2015 IEEE 56th Annual Symposium on Foundations of Computer Science

How to refute a random CSP

Sarah R. Allen, Ryan O'Donnell, David Witmer
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA
Email: srallen, odonnell, dwitmer@cs.cmu.edu

Abstract

Let P be a k-ary predicate over a finite alphabet. Consider a random CSP(P) instance $\mathcal{I}$ over n variables with m constraints. When $m \gg n$ the instance will be unsatisfiable with high probability and we want to find a certificate of unsatisfiability. When P is the 3-ary OR predicate, this is the well-studied problem of refuting random 3-SAT formulas and an efficient algorithm is known only when $m \gg n^{3/2}$. Understanding the density required for refutation of other predicates is important in cryptography, proof complexity, and learning theory. Previously, it was known that for a k-ary predicate, having $m \gg n^{\lceil k/2 \rceil}$ constraints suffices for refutation. We give a criterion for predicates that often yields efficient refutation algorithms at much lower densities. Specifically, if P fails to support a t-wise uniform distribution, then there is an efficient algorithm that refutes random CSP(P) instances whp when $m \gg n^{t/2}$. Indeed, our algorithm will "somewhat strongly" refute the instance $\mathcal{I}$, certifying $\mathrm{Opt}(\mathcal{I}) \leq 1 - \Omega_k(1)$. If $t = k$ then we get the strongest possible refutation, certifying $\mathrm{Opt}(\mathcal{I}) \leq E[P] + o(1)$. This last result is new even for random k-SAT. Prior work on SDP hierarchies has given some evidence that efficient refutation of random CSP(P) may be impossible when $m \ll n^{t/2}$; thus there is an indication that our algorithm's dependence on m is optimal for every P, at least in the context of SDP hierarchies. As an application of our result, we falsify assumptions used to show hardness-of-learning in recent work of Daniely, Linial, and Shalev-Shwartz.

I. On refutation of random CSPs

Constraint satisfaction problems (CSPs) play a major role in computer science. There is a vast theory [23] of how algebraic properties of a CSP predicate affect its worst-case satisfiability complexity; there is a similarly vast theory [68] of worst-case approximability of CSPs. Finally, there is a rich range of research — from the fields of computer science, mathematics, and physics — on the *average-case* complexity of random CSPs; see [2] for a survey just of random k-SAT. This paper is concerned with random CSPs, and in particular the problem of efficiently *refuting* satisfiability for random instances. This is a well-studied algorithmic task with connections to, e.g., proof complexity [20], inapproximability [36], SAT-solvers [1], cryptography [10], learning theory [33], statistical physics [30], and complexity theory [15].

Historically, random CSPs are probably best studied in the case of k-SAT, $k \geq 3$. The model here involves choosing a CNF formula $\mathcal{I}$ over n variables by drawing m clauses (ORs of k literals) independently and uniformly at random. This is one of the best known efficient ways of generating hard-seeming instances of NP-complete and coNP-complete problems. The computational hardness depends crucially on the *density*, $\alpha = m/n$. For each k there is (conjecturally) a constant *critical density* α_k such that $\mathcal{I}$ is satisfiable with high probability when $\alpha < \alpha_k$, and $\mathcal{I}$ is unsatisfiable with high probability when $\alpha > \alpha_k$. (Here and throughout, "with high probability (whp)" means with probability $1 - o(1)$ as $n \to \infty$.) This phenomenon occurs for all nontrivial random CSPs; in the case of k-SAT it's been rigorously proven [35] for sufficiently large k.

There is a natural algorithmic task associated with each of the two regimes. When $\alpha < \alpha_k$ one wants to find a satisfying assignment for $\mathcal{I}$. When $\alpha > \alpha_k$ one wants to *refute* $\mathcal{I}$; i.e., find a *certificate* of unsatisfiability. Most heuristic SAT-solvers use DPLL-based algorithms; on unsatisfiable instances, they produce certificates that can be viewed as refutations within the Resolution proof system. More generally, a *refutation algorithm* for density α is any algorithm that: a) outputs "unsatisfiable" or "fail"; b) never incorrectly outputs "unsatisfiable"; c) outputs "fail" with low probability (i.e., probability $o(1)$).[1] Empirical work suggests that as α increases towards α_k, finding satisfying assignments becomes more difficult; and conversely, as α increases beyond α_k, finding certificates of unsatisfiability gradually becomes easier.

A seminal paper of Chvátal and Szemerédi [25] showed that for any sufficiently large integer c (depending on k), a random k-SAT instance with $m = cn$ requires Resolution refutations of size $2^{\Omega(n)}$ (whp). On the other hand, Fu [46] showed that polynomial-size Resolution refutations exist (whp) once $m \geq O(n^{k-1})$; Beame et al. [18]

[1] We caution the reader that in this paper we do not consider the related, but distinct, scenario of distinguishing *planted* random instances from truly random ones.

0272-5428/15 $31.00 © 2015 IEEE
DOI 10.1109/FOCS.2015.48

subsequently showed that such proofs could be found efficiently.[2] A breakthrough came in 2001, when Goerdt and Krivelevich [50] abandoned combinatorial refutations for spectral ones, showing that random k-SAT instances can be efficiently refuted when $m \geq \widetilde{O}(n^{\lceil k/2 \rceil})$. Soon thereafter, Friedman and Goerdt [44] (see also [45]) showed that for 3-SAT, efficient spectral refutations exist once $m \geq n^{3/2+\epsilon}$ (for any $\epsilon > 0$). These densities for k-SAT — around $n^{3/2}$ for 3-SAT and $n^{\lceil k/2 \rceil}$ in general — have not been fundamentally improved upon in the last 14 years.[3] (See Table I for a more detailed history of results.) Improving the $n^{3/2}$ bound for 3-SAT is widely regarded as a major open problem [10], with conjectures regarding its possibility going both ways [20], [32]. See also the work of Feige, Kim, and Ofek [38] showing that polynomial-size 3-SAT refutations *exist* whp once $m \geq O(n^{1.4})$.

In a notable paper from 2002, Feige [36] made a fruitful connection between the hardness of refuting random 3-SAT instances and the inapproximability of certain optimization problems that are challenging to analyze by other means. This refers to certifying not only that a random instance $\mathcal{I}$ is unsatisfiable, but furthermore that $\mathrm{Opt}(\mathcal{I}) \leq 1 - \delta$ for some constant $\delta > 0$. Here $\mathrm{Opt}(\mathcal{I})$ denotes the maximum *fraction* of simultaneously satisfiable constraints in $\mathcal{I}$. Feige specifically introduced the following "R3SAT Hypothesis": *For all small $\delta > 0$ and large $c \in \mathbb{N}$, there is no polynomial-time δ-refutation algorithm for random 3-SAT with $m = cn$.* To stress-test Feige's R3SAT Hypothesis, one may ask if the aforementioned refutation algorithms for k-SAT can be improved to δ-refutation algorithms. Coja-Oghlan et al. [27] showed that they can be in the case of $k = 3, 4$. Indeed, they gave algorithms for what is called *strong refutation* in these cases. Here strongly refuting k-SAT refers to certifying that $\mathrm{Opt}(\mathcal{I}) \leq 1 - 2^{-k} + o(1)$ (note that $\mathrm{Opt}(\mathcal{I}) \approx 1 - 2^{-k}$ whp assuming $m \geq O(n)$).

As in the algebraic and approximation theories of CSP, it's of significant interest to consider random instances of the CSP(P) problem for general predicates $P : \mathbb{Z}_q^k \to \{0, 1\}$, besides just Boolean OR. (Though Boolean predicates are more familiar, larger domains are of interest, e.g., for q-colorability of k-uniform hypergraphs.) Specifically, we are interested in the question of how properties of P affect the number of constraints needed for efficient refutation of random CSP(P) instances. This precise question is very relevant for work in cryptography based on the candidate OWFs and PRGs of Goldreich [52]; see also [10] and the survey of Applebaum [9]. It has also proven essential for the recent exciting approach to hardness of learning due to Daniely, Linial, and Shalev-Shwartz [32]–[34]. We discuss this learning connection and our results on it in more detail in Section V.

The special case of random 3-XOR has proved particularly important: it is related to 3-SAT refutation through Feige's "3XOR Principle" (see [36], [38], [40]); it's the basis for cryptographic schemes due to Alekhnovich [3] (and is related to the "Learning Parities with Noise" problem); it's used in the best known lower bounds for the SOS SDP hierarchy [54], [69]; and, Barak and Moitra [16] have shown it to be equivalent to a certain "tensor prediction problem" in learning theory.

Prior to this work, very little was known about how the predicate P affects the complexity of refuting random CSP(P) instances. The main known result, following from the work Coja-Oghlan, Cooper, and Frieze [26], was the following: For any Boolean k-ary predicate P, one can efficiently strongly refute random CSP(P) instances $\mathcal{I}$ (i.e., certify $\mathrm{Opt}(\mathcal{I}) \leq \mathbf{E}[P] + o(1)$) provided the number of constraints m satisfies $m \geq \widetilde{O}(n^{\lceil k/2 \rceil})$. In the case of k-XOR, the very recent work of Barak and Moitra [16] showed how to improve this bound to $m \geq \widetilde{O}(n^{k/2})$.[4]

II. Our results and techniques

Here we describe our main results and techniques at a high level. Precise theorem statements appear later in the work and the terminology we use is defined in Section III. We also mention that all of our results can be generalized to the case of larger alphabets (see [4]), but we discuss Boolean predicates $P : \{0, 1\}^k \to \{0, 1\}$ for simplicity.

Our main result gives a (possibly sharp) bound on the number of constraints needed to refute random CSP(P) instances. We first describe some more concrete results that go into the main proof. All of our results rely on a strong refutation algorithm for k-XOR (actually, a slight generalization thereof). For $m \geq \widetilde{O}(n^{\lceil k/2 \rceil})$, such a result follows from [26]; however, the exponent $\lceil k/2 \rceil$ can be improved to $k/2$. We give a demonstration of this fact herein; as mentioned earlier, it was published very recently by Barak and Moitra [16, Corollary 5.5 and Extensions].

[2] In this paper we use the following not-fully-standard terminology: A statement of the form "If $f(n) \geq O(g(n))$ then X" means that there exists a certain function $h(n)$, with $h(n)$ being $O(g(n))$, such that the statement "If $f(n) \geq h(n)$ then X" is true. We also use $\widetilde{O}(f(n))$ to denote $O(f(n) \cdot \mathrm{polylog}(f(n))$, and $O_k(f(n))$ to denote that the hidden constant has a dependence on k (most often of the form $2^{O(k)}$).

[3] Actually, it is claimed in [48] that one can obtain $n^{k/2+\epsilon}$ for odd k "along the lines of [44]". On one hand, this is true, as we'll see in this paper. On the other hand, no proof was provided in [48], and we have not found the claim repeated in any paper subsequent to 2003.

[4] The present authors also obtained this result around the same time, but we credit the result to [16] as they published earlier. With their permission we repeat the proof herein, partly because we need to prove a slightly more general variant.

CSP	Poly-size refutations whp once $m \geq \cdots$	Strength	Efficient/ Existential	Reference
k-SAT	$O(n^{k-1})$	Refutation	Existential	[46]
k-SAT	$O(n^{k-1}/\log^{k-2}(n))$	Refutation	Efficient	[18], [19]
k-SAT	$\widetilde{O}(n^{\lceil k/2 \rceil})$	Refutation	Efficient	[45], [50]
3-SAT	$O(n^{3/2+\epsilon})$	Refutation	Efficient	[44], [45]
k-SAT	$O(n^{k/2+\epsilon})$	Refutation	Efficient	Claimed possible in [48], [49]
Exactly-k_1-out-of-k-SAT	$\widetilde{O}(n)$	Refutation	Efficient	[20]
2-out-of-5-SAT	$O(n^{3/2+\epsilon})$	Refutation	Efficient	[48], [49]
NAE-3-SAT	$O(n)$	$\Omega(1)$-Refutation	Efficient	[48], [49], [57]
k-SAT	$O(n^{\lceil k/2 \rceil})$	Refutation	Efficient	[29], [40]
3-SAT	$\widetilde{O}(n^{3/2})$	Refutation	Efficient	[51]
3-SAT	$\widetilde{O}(n^{3/2})$	Strong	Efficient	[27], [28]
4-SAT	$O(n^2)$	Strong	Efficient	[27], [28]
3-SAT	$O(n^{3/2})$	Refutation	Efficient	[39]
3-SAT	$O(n^{1.4})$	Refutation	Existential	[38]
3-XOR	$O(n^{3/2})$	$1/n^{\Omega(1)}$-refutation	Efficient	Implicit in [38]
3-SAT	$\Omega(n^{3/2})$	Refutation	Efficient	Claimed in [38]
Boolean k-CSP	$O(n^{\lceil k/2 \rceil})$	Strong	Efficient	[26]
k-XOR	$\widetilde{O}(n^{k/2})$	Strong	Efficient	[16] (also herein)
Any k-CSP	$\widetilde{O}(n^{k/2})$	Quasirandom ($\implies$ strong)	Efficient	This paper
Any k-CSP not supporting t-wise indep.	$\widetilde{O}(n^{t/2})$	$\Omega_k(1)$-refutation	Efficient	This paper

Table I
UP TO LOGARITHMIC FACTORS ON m, OUR WORK SUBSUMES ALL PREVIOUSLY KNOWN RESULTS.

Theorem 1. *There is an efficient algorithm that (whp) strongly refutes random k-XOR instances with at least $\widetilde{O}(n^{k/2})$ constraints; i.e., it certifies $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + o(1)$.*

The proof of Theorem 1 follows ideas from [28] and earlier works on "discrepancy" of random k-SAT instances. The case of even k is notably easier, and we present two "folklore" arguments for it. The case of odd k is trickier. Roughly speaking we view the instance as a homogeneous degree-k multilinear polynomial, which we want to certify takes on only small values on inputs in $\{-1,1\}^n$. Considering separately the contributions based on the "last" of the k variables in each constraint, and then using Cauchy–Schwarz, it suffices to bound the norm of a carefully designed quadratic form of dimension n^{k-1}, indexed by tuples of $k-1$ variables. This is done using the trace method [47], [73]. Similar techniques, including the use of the trace method, date back to the 2001 Friedgman–Goerdt work [44] refuting random 3-SAT given $m = n^{3/2+\epsilon}$ constraints.

With Theorem 1 in hand, the next step is certifying *quasirandomness* of random k-ary CSP instances having $m \geq \widetilde{O}(n^{k/2})$ constraints. Roughly speaking we say that a CSP instance is quasirandom if, for every assignment $x \in \{0,1\}^n$, the m induced k-tuples of literal values are close to being uniformly distributed over $\{0,1\}^k$. (Note that this is only a property of the instances' constraint scopes/negations, and has nothing to do with P.) Since the "Vazirani XOR Lemma" implies that a distribution on $\{-1,1\}^k$ is uniform if and only if all its 2^k XORs are have bias $\frac{1}{2}$, we are able to leverage Theorem 1 to prove:

Theorem 2. *There is an efficient algorithm that (whp) certifies that a random instance of $\mathrm{CSP}(P)$ is quasirandom, provided the number of constraints is at least $\widetilde{O}(n^{k/2})$.*

If an instance is quasirandom, then no solution can be much better than a randomly chosen one. Thus by certifying quasirandomness we are able to strongly refute random instances of any $\mathrm{CSP}(P)$:

Theorem 3. *For any k-ary predicate P, there is an efficient algorithm that (whp) strongly refutes random $\mathrm{CSP}(P)$ instances when the number of constraints is at least $\widetilde{O}(n^{k/2})$.*

In particular, this theorem improves upon [26] by a factor of $\sqrt{n}$ whenever k is odd; this savings is new even in the well-studied case of k-SAT.

The above result does not make use of any properties of the predicate P other than its arity, k. We now come to our main result, which shows that for many interesting P, random $\mathrm{CSP}(P)$ instances can be refuted with many fewer constraints than $n^{k/2}$. In the following, the phrase "t-wise uniform" (often imprecisely called "t-wise independent") refers to a distribution on $\{-1,1\}^k$ in which all marginal distributions on t out of k coordinates are uniform.

Theorem 4. *(Main.) Let P be a k-ary predicate such that there is no t-wise uniform distribution supported on its satisfying assignments, $t \geq 2$. Then there is an efficient algorithm that (whp) $\Omega_k(1)$-refutes random instances of $\mathrm{CSP}(P)$ with at least $\widetilde{O}(n^{t/2})$ constraints.*

We remark that property of a predicate P supporting a pairwise uniform distribution has played an important role in approximability theory for CSPs, ever since Austrin and Mossel [13] showed that such predicates are hereditarily approximation-resistant under the UGC. Also, note that the largest t for which a predicate P supports a t-wise uniform distribution determines the minimum number of constraints required by our algorithm. This value is closely related to the notion of distribution complexity studied by Feldman, Perkins, and Vempala [41], [42] in the context of planted random CSPs. Informally, the distribution complexity of a planted CSP is the largest t for which the distribution over constraint inputs $\{-1,1\}^k$ induced by the planted assignment is t-wise uniform. Despite this similarity, the algorithmic techniques used by Feldman, Perkins, and Vempala in the planted case [41] do not seem to directly apply to refutation.

The idea behind the proof of Theorem 4 is that with $\widetilde{O}(n^{t/2})$ constraints we can use the algorithm of Theorem 2 to certify quasirandomness (closeness to uniformity) for all subsets of t out of k coordinates. Thus for every assignment $x \in \{0,1\}^n$, the induced distribution on constraint k-tuples is ($o(1)$-close to) t-wise uniform. Since P does not support a t-wise uniform distribution, this essentially shows that no x can induce a fully-satisfying distribution on constraint inputs. To handle the $o(1)$-closeness caveat, we show that if P does not support a t-wise uniform distribution, then it is δ-far from supporting such a distribution, for $\delta = \Omega_k(1)$. The algorithm can then in fact $(\delta - o(1))$-refute random $\mathrm{CSP}(P)$ instances.

To briefly illustrate the result, consider the Exactly-k-out-of-$2k$-SAT CSP, studied previously in [20], [49]. The associated predicate supports a 1-wise uniform distribution, namely the uniform distribution over strings in $\{0,1\}^{2k}$ of Hamming weight k. However, it is not hard to show that it does not support any pairwise uniform distribution. As a consequence, random instances of this CSP can be refuted with only $\widetilde{O}(n)$ constraints, independent of k.

A. An application from learning theory

Recently, an exciting approach to proving hardness-of-learning results has been developed by Daniely, Linial, and Shalev-Shwartz [31]–[34]. The most general results appear in [33]. In this work, Daniely et al. prove computational hardness of several central learning theory problems, based on two assumptions concerning the hardness of random CSP refutation. While the assumptions made in [31], [34] appear to be plausible, our work unfortunately shows that the more general assumptions made in [33] are false.

Below we state the (admittedly strong) assumptions from [33] (up to some very minor technical details which are discussed in Section V). We will need one piece of terminology: the 0-*variability* $\mathrm{VAR}_0(P)$ of a predicate $P : \{-1,1\}^k \to \{0,1\}$ is the least c such that there exists a restriction to some c input coordinates forcing P to be 0. Essentially, the assumptions state that one can obtain hardness-of-refutation with an arbitrarily large polynomial number of constraints by using a family of predicates (P_k) that: a) have unbounded 0-variability; b) support pairwise uniformity. However, our work shows that supporting t-wise uniformity for unbounded t is also necessary.

SRCSP Assumption 1. *([33].) For all $d \in \mathbb{N}$ there is a large enough C such that the following holds: If $P : \{-1,1\}^k \to \{0,1\}$ has VAR$_0(P) \geq C$ and is hereditarily approximation resistant on satisfiable instances, then there is no polynomial-time algorithm refuting (whp) random instances of CSP(P) with $m = n^d$ constraints.*

SRCSP Assumption 2. *([33], generalizing the "RCSP Assumption" of [15] to superlinearly many constraints.) For all $d \in \mathbb{N}$ there is a large enough C such that the following holds: If $P : \{-1,1\}^k \to \{0,1\}$ has VAR$_0(P) \geq C$ and is δ-close to supporting a pairwise uniform distribution, then for all $\epsilon > 0$ there is no polynomial-time algorithm that $(\delta + \epsilon)$-refutes (whp) random instances of CSP(P) with $m = n^d$ constraints.*

In [33] it is shown how to obtain three very notable hardness-of-learning results from these assumptions. However as stated, our work falsifies the SRCSP Assumptions. Indeed, the assumptions are false even in the three specific cases used by [33] to obtain hardness-of-learning results. We now describe these cases.

Case 1. The Huang predicates (H_κ) are arity-$\Theta(\kappa^3)$ predicates introduced in [56]; they are hereditarily approximation resistant on satisfiable instances and have 0-variability $\Omega(\kappa)$. In [33] they are used in SRCSP Assumption 1

to deduce hardness of PAC-learning DNFs with $\omega(1)$ terms. However:

Theorem 5. *For all $\kappa \geq 9$, the predicate H_κ does not support a 4-wise uniform distribution.*

Thus by Theorem 4 we can efficiently refute random instances of CSP(H_κ) with just $\widetilde{O}(n^2)$ constraints, independent of κ. This contradicts SRCSP Assumption 1.

Case 2. The majority predicate Maj_k has 0-variability $\lceil k/2 \rceil$ and is shown in [33] to be $\frac{1}{k+1}$-far from supporting a pairwise uniform distribution. In [33] these predicates are used in SRCSP Assumption 2 to deduce hardness of agnsotically learning Boolean halfspaces to within any constant factor. However:

Theorem 6. *For odd $k \geq 25$, the predicate Maj_k does not support a 4-wise uniform distribution; in fact, it is .1-far from supporting a 4-wise uniform distribution.*

Theorem 4 then implies we can efficiently δ-refute random instances of CSP(Maj_k) with $\widetilde{O}(n^2)$ constraints, where $\delta = .1 \gg \frac{1}{k+1}$. This contradicts SRCSP Assumption 2.

Case 3. Finally, we also prove that SRCSP Assumption 1 is false for another family of predicates (T_k) used by [33] to show hardness of PAC-learning intersections of 4 Boolean halfspaces.

Our results described in these three cases all use linear programming duality. Specifically, in Lemma 21 we show that P is δ-far from supporting a t-wise uniform distribution if and only if there exists a k-variable multilinear polynomial Q that satisfies certain properties involving P and δ. We then explicitly construct these dual polynomials for the Huang, Majority, and T_k predicates.

We conclude this section by emphasizing the importance of the Daniely–Linial–Shalev-Shwartz hardness-of-learning program, despite the above results. Indeed, subsequently to [33], Daniely and Shalev-Shwartz [34] showed hardness of improperly learning DNF formulas with $\omega(\log n)$ terms under a much weaker assumption than SRCSP Assumption 1. Specifically, their work only assumes that for all d there is a large enough k such that refuting random k-SAT instances is hard when there are $m = n^d$ constraints. This assumption looks quite plausible to us, and may even be true with k not much larger than $2d$. Most recently, Daniely showed hardness of approximately agnostically learning halfspaces using the XOR predicate rather than majority [31]. This result shows that there is no efficient algorithm that agnostically learns halfspaces to within a constant approximation ratio under the assumption that refuting random k-XOR instances is hard when $m = n^{c\sqrt{k}\log k}$ for some $c > 0$. He also shows hardness of learning halfspaces to within an approximation factor of $2^{\log^{1-\nu} n}$ for any $\nu > 0$ assuming that there exists some constant $c > 0$ such that for all s, refuting random k-XOR instances with $k = \log^s n$ is hard when $m = n^{ck}$.

B. Evidence for the optimality of our results

It's natural to ask whether the $n^{t/2}$ dependence in our main Theorem 4 can be improved. As we don't expect to prove unconditional hardness results, we instead merely seek good evidence that refuting $(t-1)$-wise supporting predicates P is hard when $m \ll n^{t/2}$. A natural form of evidence would be showing that various strong classes of polynomial-time refutation algorithms fail when $m \ll n^{t/2}$. To make sense of this we need to talk about the form of such algorithms; i.e., propositional proof systems.

Recently, there has been significant study of the "SOS" (Sum-Of-Squares) proof system, introduced by Grigoriev and Vorobjov [55]; see, e.g., [17], [66] for discussion. It has the following virtues: a) it is very powerful, being able to efficiently simulate other proof systems (e.g., Resolution, Lovász–Schrijver); b) it is automatizable [61], [67], meaning that n-variable "degree-d SOS proofs" can be found in $n^{O(d)}$ time whenever they exist; c) we do know some examples of *lower bounds* for degree-d SOS proofs. All of our refutation algorithms for k-ary predicates can be extended to produce degree-$2k$ SOS proofs (see [4]).

We now return to the question of evidence for the optimality of constraint density used in our results. Dating back to Franco–Paull [43] and Chvátal–Szemerédi [25], there has been a long line of work in proof complexity showing lower bounds for refuting random 3-SAT instances, especially in the Resolution proof system. This culminated in the work of Ben-Sasson and Wigderson [21], which showed that for random 3-SAT (and 3-XOR) with $m = O(n^{3/2-\epsilon})$, Resolution refutations require size $2^{n^{\Omega(\epsilon)}}$ (whp). More recently, Schoenebeck [69] showed (using the expansion analysis of [21]) that random k-XOR and k-SAT instances with $m \leq n^{k/2-\epsilon}$ require SOS proofs of degree $n^{\Omega(\epsilon)}$, and hence take $2^{n^{\Omega(\epsilon)}}$ time to refute by the "SOS Method". See [24], [70] for related larger-alphabet followups. These results show that the Barak–Moitra $\widetilde{O}(n^{k/2})$ bound for refuting random k-XOR (which also works in $O(k)$-degree SOS) and our bound for random k-SAT are tight (up to subpolynomial factors) within the SOS framework.

Given the power of the SOS framework, this arguably constitutes some reasonable evidence that no polynomial-time algorithm can refute random k-SAT instances with $m \ll n^{k/2}$.

We now discuss our main theorem's $n^{t/2}$ bound for predicates P not supporting t-wise uniform distributions. Suppose P is a predicate that does support a $(t-1)$-wise uniform distribution, where $t > 2$. In the context of inapproximability and SDP-hierarchy integrality gaps, this condition on P has been significantly studied in the case of $t = 3$. For P supporting pairwise uniformity, it is known [22], [71] that the Sherali–Adams and Lovász–Schrijver$^+$ SDP hierarchies require degree $\Omega(n)$ to refute random CSP(P) instances (whp) when $m = O(n)$. This result was also recently proven for the stronger SOS proof system by Barak, Chan, and Kothari [14], except that the CSP(P) instances are not quite uniformly random; they are "slightly pruned" random instances. For any $t > 2$, the second and third authors recently essentially showed [65] that for the Sherali–Adams$^+$ SDP hierarchy, degree $n^{\Omega(\epsilon)}$ is (whp) necessary to refute random CSP(P) instances when $m \leq n^{t/2-\epsilon}$. As a caveat, again the instances are slightly pruned random instances, rather than being purely uniformly random. (The instances in [65] are also in the "Goldreich [52] style"; i.e., there are no literals, but the "right-hand sides" are random. However it is not hard to show the proofs in [65] go through in the standard random model of this paper.) Future work [63] is devoted to removing the pruning in these instances. Although the Sherali–Adams$^+$ SDP hierarchy is certainly weaker than the SOS hierarchy, these works still constitute some evidence that our main theorem's requirement of $m \geq \widetilde{O}(n^{t/2})$ for non-t-wise supporting predicates may be essentially optimal.

Further evidence for the optimality of our results is provided by the work of Feldman, Perkins, and Vempala on statistical algorithms for random planted CSPs [42]. They show that their lower bounds against statistical algorithms for solving random planted CSPs also imply lower bounds against statistical algorithms for refuting uniformly random CSPs. Specifically, they prove that when P supports a $(t-1)$-wise uniform distribution, any statistical algorithm using queries that take $\widetilde{O}(n^{t/2})$ possible values can only refute random instances of $\mathrm{CSP}(P)$ with at least $\widetilde{\Omega}(n^{t/2})$ constraints. As an application of this result, they also show that any convex program refuting such an instance of $\mathrm{CSP}(P)$ must have dimension at least $\widetilde{\Omega}(n^{t/2})$.

III. Preliminaries and notation

A. Constraint satisfaction problems

We review some definitions and facts related to constraint satisfaction problems (CSPs). We discuss the Boolean domain, which we write as $\{-1,1\}$ rather than $\{0,1\}$. For $x \in \mathbb{R}^n$ and $S \subseteq [n]$ we write $x_S \in \mathbb{R}^{|S|}$ for the restriction of x to coordinates S; i.e., $(x_i)_{i \in S}$. We also use $\circ$ to denote the entrywise product for vectors.

Definition 7. Given a predicate $P : \{-1,1\}^k \to \{0,1\}$, an instance $\mathcal{I}$ of the CSP(P) problem over variables $x_1, \dots, x_n$ is a multiset of *P-constraints*. Each P-constraint consists of a pair (c, S), where $S \in [n]^k$ is the scope and $c \in \{-1,1\}^k$ is the negation pattern; this represents the constraint $P(c \circ x_S) = 1$. We typically write $m = |\mathcal{I}|$. Let $\mathrm{Val}_{\mathcal{I}}(x)$ be the fraction of constraints satisfied by assignment $x \in \{-1,1\}^n$, i.e., $\mathrm{Val}_{\mathcal{I}}(x) = \frac{1}{m}\sum_{(c,S)\in\mathcal{I}} P(c \circ x_S)$. The objective is to find an assignment x maximizing $\mathrm{Val}_{\mathcal{I}}(x)$. The *optimum* of $\mathcal{I}$, denoted by $\mathrm{Opt}(\mathcal{I})$, is $\max_{x\in\{-1,1\}^k} \mathrm{Val}_{\mathcal{I}}(x)$. If $\mathrm{Opt}(\mathcal{I}) = 1$, we say that $\mathcal{I}$ is *satisfiable*. We also write $\overline{P}$ for the quantity $\mathbf{E}_{z\sim\{-1,1\}^k}[P(z)]$; i.e., the fraction of assignments that satisfy P. For any instance $\mathcal{I}$ in which each constraint involves k different variables, we have $\mathrm{Opt}(\mathcal{I}) \geq \overline{P}$.[5]

We next define a standard random model for CSPs. For $P : \{-1,1\}^k \to \{0,1\}$, let $\mathcal{F}_P(n,p)$ be the distribution over CSP instances given by including each of the $2^k n^k$ possible constraints independently with probability p. Note that we may include constraints on different permutations of the same set of variables, constraints on the same tuple of variables with different negations c, and constraints with the same variable occurring as more than one argument. It is reasonable to include such constraints in the case that the predicate P is not symmetric. We use $\overline{m}$ to denote the expected number of constraints, namely $2^k n^k p$. As noted in Fact 12 below, the number of constraints m in a draw from $\mathcal{F}_P(n,p)$ is tightly concentrated around $\overline{m}$, and we often blur the distinction between these parameters.

Quasirandomness. We now introduce an important notion for this paper: that of a CSP instance being *quasirandom*. Versions of this notion originate in the works of Goerdt and Lanka [51] (under the name "discrepancy"), Khot [58] ("quasi-randomness"), Austrin and Håstad [12] ("adaptive uselessness"), and Chan [24] ("low correlation"), among other places. To define it, we first define the induced distribution of an instance and an assignment.

[5]Technically, our definitions allow constraints with a variable appearing more than once, so $\mathrm{Opt}(\mathcal{I}) \geq \overline{P}$ doesn't always hold for us. But since we only consider random $\mathcal{I}$, we'll in fact have $\mathrm{Opt}(\mathcal{I}) \approx \overline{P}$ whp over $\mathcal{I}$ anyway.

Definition 8. Given a CSP instance $\mathcal{I}$ and and an assignment $x \in \{-1,1\}^n$, the *induced distribution*, denoted $\mathcal{D}_{\mathcal{I},x}$, is the probability distribution on $\{-1,1\}^k$ where the probability mass on $\alpha \in \{-1,1\}^k$ is given by $\mathcal{D}_{\mathcal{I},x}(\alpha) = \frac{1}{|\mathcal{I}|} \cdot \#\{(c,S) \in \mathcal{I} \mid c \circ x_S = \alpha\}$. In other words, it is the empirical distribution on inputs to P generated by the constraint scopes/negations on assignment x. Note that the predicate P itself is irrelevant to this notion. We will drop the subscript $\mathcal{I}$ when it is clear from the context. We define $D_{\mathcal{I},x} = 2^k \cdot \mathcal{D}_{\mathcal{I},x}$ to be the density function associated with $\mathcal{D}_{\mathcal{I},x}$.

We can now define quasirandomness.

Definition 9. A CSP instance $\mathcal{I}$ is *ϵ-quasirandom* if $\mathcal{D}_{\mathcal{I},x}$ is ϵ-close to the uniform distribution for all $x \in \{-1,1\}^n$; i.e., if $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x}, U^k) \leq \epsilon$ for all $x \in \{-1,1\}^n$.

Here we use the notation U^k for the uniform distribution on $\{-1,1\}^k$ as well as the following:

Definition 10. If $\mathcal{D}$ and $\mathcal{D}'$ are probability distributions on the same finite set A then $d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}')$ denotes their *total variation distance*, $\frac{1}{2}\sum_{\alpha \in A} |\mathcal{D}(\alpha) - \mathcal{D}'(\alpha)|$. If $d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}') \leq \epsilon$ we say that $\mathcal{D}$ and $\mathcal{D}'$ are *ϵ-close*. If $d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}') \geq \epsilon$ we say they are *ϵ-far*. (As neither inequality is strict, these notions are not quite opposites.)

An immediate consequence of an instance being quasirandom is that its optimum is close to $\overline{P}$:

Fact 11. *If $\mathcal{I}$ is ϵ-quasirandom, then $\mathrm{Opt}(\mathcal{I}) \leq \overline{P} + \epsilon$ (and in fact, $\left|\mathrm{Opt}(\mathcal{I}) - \overline{P}\right| \leq \epsilon$).*

We conclude the discussion of CSPs by recording some facts that are proven easily with the Chernoff bound:

Fact 12. *Let $\mathcal{I} \sim \mathcal{F}_P(n,p)$. Then the following statements hold with high probability.*

1) $m = |\mathcal{I}| \in \overline{m} \cdot \left(1 \pm O\left(\sqrt{\frac{\log n}{\overline{m}}}\right)\right)$.
2) $\mathrm{Opt}(\mathcal{I}) \leq \overline{P} \cdot \left(1 + O\left(\sqrt{\frac{1}{\overline{P}} \cdot \frac{n}{\overline{m}}}\right)\right)$.
3) *$\mathcal{I}$ is $O\left(\sqrt{2^k \cdot \frac{n}{\overline{m}}}\right)$-quasirandom.*

B. Algorithms and refutations on random CSPs

Definition 13. Let P be a Boolean predicate. We say that $\mathcal{A}$ is *δ-refutation algorithm for random* $\mathrm{CSP}(P)$ *with $\overline{m}$ constraints* if $\mathcal{A}$ has the following properties. First, on all instances $\mathcal{I}$ the output of $\mathcal{A}$ is either the statement "$\mathrm{Opt}(\mathcal{I}) \leq 1-\delta$" or is "fail". Second, $\mathcal{A}$ is *never* allowed to *err*, where erring means outputting "$\mathrm{Opt}(\mathcal{I}) \leq 1-\delta$" on an instance which actually has $\mathrm{Opt}(\mathcal{I}) > 1-\delta$. Finally, $\mathcal{A}$ must satisfy

$$\mathop{\mathbf{Pr}}_{\mathcal{I} \sim \mathcal{F}_P(n,p)}[\mathcal{A}(\mathcal{I}) = \text{"fail"}] < o(1) \quad (\text{as } n \to \infty),$$

where p is defined by $\overline{m} = 2^k n^k p$. Although $\mathcal{A}$ is often a deterministic algorithm, we do allow it to be randomized, in which case the above probability is also over the "internal random coins" of $\mathcal{A}$.

We refer to this notion as *weak refutation*, or simply *refutation*, when the certification statement is of the form "$\mathrm{Opt}(\mathcal{I}) < 1$" (equivalently, when $\delta = 1/|\mathcal{I}|$). We refer to the notion as *strong refutation* when the statement is of the form "$\mathrm{Opt}(\mathcal{I}) \leq \overline{P} + o(1)$" (equivalently, when $\delta = 1 - \overline{P} + o(1)$).

Remark 14. Section V a "two-sided error" variant of this definition. This is the easier algorithmic task in which we relax the condition on erring: it is only required that for each instance $\mathcal{I}$ with $\mathrm{Opt}(\mathcal{I}) > 1-\delta$, it holds that $\mathbf{Pr}[\mathcal{A}(\mathcal{I}) = \text{"}\mathrm{Opt}(\mathcal{I}) \leq 1-\delta\text{"}] \leq 1/4$, where the probability is over the random coins of $\mathcal{A}$.

Remark 15. We will also use the analogous definition for certification of related properties; e.g., we will discuss *ϵ-quasirandomness certification algorithms* in which the statement "$\mathrm{Opt}(\mathcal{I}) \leq 1-\delta$" is replaced by the statement "$\mathcal{I}$ is ϵ-quasirandom".

C. t-wise uniformity

An important notion for this paper is that of t-wise uniformity. Recall:

Definition 16. Probability distribution $\mathcal{D}$ on $\{-1,1\}^k$ is said to be *t-wise uniform*, $1 \leq t \leq k$, if for all $S \subseteq [k]$ with $|S| = t$ the random variable x_S is uniform on $\{-1,1\}^t$ when $x \sim \mathcal{D}$. (We remark that this condition is sometimes inaccurately called "t-wise independence" in the literature.)

We also consider the more general notion of (ϵ,t)-wise uniformity, typically defined using *Fourier coefficients*:

Definition 17. Probability distribution $\mathcal{D}$ on $\{-1,1\}^k$ is said to be (ϵ,t)-*wise uniform* if $|\widehat{D}(S)| \leq \epsilon$ for all $S \subseteq [k]$ with $0 < |S| \leq t$, where $D = 2^k \cdot \mathcal{D}$ is the probability density associated with distribution $\mathcal{D}$.

Here we are using standard notation from Fourier analysis of Boolean functions [64]. In particular, for any $f : \{-1,1\}^k \to \mathbb{R}$ we write $f(x) = \sum_{S \subseteq [k]} \widehat{f}(S) x^S$ for its expansion as a multilinear polynomial over $\mathbb{R}$, with x^S denoting $\prod_{i \in S} x_i$ (not to be confused with the projection $x_S \in \mathbb{R}^{|S|}$). It is a simple fact (and it follows from Lemma 18 below) that $(0,t)$-wise uniformity is equivalent to t-wise uniformity.

Also important for us is a related but distinct notion, that of being ϵ-close to a t-wise uniform distribution. It's easy to show that if $\mathcal{D}$ is ϵ-close to a t-wise uniform distribution then $\mathcal{D}$ is $(2\epsilon,t)$-wise uniform. In the other direction, we have the following (see also [5] for some quantitative improvement):

Lemma 18. (Alon–Goldreich-Mansour [7, Theorem 2.1]). *If $\mathcal{D}$ is an (ϵ,t)-wise uniform distribution on $\{-1,1\}^k$, then there exists a t-wise uniform distribution $\mathcal{D}'$ on $\{-1,1\}^k$ with*

$$d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}') \leq \Big(\sum_{i=1}^{t} \tbinom{k}{t}\Big) \cdot \epsilon \leq k^t \cdot \epsilon.$$

In particular if $t = k$ we have the bound $2^k \cdot \epsilon$ (and this can also be improved [53] to $2^{k/2-1} \cdot \epsilon$).

Finally, we make a crucial definition:

Definition 19. A predicate $P : \{-1,1\}^k \to \{0,1\}$ is said to be *t-wise supporting* if there is a t-wise uniform distribution $\mathcal{D}$ whose support is contained in $P^{-1}(1)$. We say P is *δ-far from t-wise supporting* if every t-wise uniform distribution $\mathcal{D}$ is δ-far from being supported on P; i.e., has probability mass at least δ on $P^{-1}(0)$.

D. A dual characterization of limited uniformity

It is known that the condition of P supporting a t-wise uniform distribution is equivalent to the feasibility of a certain linear program; hence one can show that P is *not* t-wise supporting by exhibiting a certain dual object, namely a polynomial. This appears, e.g., in work of Austrin and Håstad [11, Theorem 3.1]. Herein we will extend this fact by giving a dual characterization of being far from t-wise supporting.

Definition 20. Let $0 < \delta < 1$. For a multilinear polynomial $Q : \{-1,1\}^k \to \mathbb{R}$, we say that Q *δ-separates* $P : \{-1,1\}^k \to \{0,1\}$ if the following conditions hold:

- $Q(z) \geq \delta - 1 \quad \forall z \in \{-1,1\}^k$;
- $Q(z) \geq \delta \quad \forall z \in P^{-1}(1)$;
- $\widehat{Q}(\emptyset) = 0$, i.e., Q has no constant coefficient.

We now provide the quantitative version of the aforementioned [11, Theorem 3.1]:

Lemma 21. *Let $P : \{-1,1\}^k \to \{0,1\}$ and let $0 < \delta < 1$. Then P is δ-far from t-wise supporting if and only if there is a δ-separating polynomial for P of degree at most t.*

Proof: The proof is an application of linear programming duality. Consider the following LP, which has variables $\mathcal{D}(z)$ for each $z \in \{-1,1\}^k$.

$$\begin{aligned}
\text{minimize} \quad & \sum_{z \in \{-1,1\}^k} (1 - P(z)) \mathcal{D}(z) && && (1)\\
\text{s.t.} \quad & \sum_{z \in \{-1,1\}^k} \mathcal{D}(z) z^S = 2^k \cdot \widehat{\mathcal{D}}(S) = 0 && \forall S \subseteq [k] \quad 0 < |S| \leq t && (2)\\
& \sum_{z \in \{-1,1\}^k} \mathcal{D}(z) = 1 && && (3)\\
& \mathcal{D}(z) \geq 0 && \forall z \in \{-1,1\}^k &&
\end{aligned}$$

Constraint (3) and the nonnegativity constraint ensure that $\mathcal{D}$ is a probability distribution on $\{-1,1\}^k$. Constraint (2) expresses that $\mathcal{D}$ is t-wise uniform. The objective (1) is minimizing the probability mass that $\mathcal{D}$ puts on assignments in $P^{-1}(0)$. Thus the optimal value of the LP is equal to the smallest γ such that P is γ-close to t-wise supporting; equivalently, the largest γ such that P is γ-far from t-wise supporting.

The following is the dual of the above LP. It has a variable $c(S)$ for each $0 < |S| \leq t$ as well as a variable ξ corresponding to constraint (3).

$$\text{maximize} \qquad \xi \tag{4}$$

$$\text{s.t.} \qquad \sum_{\substack{S \subseteq [k] \\ 0<|S|\leq t}} c(S) z^S \leq 1 - P(z) - \xi \qquad \forall z \in \{-1,1\}^k. \tag{5}$$

Observe that a feasible solution $(\{c(S)\}_S, \xi)$ is precisely equivalent to a multilinear polynomial Q of degree at most t, namely $Q(z) = -\sum_S c(S) z^S$, that ξ-separates P.

Thus P is δ-far from t-wise supporting if and only if the LP's objective (1) is at least δ, if and only if the dual's objective (4) is at least δ, if and only if there is a δ-separating polynomial for P of degree at most t. ∎

From this proof we can also derive that if P fails to be t-wise supporting then it must in fact be $\Omega_k(1)$-far from being t-wise supporting:

Corollary 22. *Suppose $P : \{-1,1\}^k \to \{0,1\}$ is not t-wise supporting. Then it is in fact δ-far from t-wise supporting for $\delta = 2^{-\widetilde{O}(k^t)}$ (or $\delta = 2^{-\widetilde{O}(2^k)}$ when $t = k$).*

Proof: Let $K = 1 + \sum_{i=1}^{t} \binom{k}{t}$ be the number of variables in the dual LP from Lemma 21, so $K \leq k^t + 1$ in general, with $K \leq 2^k$ when $t = k$. By assumption, the objective (4) of the dual LP's optimal solution is strictly positive. This optimum occurs at a vertex, which is the solution of a linear system given by a $K \times K$ matrix of ± 1 entries and a "right-hand side" vector with $0, 1$ entries. By Cramer's rule, the solution's entries are ratios of determinants of integer matrices with entries in $\{-1, 0, 1\}$. Thus any strictly positive entry is at least $1/N$, where N is the maximum possible such determinant. By Hadamard's inequality, $N = K^{K/2}$ and the claimed result follows. ∎

IV. Quasirandomness and its implications for refutation

A. Strong refutation of k-XOR

In this section, we state our result on strong refutation of random k-XOR instances with $m = \tilde{O}\left(n^{k/2}\right)$ constraints. (Recall that essentially this result was very recently obtained by Barak and Moitra [16].) We actually have a slightly more general result, allowing variables and coefficients to take values in $[-1, 1]$ and not just in $\{-1, 1\}$. This additional freedom is used in [4] to prove refutation results for CSPs over larger alphabets as well as for independence number and chromatic number of random hypergraphs.

Theorem 23. *For $k \geq 2$ and $p \geq n^{-k/2}$, let $\{w(T)\}_{T \in [n]^k}$ be independent random variables such that for each $T \in [n]^k$,*

$$\mathbf{E}[w(T)] = 0 \tag{6}$$

$$\mathbf{Pr}[w(T) \neq 0] \leq p \tag{7}$$

$$|w(T)| \leq 1. \tag{8}$$

Then there is an efficient algorithm certifying that

$$\sum_{T \in [n]^k} w(T) x^T \leq 2^{O(k)} \sqrt{p} n^{3k/4} \log^{3/2} n.$$

for all $x \in \mathbb{R}^n$ such that $\|x\|_\infty \leq 1$ with high probability.

In this form, the theorem is not really about CSP refutation at all. It says that the value of a polynomial with random coefficients is close to its expectation when its inputs are bounded.

The full proof is given in [4]. It follows techniques from [28] fairly closely and is essentially the same as the proof of [16]. We will use this theorem to prove our results in subsequent sections.

We obtain strong refutation of k-XOR as a simple corollary.

Corollary 24. *For $k \geq 2$, let $\mathcal{I} \sim \mathcal{F}_{k\text{-XOR}}(n, p)$. Then, with high probability, there is a degree-$2k$ SOS proof that $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + \gamma$ when $\overline{m} \geq \frac{2^{O(k)} n^{k/2} \log^3 n}{\gamma^2}$.*

Proof: We can write the k-XOR predicate as

$$k\text{-XOR}(z) = \frac{1 - \prod_{i=1}^{k} z_i}{2},$$

so for a k-XOR instance $\mathcal{I} \sim \mathcal{F}_{k\text{-XOR}}(n,p)$,

$$\mathrm{Val}_{\mathcal{I}}(x) = \frac{1}{2} - \frac{1}{2m} \sum_{T \in [n]^k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} x^T \prod_{i \in [k]} c_i \;=\; \frac{1}{2} + \frac{2^{k-1}}{m} \sum_{T \in [n]^k} w(T) x^T,$$

where $w(T) = -2^{-k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} \prod_{i \in [k]} c_i$. The $w(T)$'s are random variables depending on the choice of $\mathcal{I}$; observe that $\mathbf{E}[w(T)] = 0$, $\mathbf{Pr}[w(T) \neq 0] \leq 2^k p$, and $|w(T)| \leq 1$ for all $T \in [n]^k$. By Theorem 23, there is an algorithm certifying that

$$\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + \frac{2^{O(k)} \sqrt{p} n^{3k/4} \log^{3/2} n}{m}.$$

with high probability when $p \geq n^{-k/2}$. Since $m = (1+o(1))\overline{m}$ with high probability, choosing $\overline{m} \geq \frac{2^{O(k)} n^{k/2} \log^3 n}{\gamma^2}$ gives the desired result. ∎

As an example, we can choose $\gamma = \frac{1}{\log n}$ and certify that $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + o(1)$ when $\overline{m} = \widetilde{O}_k(n^{k/2})$.

B. Quasirandomness and strong refutation of any k-CSP

Next, we will use the algorithm of Theorem 23 to certify that an instance of CSP(P) is quasirandom. This will immediately give us a strong refutation algorithm.

In order to certify quasirandomness, Lemma 18 implies that it suffices to certify each Fourier coefficient of $D_{\mathcal{I},x}$ has small magnitude.

Lemma 25. *Let $\emptyset \neq S \subseteq [k]$ with $|S| = s$. There is an algorithm that, with high probability, certifies that*

$$\left|\widehat{D_{\mathcal{I},x}}(S)\right| \leq \frac{2^{O(s)} \max\{n^{s/4}, \sqrt{n}\} \log^{5/2} n}{\overline{m}^{1/2}}$$

for all $x \in \{-1,1\}^n$, assuming also that $\overline{m} \geq \max\{n^{s/2}, n\}$.

To prove this lemma, we need another lemma certifying that polynomials whose coefficients are sums of 0-mean random variables have small value.

Lemma 26. *Let $S \subseteq [k]$ with $|S| = s > 0$. Let $\tau \in \mathbb{N}$ and let $\{w_U(i)\}_{U \in [n]^s, i \in [\tau]}$ be independent random variables satisfying conditions* (6)*,* (7)*, and* (8) *for some $p \geq \frac{1}{\tau n^{s/2}}$. Then there is an algorithm that certifies with high probability that*

$$\sum_{U \in [n]^s} x^U \sum_{j=1}^{\tau} w_U(j) \leq \begin{cases} 2^{O(s)} \sqrt{\tau p} \cdot n^{3s/4} \log^{5/2} n & \text{if } s \geq 2 \\ 4 \max\{\sqrt{\tau p}, 1\} \cdot n \log n & \text{if } s = 1. \end{cases}$$

for all $x \in \mathbb{R}^n$ such that $\|x\|_\infty \leq 1$.

The proof is straightforward and we defer it to Section IV-D.

Proof of Lemma 25: Without loss of generality, assume $1 \in S$. Applying definitions, we see that

$$\widehat{D_{\mathcal{I},x}}(S) = \mathop{\mathbf{E}}_{z \sim \mathcal{D}_{\mathcal{I},x}}\left[z^S\right] = \frac{1}{m} \sum_{U \in [n]^s} \sum_{\substack{T \in [n]^k \\ T_S = U}} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} c^S x^U = \frac{1}{m} \sum_{U \in [n]^s} x^U \sum_{\substack{T \in [n]^k \\ T_S = U}} \sum_{c' \in \{\pm 1\}^{k-1}} w_S(T,c'). \quad (9)$$

where we define $w_S(T,c') = 1_{\{(T,(1,c')) \in \mathcal{I}\}}(c')^{S \setminus \{1\}} - 1_{\{(T,(-1,c')) \in \mathcal{I}\}}(c')^{S \setminus \{1\}}$ and recall that T_S is the projection of T onto the the coordinates in S. It is clear that $\mathbf{E}[w_S(T,c')] = 0$, $\mathbf{Pr}[w_S(T,c') \neq 0] \leq p$, and $|w_S(T,c')| \leq 1$. There are $\tau = 2^{k-1} n^{k-s}$ terms in each sum of $w_S(T,c')$'s and we can apply Lemma 26. When $s = 2$, we plug in these values and see that we can certify that $\widehat{D_{\mathcal{I},x}}(S) \leq \frac{2^{O(s)} n^{s/4} \log^{5/2} n}{\overline{m}^{1/2}}$. When $s = 1$, $\overline{m} \geq n$ implies that $\tau p \geq \frac{1}{2}$ and we can certify that $\widehat{D_{\mathcal{I},x}}(S) \leq \frac{2^{O(s)} \sqrt{n} \log n}{\overline{m}^{1/2}}$. The lower bound can be proved in exactly the same way by considering the random variables $-w_S(T,c')$. ∎

The existence of an algorithm for certifying quasirandomness follows from Lemmas 18 and 25.

Theorem 27. *There is an efficient algorithm that certifies that an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\mathrm{CSP}(P)$ is γ-quasirandom with high probability when $\overline{m} \geq \frac{2^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.*

Since γ-quasirandomess implies that $\mathrm{Opt}(\mathcal{I}) \leq \overline{P} + \gamma$, this algorithm also strongly refutes $\mathrm{CSP}(P)$.

Theorem 28. *There is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq \overline{P} + \gamma$ with high probability when $\overline{m} \geq \frac{2^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.*

C. (ϵ, t)-quasirandomness and $\Omega(1)$-refutation of non-t-wise-supporting CSPs

If a predicate is not t-wise supporting, a weaker notion of quasirandomness suffices to obtain $\Omega(1)$-refutation.

Definition 29. An instance $\mathcal{I}$ of $\mathrm{CSP}(P)$ is (ϵ, t)-quasirandom if $\mathcal{D}_{\mathcal{I},x}$ is (ϵ, t)-wise uniform for every $x \in \{-1,1\}^n$.

Fact 12 shows that random instances with $\widetilde{O}(n)$ constraints are $(o(1), t)$-quasirandom for all $t \leq k$ with high probability. Lemma 25 directly implies that we can certify (ϵ, t)-quasirandomness when $m \geq \widetilde{O}(n^{t/2})$.

Theorem 30. *There is an efficient algorithm that certifies that an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\mathrm{CSP}(P)$ is (γ, t)-quasirandom with high probability when $\overline{m} \geq \frac{2^{O(t)} n^{t/2} \log^5 n}{\gamma^2}$ and $t \geq 2$.*

We now reach the main result of this section, which states that if a predicate is δ-far from t-wise supporting, then we can almost δ-refute instances of $\mathrm{CSP}(P)$.

Theorem 31. *Let P be δ-far from t-wise supporting. There is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq 1 - \delta + \gamma$ whp when $\overline{m} \geq \frac{k^{O(t)} n^{t/2} \log^5 n}{\gamma^2}$ and $t \geq 2$.*

We give two proofs of this theorem. In Proof 1, the theorem follows directly from certification of (γ, t)-quasirandomness and Lemma 18.

Proof 1: Run the algorithm of Theorem 30 to certify that $\mathcal{I}$ is $(\gamma/k^t, t)$-quasirandom with high probability. By definition, we have certified that $\mathcal{D}_{\mathcal{I},x}$ is $(\gamma/k^t, t)$-wise uniform for all $x \in \{-1,1\}^n$,. Lemma 18 then implies that for all x there exists a t-wise uniform distribution $\mathcal{D}'_x$ such that $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x}, \mathcal{D}'_x) \leq \gamma$. Now define $\mathcal{D}_{\mathrm{sat}}$ to be an arbitrary distribution over satisfying assignments to P. Since P is δ-far from being t-wise supporting, we know that $d_{\mathrm{TV}}(\mathcal{D}', \mathcal{D}_{\mathrm{sat}}) \geq \delta$ for any t-wise uniform distribution $\mathcal{D}'$. The triangle inequality then implies that $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x}, \mathcal{D}_{\mathrm{sat}}) \geq \delta - \gamma$ for all $x \in \{-1,1\}^n$ and the theorem follows. ∎

Proof 2 gives a slightly weaker version of Theorem 31, requiring the stronger assumption that $\overline{m} \geq \frac{2^{O(k)} n^{t/2} \log^5 n}{\gamma^2}$. It is based on the dual polynomial characterization of being δ-far from t-wise supporting. While perhaps less intuitive than Proof 1, Proof 2 is more direct. It only uses the XOR refutation algorithm and bypasses [7]'s connection between (ϵ, t)-wise uniformity and ϵ-closeness to a t-wise uniform distribution. Proof 2 requires Plancherel's Theorem, a fundamental result in Fourier analysis.

Theorem 32 (Plancherel's Theorem)**.** *For any $f, g : \{-1,1\}^k \to \mathbb{R}$,*

$$\mathop{\mathbf{E}}_{\boldsymbol{z} \in U^k}[f(\boldsymbol{z})g(\boldsymbol{z})] = \sum_{S \subseteq [k]} \widehat{f}(S)\widehat{g}(S).$$

Proof 2: Since P is δ-far from t-wise supporting, there exists a degree-t polynomial Q that δ-separates P. The definition of δ-separating implies that $P(z) - (1-\delta) \leq Q(z)$ for all $z \in \{-1,1\}^k$. Summing over all constraints, we get that for all $x \in \{-1,1\}^n$,

$$\sum_{T \in [n]^k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} P(x_T \circ c) - m(1-\delta) \leq \sum_{T \in [n]^k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} Q(x_T \circ c),$$

or, equivalently, $\mathrm{Val}_{\mathcal{I}}(x) - (1-\delta) \leq \mathbf{E}_{\boldsymbol{z} \in \mathcal{D}_{\mathcal{I},x}}[Q(\boldsymbol{z})]$.

It then remains to certify that $\mathbf{E}_{\boldsymbol{z} \in \mathcal{D}_{\mathcal{I},x}}[Q(\boldsymbol{z})] \leq \gamma$. Observe that

$$\mathop{\mathbf{E}}_{\boldsymbol{z} \in \mathcal{D}_{\mathcal{I},x}}[Q(\boldsymbol{z})] = \mathop{\mathbf{E}}_{\boldsymbol{z} \in U^k}[D_{\mathcal{I},x}(\boldsymbol{z})Q(\boldsymbol{z})] = \sum_{\emptyset \neq S \subseteq [k]} \widehat{D_{\mathcal{I},x}}(S)\widehat{Q}(S),$$

where the second equality follows from Plancherel's Theorem. Since $Q \geq -1$ and $\mathbf{E}[Q] = 0$, $Q \leq 2^k$ and hence $|\widehat{Q}(S)| \leq 2^k$ for all S. To finish the proof, we apply Lemma 25 to certify that $\left|\widehat{D_{\mathcal{I},x}}(S)\right| \leq \frac{\gamma}{2^{2k}}$ for all S. ∎

With Corollary 22, Theorem 31 implies that we can $\Omega_k(1)$-refute instances of $\mathrm{CSP}(P)$ with $\widetilde{O}_k(n^{t/2})$ constraints when P is not t-wise supporting.

Corollary 33. *Let P be a predicate that does not support any t-wise uniform distribution. Then there is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq 1 - 2^{-\widetilde{O}(k^t)}$ with high probability when $\overline{m} \geq 2^{\widetilde{O}(k^t)} n^{t/2} \log^5 n$ and $t \geq 2$.*

D. Proof of Lemma 26

Lemma 26. *Let $S \subseteq [k]$ with $|S| = s > 0$. Let $\tau \in \mathbb{N}$ and let $\{w_U(i)\}_{U \in [n]^s, i \in [\tau]}$ be independent random variables satisfying conditions* (6)*,* (7)*, and* (8) *for some $p \geq \frac{1}{\tau n^{s/2}}$. Then there is an algorithm that certifies with high probability that*

$$\sum_{U \in [n]^s} x^U \sum_{j=1}^{\tau} w_U(j) \leq \begin{cases} 2^{O(s)} \sqrt{\tau p} \cdot n^{3s/4} \log^{5/2} n & \text{if } s \geq 2 \\ 4 \max\{\sqrt{\tau p}, 1\} \cdot n \log n & \text{if } s = 1. \end{cases}$$

for all $x \in \mathbb{R}^n$ such that $\|x\|_\infty \leq 1$.

The proof uses Bernstein's Inequality.

Theorem 34 (Bernstein's Inequality)**.** *Let $X_1, \ldots, X_M$ be independent 0-mean random variables such that $|X_i| \leq B$. Then, for $a > 0$,*

$$\mathbf{Pr}\left[\sum_{i=1}^{M} X_i > a\right] \leq \exp\left(\frac{-\frac{1}{2}a^2}{\sum_{i=1}^{M} \mathbf{E}[X_i^2] + \frac{1}{3}Ba}\right).$$

Proof of Lemma 26: First, we define

$$v_U = \sum_{j=1}^{\tau} w_U(i).$$

Observe that the v_U's are independent and that each one is the sum of τ mean-0, i.i.d. random variables with magnitude at most 1. Noting that $\sum_{i=1}^{\tau} \mathbf{E}[w_U(i)^2] \leq \tau p$, we can use Bernstein's Inequality to show that the $|v_U|$'s are not too big with high probability. If $s \geq 2$, Theorem 23 then implies that the desired algorithm exists. If $s = 1$, we are simply bounding a linear function over ± 1 variables. We consider two cases: Small p and large p.

Case 1: $p \leq \frac{1}{4\tau}$.

Choosing $a = 2s \log n$ in Bernstein's Inequality, we see that $\mathbf{Pr}[|v_U| \geq 2s \log n] \leq n^{-2s}$. A union bound over all U then implies that $\mathbf{Pr}[\text{any } |v_U| > 2s \log n] \leq n^{-s}$. If $s \geq 2$, we observe that $\mathbf{Pr}[v_U \neq 0] \leq \tau p$, scale the v_U's down by $2s \log n$, and apply Theorem 23 to get the stated result. If $s = 1$, we obtain the second bound by observing that

$$\sum_{i \in [n]} v_i x_i \leq \sum_{i \in [n]} |v_i| \leq 2n \log n. \tag{10}$$

Case 2: $p > \frac{1}{4\tau}$.

We set $a = 4s\sqrt{\tau p} \log n$ and get that $\mathbf{Pr}[\text{any } |v_U| > 4s\sqrt{\tau p} \log n] \leq n^{-s}$ as above. If $s \geq 2$, we then divide the v_U's by $4s\sqrt{\tau p} \log n$ and apply Theorem 23. If $s = 1$, we get a bound of $4\sqrt{\tau p} \cdot n \log n$ as in (10). ∎

V. Hardness of learning implications

Recent work by Daniely et al. [33] reduces the problem of refuting specific instances of $\mathrm{CSP}(P)$ to the problem of improperly learning certain hypothesis classes in the Probably Approximately Correct (PAC) model [72]. A definition of the model is also given in [33]. Daniely et al. reduce the problem of distinguishing between random instances of $\mathrm{CSP}(P)$ and instances with value at least α as a PAC learning problem by transforming each constraint into a labeled example. To show hardness of improperly learning a certain hypothesis class in the PAC model, they define a predicate P that is specific to the hypothesis class and assume hardness of distinguishing between random instances of $\mathrm{CSP}(P)$ and instances with n^d constraints and value at least α for all $d > 0$. They then demonstrate that the sample can be realized (or approximately realized) by some function in the hypothesis class if the CSP instance is satisfiable (or has value at least α). They also show that if the given CSP instance is random, the set of examples will have error at least $\frac{1}{4}$ (in the agnostic case $\frac{1}{5}$) for all h in the hypothesis class with high probability. Using this approach, they obtain hardness results for the following problems: improperly learning DNF formulas, improperly learning intersections of 4 halfspaces, and improperly approximately agnostically learning halfspaces for any approximation factor.

The hardness assumptions made in [33] are the same as those presented in Section II-A, except their model fixes the number of constraints rather than the probability with which each constraint is included in. It is well-known that results in one model easily translate to the other (see [4] for a proof). Additionally, SRCSP Assumptions 1 and 2 purport hardness of distinguishing random instances of CSP(P) from satisfiable instances, even when the algorithm is allowed to err with probability $\frac{1}{4}$ over its internal coins. The algorithms presented in the preceding sections never err on satisfiable instances; further, they only fail to certify random instances with probability $o(1)$. As a result, our refutation algorithms also falsify weaker versions of both SRCSP Assumptions, wherein the allowed probability of error is both lower and one-sided. For each predicate P presented in [33] and a corresponding $\delta > 0$, we define a degree-t polynomial that δ-separates P. Using the preceding sections, we deduce that $\widetilde{O}(n^{t/2})$ constraints are sufficient to distinguish random instances of CSP(P) from those that are satisfiable (or have value at least α). In order to simplify the presentation, we begin with simpler versions of the polynomials and then scale them to attain the appropriate values of δ. The following lemma will be of use for this scaling.

Lemma 35. *For predicate $P : \{-1,1\}^k \to \{0,1\}$, let $Q : \{-1,1\}^k \to \mathbb{R}$ be an unbiased multilinear polynomial of degree t such that there exist $\theta_1 > 0, \theta_0 < 0$ not dependent on z for which the following holds: $Q(z) \geq \theta_1$ for all $z \in P^{-1}(1)$ and $Q(z) \geq \theta_0$ for all $z \in \{-1,1\}^k$. Then there exists a degree-t polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that $\frac{\theta_1}{\theta_1 - \theta_0}$-separates P.*

Proof: Define $\mathcal{Q}(z) = \frac{Q(z)}{\theta_1 - \theta_0}$. Clearly $\mathcal{Q}$ is also unbiased and has degree t. Then for all $z \in P_1$, $\frac{Q(z)}{\theta_1 - \theta_0} \geq \frac{\theta_1}{\theta_1 - \theta_0}$. Similarly, for all z, $\frac{Q(z)}{\theta_1 - \theta_0} \geq \frac{\theta_0}{\theta_1 - \theta_0} = -\frac{\theta_1 - \theta_0}{\theta_1 - \theta_0} + \frac{\theta_1}{\theta_1 - \theta_0} = -1 + \frac{\theta_1}{\theta_1 - \theta_0}$. ■

We now demonstrate that the above can be applied to the predicates suggested in [33] by defining separating polynomials and applying Theorem 31.

A. Huang's predicate and hardness of learning DNF formulas

In order to obtain hardness of improperly learning DNF formulas with $\omega(1)$ terms, Daniely et al. use the following predicate, introduced by Huang [56]. Huang showed that it is hereditarily approximation resistant; Daniely et al. also observed that its 0-variability is $\Omega(k^{1/3})$ [33].

Definition 36. Let $k = \kappa + \binom{\kappa}{3}$ for some integer $\kappa \geq 3$. For $z \in \{-1,1\}^k$, index z as follows. Label the first κ bits of z as $z_1, \ldots, z_\kappa$. The remaining $\binom{\kappa}{3}$ bits are indexed by unordered triples of integers between 1 and κ. Each $T \subseteq [\kappa]$ with $|T| = 3$ is associated with a distinct bit of the remaining $\binom{\kappa}{3}$ bits, which is indexed by z_T. We say that z *strongly satisfies* the Huang predicate iff for every $T = \{z_i, z_j, z_\ell\}$ such that z_i, z_j, z_ℓ are distinct elements of $[\kappa]$, $z_i z_j z_\ell = z_{\{i,j,\ell\}}$. Additionally, we say that z *satisfies* the Huang predicate iff there exists some $z' \in \{-1,1\}^k$ such that z has Hamming distance at most κ from z' and z' strongly satisfies the Huang predicate. Define $H_\kappa : \{-1,1\}^k \to \{0,1\}$ as follows: $H_\kappa(z) = 1$ if z satisfies the Huang predicate and $H_\kappa(z) = 0$ otherwise.

Daniely et al. reduce the problem of distinguishing between random instances of CSP(H_κ) with $2n^d$ constraints and satisfiable instances to the problem of improperly PAC learning the class of DNF formulas with $\omega(1)$ terms on a sample of $O(n^d)$ training examples with error $\epsilon = 1/5$ with probability at least $\frac{3}{4}$. Here we show that there exists a polynomial time algorithm that refutes random instances of CSP(H_κ) by demonstrating that H_k does not support a 4-wise uniform distribution and applying Theorem 31.

Theorem 37. *Assume $\kappa \geq 9$. There exists a degree-4 polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that $\frac{1}{8}$-separates H_κ. Consequently, H_κ is $\frac{1}{8}$-far from supporting a 4-wise uniform distribution.*

Proof: As a notational shorthand, write z_{abc} for $z_{\{i_a, i_b, i_c\}}$. Define $\zeta : [\kappa]^6 \times \{-1,1\}^k \to [-5,5]$ as follows:

$$\begin{aligned}\zeta(i_1,i_2,i_3,i_4,i_5,i_6,z) = {} & z_{126}z_{134}z_{235}z_{456} + z_{256}z_{146}z_{345}z_{123} + z_{136}z_{236}z_{145}z_{245} \\ & + z_{124}z_{234}z_{356}z_{156} + z_{125}z_{135}z_{346}z_{246}.\end{aligned} \tag{11}$$

Observe that for each monomial $z_{T_1}z_{T_2}z_{T_3}z_{T_4}$ of ζ, for every $j \in [6]$, $\sum_{i=1}^4 1_{\{T_i \ni j\}} = 2$. Further, for each $T \subseteq [6]$ with $|T| = 3$, z_T appears exactly once in ζ. Let $\mathcal{Z}_6$ be the set of all ordered 6-tuples of distinct elements of $[\kappa]$. For an ordered tuple I, we use $\in_()$ to denote membership in I.

Define $Q : \{-1,1\}^k \to \mathbb{R}$ as follows. Our final polynomial $\mathcal{Q}$ will be a scaled version of Q.

$$Q(z) = \operatorname*{avg}_{I \in \mathcal{Z}_6} \zeta(I, z).$$

Observe that Q does not depend on any of $z_{\{1\}}, \ldots z_{\{\kappa\}}$. By construction, Q contains no constant term, so $\widehat{Q}(\emptyset) = 0$. Clearly $Q(z) \geq -5$ for all z because (11) is always at least -5.

Now we lower bound the value of Q on all z that satisfy H_κ. We first show that for any z' that strongly satisfies the Huang predicate, $Q(z') = 5$, then bound $Q(z') - Q(z)$ for any z with Hamming distance at most κ from z'. By definition, for each z'_{T_i}, we have that $z'_{T_i} \prod_{j \in T_i} z'_j = 1$. So for each monomial of Q,

$$\begin{aligned}\frac{1}{|\mathcal{Z}_6|} z'_{T_1} z'_{T_2} z'_{T_3} z'_{T_4} &= \frac{1}{|\mathcal{Z}_6|} \prod_{i=1}^{4} \prod_{j \in T_i} z'_j \\ &= \frac{1}{|\mathcal{Z}_6|} \prod_{j \in T_1 \cup T_2 \cup T_3 \cup T_4} (z'_j)^2 = \frac{1}{|\mathcal{Z}_6|},\end{aligned}$$

where the last line follows from the fact that $\sum_{1=1}^{4} 1_{\{T_i \ni j\}} = 2$. Since there are $5 \cdot |\mathcal{Z}_6|$ monomials, their sum is 5.

Now we consider the case where z does not strongly satisfy the Huang Predicate, but $H_\kappa(z) = 1$. Any singleton index on which z and z' differ will not change the value of Q. Let $N = \{T : z_T \neq z'_T\}$. We lower bound Q by counting the number of monomials in which each z_T appears and

$$Q(z) \geq 5 - \frac{2}{|\mathcal{Z}_6|} \sum_{T \in N} \sum_{I \in \mathcal{Z}} 1_{\{\bigwedge_{z_i \in T} i \in_{()} I\}}.$$

For fixed T, the number of monomials containing the variables of z_T is

$$\sum_{I \in \mathcal{Z}} 1_{\{\bigwedge_{z_i \in T} i \in_{()} I\}} = 120(\kappa - 3)(\kappa - 4)(\kappa - 5)$$

because there are exactly 120 ways to permute the three indices of T in I and the remaining $\kappa - 3$ indices are permuted in the remaining 3 positions of I. So

$$Q(z) \geq 5 - \frac{240\kappa}{|\mathcal{Z}_6|}(\kappa - 3)(\kappa - 4)(\kappa - 5) = 5 - \frac{240}{(\kappa - 1)(\kappa - 2)}. \tag{12}$$

For $\kappa \geq 9$, (12) is at least $5 - \frac{30}{7}$. Applying Lemma 35, there exists $\mathcal{Q} : \{-1, 1\}^k \to \mathbb{R}$ that $\frac{1}{8}$-separates H_κ. ■

B. Hamming weight predicates

The remaining predicates we would like to examine are symmetric, meaning they are functions only of their Hamming weights. Again for each predicate P we present a multivariate polynomial that δ-separates P for some $0 < \delta \leq 1$. Each of these polynomials can also be written as a univariate polynomial on the Hamming weight of its input, which we will use to show that each of the following polynomials δ-separates its predicate for the appropriate value of δ. We give the construction below.

Definition 38. For $z \in \{-1, 1\}^k$ where $z = z_1, \ldots, z_k$, let $S_z = \sum_{i=1}^{k} z_i$ and call S_z the Hamming weight of z. For all odd k and any $\theta \in \{-k, -k + 2, \ldots, k - 2, k\}$, define the predicate $\mathrm{Thr}_k^\theta : \{-1, 1\}^k \to \{0, 1\}$ as follows:

$$\mathrm{Thr}_k^\theta(z) = \begin{cases} 1 & \text{if } S_z \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

For example, Maj_k is the same as Thr_k^1 and Thr_k^{-k} is the trivial predicate satisfied by all $z \in \{-1, 1\}^k$.

Because the multilinear separating polynomials we will use are symmetric, we present a transformation to an equivalent univariate polynomial on the Hamming weight of the original input.

Lemma 39. *Let $Q : \{-1, 1\}^k \to \mathbb{R}$ be of the following form for some $a, b, c, d \in \mathbb{R}$:*

$$Q(z) = a \sum_{\substack{T \subseteq [n] \\ |T| = 1}} z^T + b \sum_{\substack{T \subseteq [n] \\ |T| = 2}} z^T + c \sum_{\substack{T \subseteq [n] \\ |T| = 3}} z^T + d \sum_{\substack{T \subseteq [n] \\ |T| = 4}} z^T. \tag{13}$$

Define $Q_u : \mathbb{R} \to \mathbb{R}$ as follows:

$$Q_u(z) = \tfrac{d}{24} S_z^4 + \tfrac{c}{6} S_z^3 + \left(\tfrac{b}{2} + \tfrac{d}{3} - \tfrac{dk}{4}\right) S_z^2 + \left(a + \tfrac{c}{6} \cdot (-3k + 2)\right) S_z - \tfrac{bk}{2} + \tfrac{dk}{24}(3k - 6). \tag{14}$$

Then $Q(z) = Q_u(S_z)$ for all $z \in \{-1, 1\}^k$.

Proof: We can write (13) as follows:

$$Q(z) = a\mathscr{K}_1\left(\tfrac{k-S_z}{2};k\right) + b\mathscr{K}_2\left(\tfrac{k-S_z}{2};k\right) + c\mathscr{K}_3\left(\tfrac{k-S_z}{2};k\right) + d\mathscr{K}_4\left(\tfrac{k-S_z}{2};k\right), \tag{15}$$

where $\mathscr{K}_i(\nu;k) = \sum_{j=0}^{i}(-1)^j\binom{\nu}{i}\binom{k-\nu}{i-j}$ denotes the Krawtchouk polynomial of degree i [59], [60]. Substituting $\nu = \frac{k-S_z}{2}$ yields the expression in (14). ∎

As a consequence, by choosing values of a, b, c, and d, we can work with a univariate polynomial while ensuring that its multivariate analogue is unbiased and has degree at most 4 (degree 3 when $d = 0$).

1) Almost-Majority and hardness of learning intersections of halfspaces

Definition 40. Daniely et al. define the following predicate in order to show hardness of improperly learning intersections of four halfspaces.

$$I_{8k} = \left(\wedge_{i=0}^{3}\mathrm{Thr}_k^{-1}(z_{ki+1}\dots z_{ki+k})\right) \wedge \neg\left(\wedge_{i=4}^{7}\mathrm{Thr}_k^{-1}(z_{ki+1}\dots z_{ki+k})\right).$$

The reduction relies on the assumption that for all $d > 0$, it is hard to distinguish random instances of CSP(I_{8k}) with n^d constraints from satisfiable instances. Because the input variables to each instance of Thr_k^{-1} above are disjoint, it is sufficient to show that each of the first four groups of k variables cannot support a 3-wise uniform distribution and consequently neither can I_{8k}; therefore, from Theorem 31 we deduce that there exists an efficient algorithm that refutes random instances of CSP(I_{8k}) with $\widetilde{O}(n^{3/2})$ constraints with high probability.

Theorem 41. *Assume $k \geq 5$ and k is odd. There exist $\delta = \delta(k) > 0$ where δ is $\Omega(k^{-4})$ and a degree-3 multilinear polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that δ-separates Thr_k^{-1}. Consequently, Thr_k^{-1} is not 3-wise supporting.*

Proof: Let

$$Q(z) = (k^2-k-1)\sum_{\substack{T\subseteq[n]\\|T|=1}} z^T + (1-k)\sum_{\substack{T\subseteq[n]\\|T|=2}} z^T + (1+k)\sum_{\substack{T\subseteq[n]\\|T|=3}} z^T$$

and define $Q_u : \mathbb{R} \to \mathbb{R}$ as follows:

$$Q_u(s) = \tfrac{1+k}{6}s^3 + \left(\tfrac{1-k}{2}\right)s^2 + \left(\tfrac{3k^2-7k-4}{6}\right)s - \tfrac{(1-k)k}{2}.$$

By Lemma 39, for all $z \in \{-1,1\}^k$, $Q(z) = Q_u(S_z)$. We lower bound $Q_u(s)$ for $s \geq -1$ and for all $s \in [-k,k]$.

First we show that Q_u is monotonically increasing in s.

$$\frac{dQ_u}{ds} = \tfrac{k+1}{2}s^2 + (1-k)s + \tfrac{3k^2-7k-4}{6} = \tfrac{1}{6}\left[(k-4)\left(3(s-1)^2 + \tfrac{2}{3} + 3k\right) + 15\left(\left(s-\tfrac{3}{5}\right)^2 + \tfrac{53}{75}\right)\right],$$

which is evidently positive for $k \geq 5$.

Because Q is monotonically increasing in s, $Q_u(s) \geq Q_u(-k)$ for all $s \in [-k,k]$.

$$\begin{aligned} Q_u(-k) &= \left(\tfrac{-k-1}{6}\right)k^3 + \left(\tfrac{1-k}{2}\right)k^2 - \left(\tfrac{3k^2-7k-4}{6}\right)k - \tfrac{(1-k)k}{2} \\ &= -\tfrac{1}{6}\left[k(k-2)(k^2+9k+5) + 9k\right], \end{aligned} \tag{16}$$

which is clearly negative for $k \geq 5$. Now it just remains to lower-bound $Q_u(s)$ for $s \geq -1$. Again, since Q_u is monotonically increasing in s, we use the value $Q_u(-1)$:

$$Q_u(-1) = \tfrac{-k-1}{6} + \left(\tfrac{1-k}{2}\right) - \left(\tfrac{3k^2-7k-4}{6}\right) - \tfrac{(1-k)k}{2} = 1.$$

By applying Lemma 35, there exists an unbiased multilinear polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ of degree 3 that $\frac{6}{k^4+7k^3-13k^2-k+6}$-separates Thr_k^{-1}. ∎

C. Majority and hardness of approximately agnostically learning halfspaces

Daniely et al. show that approximate agnostic improper learning of halfspaces is hard for all approximation factors $\phi \geq 1$ based on the assumption that for all $d > 0$ and for sufficiently large odd k, it is hard to distinguish between random instances of CSP(Thr_k^1) with n^d constraints and instances with value at least $1 - \frac{1}{10\phi}$. This is based on the fact that $\max_{\mathcal{D}} \mathbf{E}_{z\sim\mathcal{D}}\left[\mathrm{Thr}_1(z)\right] = 1 - \frac{1}{k+1}$, where $\mathcal{D}$ is a pairwise independent distribution on $\{-1,1\}^k$, and applying SRCSP Assumption 2. Here we show that for odd $k \geq 25$, Thr_k^1 is 0.1-far from supporting a 4-wise

uniform distribution. The value 0.1 is not sharp, but is chosen as a compromise between a reasonably large value and a reasonably simple proof.

Theorem 42. *There exists a degree-4 multilinear polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that 0.1-separates Thr_k^1 for all odd $k \geq 25$.*

Proof: Let

$$Q(z) = \tfrac{8}{27\sqrt{k}} \sum_{\substack{T\subseteq[n]\\|T|=1}} z^T - \tfrac{5}{9k^{3/2}} \sum_{\substack{T\subseteq[n]\\|T|=3}} z^T + \tfrac{4}{3k^2} \sum_{\substack{T\subseteq[n]\\|T|=4}} z^T$$

and let

$$Q_u(s) = \tfrac{1}{54}\left[\tfrac{3}{k^2}s^4 - \tfrac{5}{k^{3/2}}s^3 + \left(-\tfrac{18}{k} + \tfrac{24}{k^2}\right)s^2 + \left(\tfrac{31}{\sqrt{k}} - \tfrac{10}{k^{3/2}}\right)s + 9 - \tfrac{18}{k}\right]. \tag{17}$$

Then by Lemma 39 and some algebra, for all $z \in \{-1,1\}^k$, $Q(z) = Q_u(S_z)$. To simplify Q, let $\sigma = sk^{-1/2}$. Then we can rewrite (17) as follows:

$$Q_u(s) = \tfrac{1}{54}\left[3\sigma^4 - 5\sigma^3 + \left(-18 + \tfrac{24}{k}\right)\sigma^2 + \left(31 - \tfrac{10}{k}\right)\sigma + 9 - \tfrac{18}{k}\right]. \tag{18}$$

First we lower-bound $Q_u(s)$ for all $\sigma \in \mathbb{R}$ using the following expression, which is equivalent to (18).

$$\begin{aligned} Q_u(s) &= \tfrac{1}{54}\left[3(\sigma + \tfrac{29}{18})^2(\sigma - \tfrac{22}{9})^2 + \tfrac{383}{108}\left(\sigma + \tfrac{1832}{1149}\right)^2 - \tfrac{38987378}{837621} + \tfrac{24}{k}\left((\sigma - \tfrac{5}{24})^2 - \tfrac{457}{576}\right)\right] \\ &> \tfrac{1}{54}\left[3(\sigma + \tfrac{29}{18})^2(\sigma - \tfrac{22}{9})^2 + \tfrac{383}{108}\left(\sigma + \tfrac{1832}{1149}\right)^2 - 47 + \tfrac{24}{k}\left(-\tfrac{457}{576}\right)\right] > -\tfrac{48}{54} = -\tfrac{8}{9}, \end{aligned}$$

where the last inequality follows from the fact that $k \geq 24$ and the first two terms are always nonnegative.

Next we lower-bound $Q_u(s)$ for $s > 0$.

$$\begin{aligned} Q_u(s) &= \tfrac{1}{54}\left[3\sigma^4 - 5\sigma^3 + \left(-18 + \tfrac{24}{k}\right)\sigma^2 + \left(31 - \tfrac{10}{k}\right)\sigma + 9 - \tfrac{18}{k}\right] \\ &= \tfrac{1}{54}\left[3(\sigma - \tfrac{1}{4})^2(\sigma - \tfrac{25}{12})^2 + \tfrac{41}{120}((\sigma - \tfrac{839}{410})^2 + \tfrac{21507}{1344800} + \tfrac{27}{4} + 9\sigma(\sigma - \tfrac{21}{10})^2\right] > \tfrac{1}{8}. \end{aligned}$$

Applying Lemma 35, there exists $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ such that $\mathcal{Q}$ has degree 4 and $\mathcal{Q}$ $\frac{9}{73}$-separates Thr_k^1. ∎

D. Predicates satisfied by strings with Hamming weight at least $-\Theta(\sqrt{k})$.

In light of the fact that the threshold based predicates above are not 4-wise supporting, one may attempt to find another threshold-based predicate. Here we show that a symmetric threshold predicate that is 4-wise supporting must be satisfied by all strings with Hamming weight at least $-\frac{\sqrt{k}}{2}$. Furthermore, there exists a symmetric threshold predicate that is 4-wise supporting with a threshold of $-\Theta(\sqrt{k})$ and we sketch its construction.

We also consider the predicate $\mathrm{Thr}_k^{-\frac{1}{2}\sqrt{k}}$. While it is not used in [33], we show that it does not support a 4-wise uniform distribution in the interest of obtaining a tighter bound for the Hamming weight above which an unbiased, symmetric predicate is not 4-wise supporting. The threshold of $-\frac{1}{2}\sqrt{k}$ is particularly interesting in that it asymptotically matches the threshold θ below which Thr_k^θ is 4-wise supporting. The proof is similar to that of Theorem 42 and a full version is given in [4].

Theorem 43. *Assume $k \geq 99$ and k is odd. Then there exists a degree-4 polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that $\frac{1}{225}$-separates $\mathrm{Thr}_k^{-\frac{1}{2}\sqrt{k}}$. Consequently, $\mathrm{Thr}_k^{-\frac{1}{2}\sqrt{k}}$ is $\frac{1}{255}$-far from 4-wise supporting.*

Claim 44. *Assume $k = 2^m - 1$ for some integer $m \geq 3$. Then there exists a 4-wise uniform distribution supported only on $z \in \{-1,1\}^k$ such that $S_z \geq 1 - 2\sqrt{k+1}$.*

Proof: Let $\mathcal{C}$ be a binary BCH code of length k with designed distance $2\iota + 1$ and let $\mathcal{C}^\perp$ be its dual. Then the uniform distribution on the codewords of $\mathcal{C}$ is 2ι-wise uniform [6], [62]; see also [8, Ch 16.2]. Let $c = c_1 \ldots c_k$ be a codeword of $\mathcal{C}^\perp$, where each $c_i \in \{-1,1\}$. The Carlitz-Uchiyama bound [62] states that for all $c \in \mathcal{C}^\perp$,

$$\sum_{i=1}^{k} \tfrac{1}{2}(1 - c_i) \leq \tfrac{k+1}{2} + (\iota - 1)\sqrt{k+1}.$$

Therefore,

$$S_c = k - 2\sum_{i=1}^{k} \tfrac{1}{2}(1 - c_i) \geq k - (k+1) - (2\iota - 2)\sqrt{k+1} = -1 - (2\iota - 2)\sqrt{k+1}.$$

Setting $\iota = 2$, we can obtain 4-wise uniformity on this distribution and each string in the support of the distribution has Hamming weight at least $-1 - 2\sqrt{k+1}$. ∎

VI. Directions for future work

It would be interesting to show analogous efficient refutation results for models of random $\mathrm{CSP}(P)$ in which literals are not used. This would allow for results on, say, refuting q-colorability for random k-uniform hypergraphs. For some predicates (e.g., monotone Boolean predicates), random CSP instances are trivially satisfiable when there are no literals. However for such predicates one could consider a "Goldreich [52]-style" model in which each constraint is randomly either P or $\neg P$ applied to k random variables. Additionally, it would be good to investigate whether our refutation algorithms can be extended from the purely random $\mathrm{CSP}(P)$ setting to the "smoothed"/"semi-random" setting of Feige [37], in which the m constraints scopes are worst-case and only the negation pattern for literals is random. Feige showed how to efficiently refute random 3-SAT instances with $m \geq \widetilde{O}(n^{3/2})$ constraints even in this model.

Another valuable open direction would be to shore up the known proof-complexity evidence suggesting that $\widetilde{\Theta}(n^{t/2})$ constraints might be necessary to refute random $\mathrm{CSP}(P)$ when P is not t-wise supporting. The natural question here is whether the SOS lower bound of [14] can be extended from non-pairwise uniform supporting and $m = O(n)$ constraints, to non-t-wise uniform supporting and $m = O(n^{t/2-\epsilon})$ constraints. (Of course, it would also be good to eliminate the pruning step from their random instances.) One might also investigate the more refined question of whether, for P that are δ-far from t-wise supporting, one can improve on δ-refutation when there are $m \geq \widetilde{O}(n^{t/2})$ constraints.

Followup work on the very interesting paper [38] of Feige, Kim, and Ofek also seems warranted. Recall that it gives a *nondeterministic* refutation algorithm for random 3-SAT when $m \geq O(n^{1.4})$ (as well as a subexponential-time deterministic algorithm). This raises the question of whether there exist polynomial-size refutations for random $\mathrm{CSP}(P)$ instances that are nevertheless hard to find efficiently.

Finally, we suggest trying to rehabilitate the hardness-of-learning results in [33], given our new knowledge about what random $\mathrm{CSP}(P)$ instances seem hard to refute. As mentioned, the followup work of Daniely and Shalev-Shwartz [34] shows hardness of PAC-learning DNFs with $\omega(\log n)$ terms based on the very reasonable assumption that refuting random k-SAT requires $n^{f(k)}$ constraints for some $f(k) = \omega(1)$. Subsequent work by Daniely [31] shows hardness of approximately PAC-learning halfspaces assuming that refuting random k-XOR is hard both when $m = n^{c\sqrt{k}\log k}$ and when k is polylogarithmic in n and $m = n^{ck}$ for some $c > 0$.

Acknowledgments

Supported by NSF grants CCF-0747250 and CCF-1116594. Some of this work performed while the second-named author was at the Boğaziçi University Computer Engineering Department, supported by Marie Curie International Incoming Fellowship project number 626373. The first and third named authors were partially supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1252522. The authors would like to thank Amin Coja–Oghlan for help with the literature, and Boaz Barak and Ankur Moitra for permission to reprint the proof of the strong k-XOR refutation result. The last author would like to thank Anupam Gupta for several helpful discussions.

References

[1] http://satcompetition.org/2014/certunsat.shtml.

[2] D. Achlioptas, *Handbook of Satisfiability*, ser. Frontiers in Artificial Intelligence and Applications. IOS Press, 2009, vol. 185, ch. 8: Random Satisfiability, pp. 243–268.

[3] M. Alekhnovich, "More on average case vs. approximation complexity," in *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp. 298–307.

[4] S. R. Allen, R. O'Donnell, and D. Witmer, "How to refute a random CSP," *CoRR*, vol. abs/1505.04383, 2015. [Online]. Available: http://arxiv.org/abs/1505.04383

[5] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie, "Testing k-wise and almost k-wise independence," in *Proceedings of the 39th ACM Symposium on Theory of Computing*, 2007, pp. 496–505.

[6] N. Alon, L. Babai, and A. Itai, "A fast and simple randomized parallel algorithm for the maximal independent set problem," *Journal of Algorithms*, vol. 7, no. 4, pp. 567–583, 1986.

[7] N. Alon, O. Goldreich, and Y. Mansour, "Almost k-wise independence versus k-wise independence," *Information Processing Letters*, vol. 88, no. 3, pp. 107–110, 2003.

[8] N. Alon and J. H. Spencer, *The probabilistic method*. John Wiley & Sons, 2004.

[9] B. Applebaum, "Cryptographic hardness of random local functions – survey," in *tcc13*, 2013, pp. 599–599.

[10] B. Applebaum, B. Barak, and A. Wigderson, "Public-key cryptography from different assumptions," in *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, 2010, pp. 171–180.

[11] P. Austrin and J. Håstad, "Randomly supported independence and resistance," in *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*. ACM, New York, 2009, pp. 483–492.

[12] ——, "On the usefulness of predicates," *Transactions on Computation Theory*, vol. 5, no. 1, p. 1, 2013.

[13] P. Austrin and E. Mossel, "Approximation resistant predicates from pairwise independence," *Computational Complexity*, vol. 18, no. 2, pp. 249–271, 2009.

[14] B. Barak, S. O. Chan, and P. Kothari, "Sum of Squares lower bounds from pairwise independence," in *stoc15*, 2015.

[15] B. Barak, G. Kindler, and D. Steurer, "On the optimality of relaxations for average-case and generalized constraint satisfaction problems," in *Proceedings of the 4th Annual Innovations in Theoretical Computer Science conference*, 2013.

[16] B. Barak and A. Moitra, "Tensor prediction, rademacher complexity and random 3-xor," *CoRR*, vol. abs/1501.06521, 2015. [Online]. Available: http://arxiv.org/abs/1501.06521

[17] B. Barak and D. Steurer, "Sum-of-squares proofs and the quest toward optimal algorithms," in *Proceedings of International Congress of Mathematicians*, 2014.

[18] P. Beame, R. Karp, T. Pitassi, and M. Saks, "On the complexity of unsatisfiability proofs for random k-CNF formulas," in *stoc98*, 1998, pp. 561–571.

[19] ——, "The efficiency of resolution and Davis-Putnam procedures," *SIAM J. Comput.*, vol. 31, no. 4, pp. 1048–1075, 2002. [Online]. Available: http://dx.doi.org/10.1137/S0097539700369156

[20] E. Ben-Sasson and Y. Bilu, "A gap in average proof complexity," *Electronic Colloquium on Computational Complexity (ECCC)*, no. 003, 2002. [Online]. Available: http://eccc.hpi-web.de/eccc-reports/2002/TR02-003/index.html

[21] E. Ben-Sasson and A. Wigderson, "Short proofs are narrow—resolution made simple," in *Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999)*. ACM, New York, 1999, pp. 517–526 (electronic). [Online]. Available: http://dx.doi.org/10.1145/301250.301392

[22] S. Benabbas, K. Georgiou, A. Magen, and M. Tulsiani, "SDP gaps from pairwise independence," *Theory of Computing*, vol. 8, no. 1, pp. 269–289, 2012.

[23] A. Bulatov, P. Jeavons, and A. Krokhin, "Classifying the complexity of constraints using finite algebras," *SIAM Journal on Computing*, vol. 34, no. 3, pp. 720–742, 2005.

[24] S. O. Chan, "Approximation resistance from pairwise independent subgroups," in *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, 2013, pp. 447–456.

[25] V. Chvátal and E. Szemerédi, "Many hard examples for resolution," *J. Assoc. Comput. Mach.*, vol. 35, no. 4, pp. 759–768, 1988. [Online]. Available: http://dx.doi.org/10.1145/48014.48016

[26] A. Coja-Oghlan, C. Cooper, and A. Frieze, "An efficient sparse regularity concept," *SIAM J. Discrete Math.*, vol. 23, no. 4, pp. 2000–2034, 2009/10. [Online]. Available: http://dx.doi.org/10.1137/080730160

[27] A. Coja-Oghlan, A. Goerdt, and A. Lanka, "Strong Refutation Heuristics for Random k-SAT," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, ser. Lecture Notes in Computer Science, K. Jansen, S. Khanna, J. D. Rolim, and D. Ron, Eds. Springer Berlin Heidelberg, 2004, vol. 3122, pp. 310–321. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-27821-4_28

[28] ——, "Strong refutation heuristics for random k-SAT," *Combin. Probab. Comput.*, vol. 16, no. 1, pp. 5–28, 2007. [Online]. Available: http://dx.doi.org/10.1017/S096354830600784X

[29] A. Coja-Oghlan, A. Goerdt, A. Lanka, and F. Schädlich, "Techniques from combinatorial approximation algorithms yield efficient algorithms for random $2k$-SAT," *Theoretical Computer Science*, vol. 329, no. 1, pp. 1–45, 2004.

[30] A. Crisanti, L. Leuzzi, and G. Parisi, "The 3-SAT problem with large number of clauses in the ∞-replica symmetry breaking scheme," *Journal of Physics A: Mathematical and General*, vol. 35, no. 3, p. 481, 2002.

[31] A. Daniely, "Complexity theoretic limitations on learning halfspaces," *CoRR*, vol. abs/1505.05800, 2015. [Online]. Available: http://arxiv.org/abs/1505.05800

[32] A. Daniely, N. Linial, and S. Shalev-Shwartz, "More data speeds up training time in learning halfspaces over sparse vectors," in *Advances in Neural Information Processing Systems*, 2013, vol. 26, pp. 145–153. [Online]. Available: http://papers.nips.cc/paper/4905-more-data-speeds-up-training-time-in-learning-halfspaces-over-sparse-vectors.pdf

[33] ——, "From average case complexity to improper learning complexity," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, 2014, pp. 441–448. [Online]. Available: http://doi.acm.org/10.1145/2591796.2591820

[34] A. Daniely and S. Shalev-Shwartz, "Complexity theoretic limitations on learning DNFs," *CoRR*, vol. abs/1404.3378, 2014.

[35] J. Ding, A. Sly, and N. Sun, "Proof of the satisfiability conjecture for large k," in *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, 2015.

[36] U. Feige, "Relations between average case complexity and approximation complexity," in *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 2002, pp. 543–543.

[37] ——, "Refuting smoothed 3CNF formulas," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 2007, pp. 407–417.

[38] U. Feige, J. H. Kim, and E. Ofek, "Witnesses for non-satisfiability of dense random 3CNF formulas," in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2006, pp. 497–508.

[39] U. Feige and E. Ofek, "Easily refutable subformulas of large random 3CNF formulas," in *Automata, languages and programming*, ser. Lecture Notes in Comput. Sci., vol. 3142. Springer, Berlin, 2004, pp. 519–530. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-27836-8_45

[40] ——, "Spectral techniques applied to sparse random graphs," *Random Structures & Algorithms*, vol. 27, no. 2, pp. 251–275, 2005.

[41] V. Feldman, W. Perkins, and S. Vempala, "Subsampled Power Iteration: a Unified Algorithm for Block Models and Planted CSP's," *CoRR*, vol. abs/1407.2774, 2014. [Online]. Available: http://arxiv.org/abs/1407.2774

[42] ——, "On the Complexity of Random Satisfiability Problems with Planted Solutions," in *Proceedings of the 47th ACM Symposium on Theory of Computing*, 2015.

[43] J. Franco and M. Paull, "Probabilistic analysis of the Davis Putnam procedure for solving the satisfiability problem," *Discrete Applied Mathematics*, vol. 5, no. 1, pp. 77–87, 1983.

[44] J. Friedman and A. Goerdt, "Recognizing more unsatisfiable random 3-SAT instances efficiently," in *Automata, languages and programming*, ser. Lecture Notes in Comput. Sci. Springer, Berlin, 2001, vol. 2076, pp. 310–321. [Online]. Available: http://dx.doi.org/10.1007/3-540-48224-5_26

[45] J. Friedman, A. Goerdt, and M. Krivelevich, "Recognizing more unsatisfiable random k-SAT instances efficiently," *SIAM J. Comput.*, vol. 35, no. 2, pp. 408–430, 2005. [Online]. Available: http://dx.doi.org/10.1137/S009753970444096X

[46] X. Fu, "On the complexity of proof systems," Ph.D. dissertation, University of Toronto, 1996.

[47] Z. Füredi and J. Komlós, "The eigenvalues of random symmetric matrices," *Combinatorica*, vol. 1, no. 3, pp. 233–241, 1981.

[48] A. Goerdt and T. Jurdziński, "Some results on random unsatisfiable k-Sat instances and approximation algorithms applied to random structures," in *Mathematical Foundations of Computer Science 2002*, ser. Lecture Notes in Comput. Sci. Springer, Berlin, 2002, vol. 2420, pp. 280–291. [Online]. Available: http://dx.doi.org/10.1007/3-540-45687-2_23

[49] ——, "Some results on random unsatisfiable k-Sat instances and approximation algorithms applied to random structures," *Combin. Probab. Comput.*, vol. 12, no. 3, pp. 245–267, 2003, combinatorics, probability and computing (Oberwolfach, 2001). [Online]. Available: http://dx.doi.org/10.1017/S0963548303005637

[50] A. Goerdt and M. Krivelevich, "Efficient recognition of random unsatisfiable k-sat instances by spectral methods," in *stacs01*, 2001, pp. 294–304.

[51] A. Goerdt and A. Lanka, "Recognizing more random unsatisfiable 3-SAT instances efficiently," *Electronic Notes in Discrete Mathematics*, vol. 16, pp. 21–46, 2003.

[52] O. Goldreich, "Candidate One-Way Functions Based on Expander Graphs," in *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 7, no. 90, 2000.

[53] ——, "Three XOR-Lemmas — an exposition," in *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, 2011, pp. 248–272.

[54] D. Grigoriev, "Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity," *Theoretical Computer Science*, vol. 259, no. 1-2, pp. 613–622, 2001.

[55] D. Grigoriev and N. Vorobjov, "Complexity of Null- and Positivstellensatz proofs," *Annals of Pure and Applied Logic*, vol. 113, no. 1, pp. 153–160, 2001.

[56] S. Huang, "Approximation resistance on satisfiable instances for predicates with few accepting inputs," in *Proceedings of the 45th ACM Symposium on Theory of Computing*. ACM, 2013, pp. 457–466. [Online]. Available: http://doi.acm.org/10.1145/2488608.2488666

[57] V. Kann, J. Lagergren, and A. Panconesi, "Approximability of Maximum Splitting of k-sets and some other APX-complete problems," *Inf. Proc. Lett.*, vol. 58, no. 3, pp. 105–110, 1996.

[58] S. Khot, "Ruling out PTAS for Graph Min-Bisection, Dense k-Subgraph, and Bipartite Clique," *SIAM Journal on Computing*, vol. 36, no. 4, pp. 1025–1071, 2006.

[59] I. Krasikov and S. Litsyn, "On integral zeros of Krawtchouk polynomials," *Journal of Combinatorial Theory, Series A*, vol. 74, no. 1, pp. 71–99, 1996.

[60] M. Krawtchouk, "Sur une généralisation des polynomes d'Hermite," *Comptes Rendus*, vol. 189, pp. 620–622, 1929.

[61] J. Lasserre, "Optimisation globale et théorie des moments," *Comptes Rendus de l'Académie des Sciences*, vol. 331, no. 11, pp. 929–934, 2000.

[62] F. J. MacWilliams and N. Sloane, *The theory of error-correcting codes*. North-Holland, 1977.

[63] R. Mori, O. Watanabe, and D. Witmer, "Improved SDP integrality gaps for random CSPs," 2015, in progress.

[64] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[65] R. O'Donnell and D. Witmer, "Goldreich's PRG: Evidence for near-optimal polynomial stretch," in *ccc14*, 2014, pp. 1–12.

[66] R. O'Donnell and Y. Zhou, "Approximability and proof complexity," in *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2013, pp. 1537–1556.

[67] P. Parrilo, "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization," Ph.D. dissertation, California Institute of Technology, 2000.

[68] P. Raghavendra, "Approximating NP-hard problems: efficient algorithms and their limits," Ph.D. dissertation, University of Washington, 2009.

[69] G. Schoenebeck, "Linear level Lasserre lower bounds for certain k-CSPs," in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008, pp. 593–602.

[70] M. Tulsiani, "CSP gaps and reductions in the Lasserre hierarchy," in *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, 2009, pp. 303–312.

[71] M. Tulsiani and P. Worah, "LS_+ lower bounds from pairwise independence," in *ccc13*, 2013, pp. 121–132.

[72] L. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, Nov. 1984. [Online]. Available: http://doi.acm.org/10.1145/1968.1972

[73] E. P. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *Ann. of Math. (2)*, vol. 62, pp. 548–564, 1955.