ define $h(\mathcal{F}) \subseteq \mathcal{P}(Y)$ as

$$h(\mathcal{F}) = \left\{ h(U) : U \in \mathcal{F} \text{ and } h \text{ is injective on } U \right\}.$$

Lemma 4.4. *Let X and Y be sets, $h : X \to Y$ a map, $\mathcal{F} \subset \mathcal{P}(X)$. Then*

$$\Pr_{W_Y \sim Y_p}[\exists U \in h(\mathcal{F}), U \subseteq W_Y] \leq \Pr_{W_X \sim X_p}[\exists U \in \mathcal{F}, U \subseteq W_X].$$

Proof. Without loss of generality, we can assume the map h is surjective, because elements $y \in Y \setminus h(X)$ do not affect the events. If $|Y| = |X|$ then h is a bijection and hence $\mathcal{F}$ and $h(\mathcal{F})$ are the same, up to renaming the elements. So, assume $|Y| < |X|$. It suffices to prove the lemma for the case that $|Y| = |X| - 1$, as the general case follows from applying this case iteratively (namely, decompose h as a sequence of maps, each reduces the domain size by one).

So, assume $|Y| = |X| - 1$. In this case, there is a unique pair $x_1, x_2 \in X$ such that $h(x_1) = h(x_2) = y$. We may assume without loss of generality (by renaming the elements of Y) that h is the identity map on $X' = X \setminus \{x_1, x_2\}$. This allows us to jointly sample (W_X, W_Y) as follows. Sample $W' \sim X'_p, W'_X \sim \{x_1, x_2\}_p, W'_Y \sim \{y\}_p$ and set $W_X = W' \cup W'_X, W_Y = W' \cup W'_Y$. We will show that for every fixed $W' = w'$,

$$\Pr_{W_Y \sim Y_p}[\exists U \in h(\mathcal{F}), U \subseteq W_Y \mid W' = w'] \leq \Pr_{W_X \sim X_p}[\exists U \in \mathcal{F}, U \subseteq W_X \mid W' = w']. \tag{4.1}$$

The lemma then follows by averaging over W'.

To that end, fix W'. Let $\mathcal{F}' = \{U \setminus X' : U \in \mathcal{F}, (U \cap X') \subset W'\}$. Note that $\mathcal{F}' \subseteq \mathcal{P}(\{x_1, x_2\})$. Similarly, define $\mathcal{F}'' = \{U \setminus X' : U \in h(\mathcal{F}), (U \cap X') \subset W'\}$. Note that $\mathcal{F}'' \subseteq \mathcal{P}(\{y\})$. Equation (4.1) is equivalent to

$$\Pr_{W'_Y \sim \{y\}_p}[\exists U \in \mathcal{F}'', U \subseteq W'_Y] \leq \Pr_{W'_X \sim \{x_1, x_2\}_p}[\exists U \in \mathcal{F}', U \subseteq W'_X]. \tag{4.2}$$

We verify Equation (4.2) by a case analysis.

(i) If $\mathcal{F}''$ is empty then the LHS of Equation (4.2) is 0, while the RHS is non-negative.

(ii) If $\emptyset \in \mathcal{F}''$ then $\emptyset \in \mathcal{F}'$. In this case, both the LHS and RHS of Equation (4.2) equal 1.

(iii) If $\mathcal{F}'' = \{\{y\}\}$ then either $\{x_1\} \in \mathcal{F}'$ or $\{x_2\} \in \mathcal{F}'$ (or possibly both). In either case, the LHS of Equation (4.2) equals p, while the RHS is at least p.

□

We now prove Lemma 4.2. Let $\mathcal{F}$ be a family that satisfies the assumptions. We will show there is a function h such that $h(\mathcal{F}) = \mathcal{U}^*$. The extractor from Theorem 1.9, with an appropriate choice of seed, provides such a function h.

Proof of Lemma 4.2. Let $\mathcal{F}$ be a w-partite family that satisfies the spread condition. We first define the function h. Since $\mathcal{F}$ is a w-partite family, there exists a partition of X to $X_1, \ldots, X_w$ such that for each $U \in \mathcal{F}$ and $j \in [w]$, $|X_j \cap U| = 1$.

Define the sample space as $X_1 \times \cdots \times X_w$. With a slight abuse of notation, we identify $\mathcal{F} \subseteq \mathcal{P}(X_1 \times \cdots \times X_w)$, and let D be a uniform distribution over $\mathcal{F}$. Since $\mathcal{F}$ is u^c-spread, the distribution D is a $(w, \log X, k)$ block min-entropy source with $k = c \log u \geq c \log w$. Then by Theorem 1.9, there exists seeds $r_1, \ldots, r_w$ such that $(\mathrm{IP}(D_1, r_1), \ldots, \mathrm{IP}(D_w, r_w))$ has full support, where $D = (D_1, \ldots, D_w)$. Note that the output of $\mathrm{IP}(\cdot, \cdot)$ is in $\{0,1\}^{k/c} \cong [u]$. We can now define h as follows:

$$h(x) = (\mathrm{IP}(x, r_j), j) \qquad \forall x \in X_j.$$

Note that by definition, h is injective on any $U \in X_1 \times \cdots \times X_w$. We identify elements of $\mathcal{U}^*$ with $\{(a_1, 1), \ldots, (a_w, w)\}$ with $a_i \in [u]$. Thus $h(\mathcal{F}) = \mathcal{U}^*$. The lemma now follows from Lemma 4.4 and Claim 4.3. □

We will also need the following lemma.

Lemma 4.5. *Let $u \geq w$. Let c be the constant from Theorem 1.9. Then for every w-partite family $\mathcal{F}$ which is u^c-spread (recall Definition 1.13), it holds that $\mathcal{F}$ contains u pairwise disjoint sets.*

Proof. The proof is very similar to the proof of Lemma 4.2. There is a map h for which $h(\mathcal{F}) = \mathcal{U}^*$. Note that $\mathcal{U}^*$ contains u pairwise disjoint sets, $U'_1, \ldots, U'_u$. By definition, $U'_i = h(U_i)$. But then also $U_1, \ldots, U_u$ must be pairwise disjoint. □

4.1 Sunflowers and robust sunflowers from compression

Now we can prove Theorem 1.8.

Theorem 4.6 (Theorem 1.8 restated)**.** *Suppose that there exists a strong $(k, 0, d, s)$-block min-entropy disperser, $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^d$ for any (w, n, k)-block min-entropy source. Then the following holds.*

Let $\mathcal{F}$ be a w-uniform family. Assume that $|\mathcal{F}| \geq 2^{(k+2)w}$. Then:

(i) $\mathcal{F}$ contains a 2^d-sunflower.

(ii) $\mathcal{F}$ *contains a* $\left(p, w(1-p)^{2^d}\right)$*-robust sunflower.*

Proof. By Lemma 4.1, there is a w-partite subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size $|\mathcal{F}'| \geq 2^{kw}$. There are two possible cases.
Case 1: There is a subset $S \subseteq X$ such that

$$|\{U \in \mathcal{F}' : S \subseteq U\}| \geq |\mathcal{F}'| \cdot 2^{-k|S|}.$$

Define the family $\mathcal{F}'_S := \{U \setminus S : S \subseteq U \in \mathcal{F}'\}$. Notice that

- $\mathcal{F}'_S$ is $(w - |S|)$-partite;
- $|\mathcal{F}'_S| \geq |\mathcal{F}'| \cdot 2^{-k|S|} \geq 2^{k(w-|S|)}$.

By induction both (i) and (ii) hold.
Case 2: For all $S \subseteq X$,

$$|\{U \in \mathcal{F}' : S \subseteq U\}| \leq |\mathcal{F}'| \cdot 2^{-k|S|}$$

Notice that this is the spread condition for Lemma 4.2 and Lemma 4.5. Their conclusions are precisely (i) and (ii). □

References

[1] NOGA ALON, AMIR SHPILKA, AND CHRISTOPHER UMANS: On sunflowers and matrix multiplication. *Comput. Complexity*, 22(2):219–243, 2013. Preliminary version in CCC'12. [doi:10.1007/s00037-013-0060-1, ECCC:TR11-067] 3

[2] RYAN ALWEISS, SHACHAR LOVETT, KEWEN WU, AND JIAPENG ZHANG: Improved bounds for the sunflower lemma. *Ann. Math.*, 194(3):795–815, 2021. Preliminary version in STOC'20. [doi:10.4007/annals.2021.194.3.5, ECCC:TR19-110, arXiv:1908.08483] 4, 6, 8

[3] BOAZ BARAK, GUY KINDLER, RONEN SHALTIEL, BENNY SUDAKOV, AND AVI WIGDERSON: Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. *J. ACM*, 57(4):20:1–52, 2010. Preliminary version in STOC'05. [doi:10.1145/1734213.1734214, ECCC:TR10-037] 2

[4] AVRAHAM BEN-AROYA, DEAN DORON, AND AMNON TA-SHMA: An efficient reduction from two-source to non-malleable extractors: Achieving near-logarithmic min-entropy. In *Proc. 49th STOC*, pp. 1185–1194. ACM Press, 2017. [doi:10.1145/3055399.3055423, ECCC:TR16-088] 2

[5] ESHAN CHATTOPADHYAY AND DAVID ZUCKERMAN: Explicit two-source extractors and resilient functions. In *Proc. 48th STOC*, pp. 670–683. ACM Press, 2016. [doi:10.1145/2897518.2897528, ECCC:TR15-119] 2

[6] BENNY CHOR AND ODED GOLDREICH: Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988. Preliminary version in FOCS'85. [doi:10.1137/0217015] 9

[7] GIL COHEN: Towards optimal two-source extractors and Ramsey graphs. In *Proc. 49th STOC*, pp. 1157–1170. ACM Press, 2017. [doi:10.1145/3055399.3055429, ECCC:TR16-114] 2

[8] ERNIE CROOT, VSEVOLOD F. LEV, AND PÉTER PÁL PACH: Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small. *Ann. Math.*, 185(1):331–337, 2017. [doi:10.4007/annals.2017.185.1.7] 3

[9] JORDAN S. ELLENBERG AND DION GIJSWIJT: On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression. *Ann. Math.*, 185(1):339–343, 2017. [doi:10.4007/annals.2017.185.1.8] 3

[10] PAUL ERDŐS: Some remarks on the theory of graphs. *Bull. AMS*, 53(4):292–294, 1947. project Euclid. 2

[11] PAUL ERDŐS AND RICHARD RADO: Intersection theorems for systems of sets. *J. London Math. Soc.*, 35(1):85–90, 1960. [doi:10.1112/jlms/s1-35.1.85] 2

[12] PAUL ERDŐS AND GEORGE SZEKERES: A combinatorial problem in geometry. *Compositio Math.*, 2:463–470, 1935. NUMDAM. 2

[13] MIKA GÖÖS, SHACHAR LOVETT, RAGHU MEKA, THOMAS WATSON, AND DAVID ZUCKERMAN: Rectangles are nonnegative juntas. *SIAM J. Comput.*, 45(5):1835–1869, 2016. Preliminary version in STOC'15. [doi:10.1137/15M103145X, ECCC:TR14-147] 4, 7

[14] PARIKSHIT GOPALAN, RAGHU MEKA, AND OMER REINGOLD: DNF sparsification and a faster deterministic counting algorithm. *Comput. Complexity*, 22(2):275–310, 2013. Preliminary version in CCC'12. [doi:10.1007/s00037-013-0068-6, arXiv:1205.3534] 3, 7

[15] BEN GREEN AND TERENCE TAO: The primes contain arbitrarily long arithmetic progressions. *Ann. Math.*, 167(2):481–547, 2008. [doi:10.4007/annals.2008.167.481] 2

[16] PRAVESH K. KOTHARI, RAGHU MEKA, AND PRASAD RAGHAVENDRA: Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of CSPs. In *Proc. 49th STOC*, pp. 590–603. ACM Press, 2017. [doi:10.1145/3055399.3055438, arXiv:1610.02704] 7

[17] XIN LI: Three source extractors for polylogarithmic min-entropy. In *Proc. 56th FOCS*, pp. 863–882. IEEE Comp. Soc., 2015. [doi:10.1109/FOCS.2015.58, ECCC:TR15-034, arXiv:1503.02286] 2

[18] XIN LI: Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proc. 49th STOC*, pp. 1144–1156. ACM Press, 2017. [doi:10.1145/3055399.3055486, ECCC:TR16-115, arXiv:1608.00127] 2

[19] XIN LI, SHACHAR LOVETT, AND JIAPENG ZHANG: Sunflowers and quasi-sunflowers from randomness extractors. In *Proc. 22nd Internat. Conf. on Randomization and Computation (RANDOM'18)*, pp. 51:1–13. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.APPROX-RANDOM.2018.51] 1, 7, 8

[20] SHACHAR LOVETT, NOAM SOLOMON, AND JIAPENG ZHANG: From DNF compression to sunflower theorems via regularity. In *Proc. 34th Comput. Complexity Conf. (CCC'19)*, volume 137, pp. 5:1–14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.CCC.2019.5, ECCC:TR19-028, arXiv:1903.00580] 7

[21] SHACHAR LOVETT, KEWEN WU, AND JIAPENG ZHANG: Decision list compression by mild random restrictions. In *Proc. 52nd STOC*, pp. 247–254. ACM Press, 2020. [doi:10.1145/3357713.3384241, ECCC:TR19-137, arXiv:1909.10658] 7

[22] SHACHAR LOVETT AND JIAPENG ZHANG: DNF sparsification beyond sunflowers. In *Proc. 51st STOC*, pp. 454–460. ACM Press, 2019. [doi:10.1145/3313276.3316323, ECCC:TR18-190] 7

[23] RAGHU MEKA: Personal communication, 2018. 7, 8

[24] ERIC NASLUND AND WILL SAWIN: Upper bounds for sunflower-free sets. In *Forum of Mathematics, Sigma*, volume 5. Cambridge Univ. Press, 2017. [doi:10.1017/fms.2017.12] 3

[25] ANUP RAO: Coding for sunflowers. *Discrete Analysis*, pp. 2:1–8, 2020. [doi:10.19086/da.11887, arXiv:1909.04774] 4, 6, 8

[26] ALEXANDER A. RAZBOROV: Lower bounds for the monotone complexity of some Boolean functions. *Dokl. Math.*, 31(4):354–357, 1985. Link at Math-Net.ru. 3

[27] BENJAMIN ROSSMAN: The monotone complexity of k-clique on random graphs. *SIAM J. Comput.*, 43(1):256–279, 2014. Preliminary version in FOCS'10. [doi:10.1137/110839059] 2, 3

[28] ENDRE SZEMERÉDI: On sets of integers containing no k elements in arithmetic progression. *Acta Arithm.*, 27(1):199–245, 1975. DML PL. 2

[29] ENDRE SZEMERÉDI: Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Proc. conf. Univ. Orsay 1976)*, volume 260, pp. 399–401. C.N.R.S., 1978. 2

AUTHORS

Xin Li
Associate professor
Johns Hopkins University
Baltimore, MD, USA
lixints@cs.jhu.edu
https://www.cs.jhu.edu/~lixints/

Shachar Lovett
Associate professor
University of California San Diego
La Jolla, CA, USA
slovett@cs.ucsd.edu
http://cseweb.ucsd.edu/~slovett

Jiapeng Zhang
Assistant professor
University of Southern California
Los Angeles, CA, USA
jiapengz@usc.edu
https://sites.google.com/site/jiapeng0708/home

ABOUT THE AUTHORS

XIN LI is an associate professor in the Computer Science Department at Johns Hopkins University. He received his Ph. D. in 2011 from the University of Texas at Austin under the supervision of David Zuckerman. After his Ph. D., Xin was a Simons Postdoctoral Fellow at the University of Washington. His research interests include the use of randomness in computation, complexity theory, coding theory, and cryptography. A significant part of his work has been on explicit constructions of randomness extractors. Previously he did some work on quantum computing and human–computer interaction.

SHACHAR LOVETT graduated from the Weizmann Institute of Science in 2010; his advisors were Omer Reingold and Ran Raz. He was a member of the Institute for Advanced Study, School of Mathematics between 2010-2012. Since then, he has been a faculty member at the University of California, San Diego. He is interested in the role that structure and randomness play in computation and mathematics, and in particular in computational complexity, coding theory, pseudorandomness, and algebraic constructions.

JIAPENG ZHANG is an assistant professor in the Computer Science Department at the University of Southern California. He received his B. S. in 2011 from Shanghai Jiao Tong University, and his Ph. D. in 2019 from the University of California, San Diego, under the supervision of Shachar Lovett. He spent the year 2019–2020 as a postdoc at Harvard. His research interests include the analysis of Boolean functions, machine learning theory, computational complexity, and cryptography.