How to refute a random CSP

Sarah R. Allen[*]   Ryan O'Donnell[*]   David Witmer[*]

June 28, 2015

Abstract

Let P be a nontrivial k-ary predicate over a finite alphabet. Consider a random $\mathrm{CSP}(P)$ instance $\mathcal{I}$ over n variables with m constraints, each being P applied to k random literals. When $m \gg n$ the instance $\mathcal{I}$ will be unsatisfiable with high probability, and the natural associated algorithmic task is to find a *refutation* of $\mathcal{I}$ — i.e., a certificate of unsatisfiability. When P is the 3-ary Boolean OR predicate, this is the well studied problem of refuting random 3-SAT formulas; in this case, an efficient algorithm is known only when $m \gg n^{3/2}$. Understanding the density required for average-case refutation of other predicates is of importance for various areas of complexity, including cryptography, proof complexity, and learning theory. The main previously-known result is that for a general Boolean k-ary predicate P, having $m \gg n^{\lceil k/2 \rceil}$ random constraints suffices for efficient refutation.

In this work we give a general criterion for arbitrary k-ary predicates, one that often yields efficient refutation algorithms at much lower densities. Specifically, if P fails to support a t-wise independent (uniform) probability distribution ($2 \le t \le k$), then there is an efficient algorithm that refutes random $\mathrm{CSP}(P)$ instances $\mathcal{I}$ with high probability, provided $m \gg n^{t/2}$. Indeed, our algorithm will "somewhat strongly" refute $\mathcal{I}$, certifying $\mathrm{Opt}(\mathcal{I}) \le 1 - \Omega_k(1)$; if $t = k$ then we furthermore get the strongest possible refutation, certifying $\mathrm{Opt}(\mathcal{I}) \le \mathbf{E}[P] + o(1)$. This last result is new even in the context of random k-SAT.

Regarding the optimality of our $m \gg n^{t/2}$ density requirement, prior work on SDP hierarchies has given some evidence that efficient refutation of random $\mathrm{CSP}(P)$ may be impossible when $m \ll n^{t/2}$. Thus there is an indication our algorithm's dependence on m is optimal for every P, at least in the context of SDP hierarchies. Along these lines, we show that our refutation algorithm can be carried out by the $O(1)$-round SOS SDP hierarchy.

Finally, as an application of our result, we falsify the "SRCSP assumptions" used to show various hardness-of-learning results in the recent (STOC 2014) work of Daniely, Linial, and Shalev–Shwartz.

1 On refutation of random CSPs

Constraint satisfaction problems (CSPs) play a major role in computer science. There is a vast theory [BJK05] of how algebraic properties of a CSP predicate affect its worst-case satisfiability complexity; there is a similarly vast theory [Rag09] 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 [Ach09] 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 [BB02], inapproximability [Fei02], SAT-solvers [SAT], cryptography [ABW10], learning theory [DLSS14], statistical physics [CLP02], and complexity theory [BKS13].

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. (The precise details of the random model are inessential; see Section 3.1 for more information.) 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 [DSS15] for sufficiently large k.

There is a natural algorithmic task associated with 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 [CS88] 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 [Fu96] showed that polynomial-size Resolution refutations exist (whp) once $m \geq O(n^{k-1})$; Beame et al. [BKPS99] subsequently showed that such proofs could be found efficiently.[2] A breakthrough came in 2001, when Goerdt and Krivelevich [GK01] 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 [FG01] (see also [FGK05]) 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 1 for a more detailed history of results in this

---

[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.

[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 [GJ02] that one can obtain $n^{k/2+\epsilon}$ for odd k "along the lines of [FG01]". On one hand, this is true, as we'll see in this paper. On the other hand, no proof was provided in [GJ02], and we have not found the claim repeated in any paper subsequent to 2003.

area.) Improving the $n^{3/2}$ bound for 3-SAT is widely regarded as a major open problem [ABW10], with conjectures regarding its possibility going both ways [BB02, DLSS13]. See also the intriguing work of Feige, Kim, and Ofek [FKO06] showing that polynomial-size 3-SAT refutations *exist* whp once $m \geq O(n^{1.4})$.

Strong refutation. In a notable paper from 2002, Feige [Fei02] 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. [COGL04] 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 $m \geq O(n)$).

Beyond k-SAT. As in the algebraic and approximation theories of CSP, it's of significant interest to consider random instances of the $\mathrm{CSP}(P)$ 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 $\mathrm{CSP}(P)$ instances. This precise question is very relevant for work in cryptography based on the candidate OWFs and PRGs of Goldreich [Gol00]; see also [ABW10] and the survey of Applebaum [App13]. It has also proven essential for the recent exciting approach to hardness of learning due to Daniely, Linial, and Shalev-Shwartz [DLSS13, DLSS14, DSS14]. We discuss this learning connection and our results on it in more detail in Section 5.

The special case of random 3-XOR has proved particularly important: it is related to 3-SAT refutation through Feige's "3XOR Principle" (see [Fei02, FO05, FKO06]); it's the basis for cryptographic schemes due to Alekhnovich [Ale03] (and is related to the "Learning Parities with Noise" problem); it's used in the best known lower bounds for the SOS SDP hierarchy [Gri01, Sch08], which we discuss further in Section 6; and, Barak and Moitra [BM15] 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 $\mathrm{CSP}(P)$ instances. The main known result, following from the work Coja-Oghlan, Cooper, and Frieze [COCF10], was the following: For any Boolean k-ary predicate P, one can efficiently strongly refute random $\mathrm{CSP}(P)$ 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 [BM15] showed how to improve this bound to $m \geq \widetilde{O}(n^{k/2})$.[4]

---

[4]The present authors also obtained this result around the same time, but we credit the result to [BM15] 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	[Fu96]
k-SAT	$O(n^{k-1}/_{\log^{k-2}(n)})$	Refutation	Efficient	[BKPS99, BKPS02]
k-SAT	$\widetilde{O}(n^{\lceil k/2 \rceil})$	Refutation	Efficient	[GK01, FGK05]
3-SAT	$O(n^{3/2+\epsilon})$	Refutation	Efficient	[FG01, FGK05]
k-SAT	$O(n^{k/2+\epsilon})$	Refutation	Efficient	Claimed possible in [GJ02, GJ03]
Exactly-k_1-out-of-k-SAT	$\widetilde{O}(n)$	Refutation	Efficient	[BB02]
2-out-of-5-SAT	$O(n^{3/2+\epsilon})$	Refutation	Efficient	[GJ02, GJ03]
NAE-3-SAT	$O(n)$	$\Omega(1)$-Refutation	Efficient	[KLP96, GJ02, GJ03]
k-SAT	$O(n^{\lceil k/2 \rceil})$	Refutation	Efficient	[COGLS04, FO05]
3-SAT	$\widetilde{O}(n^{3/2})$	Refutation	Efficient	[GL03]
3-SAT	$\widetilde{O}(n^{3/2})$	Strong	Efficient	[COGL04, COGL07]
4-SAT	$O(n^2)$	Strong	Efficient	[COGL04, COGL07]
3-SAT	$O(n^{3/2})$	Refutation	Efficient	[FO04]
3-SAT	$O(n^{1.4})$	Refutation	Existential	[FKO06]
3-XOR	$O(n^{3/2})$	$1/n^{\Omega(1)}$-refutation	Efficient	Implicit in [FKO06]
3-SAT	$\Omega(n^{3/2})$	Refutation	Efficient	Claimed in [FKO06]
Boolean k-CSP	$O(n^{\lceil k/2 \rceil})$	Strong	Efficient	[COCF10]
k-XOR	$\widetilde{O}(n^{k/2})$	Strong	Efficient	[BM15] (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 1: Up to logarithmic factors on m, our work subsumes all previously known results.

2   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 definitions of the terminology we use is given in Section 3. We also mention that in Section B we will generalize all of our results to the case of larger alphabets; but we'll just 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. Before getting to it, 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 [COCF10]; 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 [BM15, Corollary 5.5 and Extensions].

Theorem 2.1. *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 2.1 follows ideas from [COGL07] 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 [Wig55, FK81]. Similar techniques, including the use of the trace method, date back to the 2001 Friedgman–Goerdt work [FG01] refuting random 3-SAT given $m = n^{3/2+\epsilon}$ constraints.

With Theorem 2.1 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 2.1 to prove:

Theorem 2.2. *There is an efficient algorithm that (whp) certifies that a random instance of* 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 CSP(P):

Theorem 2.3. *For any k-ary predicate P, there is an efficient algorithm that (whp) strongly refutes random* CSP(P) *instances when the number of constraints is at least* $\widetilde{O}(n^{k/2})$.

In particular, this theorem improves upon [COCF10] 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 CSP(P) instances can be refuted with many fewer constraints than $n^{k/2}$. In the following theorem, 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 2.4. *(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 [AM09] showed that such predicates are hereditarily approximation-resistant under the UGC. In addition, the largest t for which a predicate P supports a t-wise uniform distribution over satisfying assignments is closely related to the notion of distribution complexity studied by Feldman, Perkins, and Vempala [FPV15] in the context of planted random CSPs. Informally, the distribution complexity of a planted CSP is the largest t for which the distribution over inputs to predicates induced by the planted assignment is t-wise uniform. Despite this similarity, the algorithmic techniques used in the planted case [FPV14] do not seem to apply.

The idea behind the proof of Theorem 2.4 is that with $\widetilde{O}(n^{t/2})$ constraints we can use the algorithm of Theorem 2.2 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.

Example 2.5. To briefly illustrate the result, consider the Exactly-k-out-of-$2k$-SAT CSP, studied previously in [BB02,GJ03]. 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.

2.1 An application from learning theory

Recently, an exciting approach to proving hardness-of-learning results has been developed by Daniely, Linial, and Shalev-Shwartz [DLSS13, DLSS14, DSS14]. The most general results appear in [DLSS14]. 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 assumption made in [DSS14] appears to be plausible, our work unfortunately shows that the more general assumptions made in [DLSS14] are false.

Below we state the (admittedly strong) assumptions from [DLSS14] (up to some very minor technical details which are discussed and treated in Section 5). 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) don't support pairwise uniformity. However, our work shows that avoiding t-wise uniformity for unbounded t is also necessary.

SRCSP Assumption 1. *([DLSS14].) 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. *([DLSS14], generalizing the "RCSP Assumption" of [BKS13] to super-linearly many constraints.) For all $d \in \mathbb{N}$ there is a large enough C such that the following holds: If $P : \{-1,1\}^k \rightarrow \{0,1\}$ has $VAR_0(P) \geq C$ and is δ-far from 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 [DLSS14] 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 [DLSS14] 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 [Hua13]; they are hereditarily approximation resistant on satisfiable instances and have 0-variability $\Omega(\kappa)$. In [DLSS14] they are used in SRCSP Assumption 1 to deduce hardness of PAC-learning DNFs with $\omega(1)$ terms. However:

Theorem 2.6. *For all $\kappa \geq 9$, the predicate H_κ does not support a 4-wise uniform distribution.*

Thus by Theorem 2.4 we can efficiently refute random instances of $CSP(H_\kappa)$ with just $\tilde{O}(n^2)$ 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 [DLSS14] to be $\frac{1}{k+1}$-far from supporting a pairwise uniform distribution. In [DLSS14] these predicates are used in SRCSP Assumption 2 to deduce hardness of agnsotically learning Boolean halfspaces to within any constant factor. However:

Theorem 2.7. *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 2.4 then implies we can efficiently δ-refute random instances of $CSP(\text{Maj}_k)$ with $\tilde{O}(n^2)$ 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 [DLSS14] 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 3.16 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 [DLSS14], Daniely and Shalev-Shwartz [DSS14] 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 $2d$.

2.2 Evidence for the optimality of our results

It's natural to ask whether the $n^{t/2}$ dependence in our main Theorem 2.4 can be improved. As we don't expect to prove unconditional hardness results, we instead merely seek good evidence that refuting t-wise non-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 [GV01]; see, e.g., [OZ13, BS14] 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 [Las00, Par00], 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. In Section 6 of this paper we show the following:

Theorem 2.8. *All of our refutation algorithms for k-ary predicates can be extended to produce degree-$2k$ SOS proofs.*

We now return to the question of evidence for the optimality of constraint density used in our results. Dating back to Franco–Paull [FP83] and Chvátal–Szemerédi [CS88], 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 [BSW99], 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, Schoenbeck [Sch08] showed (using the expansion analysis of [BSW99]) 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 [Tul09, Cha13] for related larger-alphabet followups. These results how 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. In the context of inapproximability and SDP-hierarchy integrality gaps, this condition on P has been significantly studied in the case of $t = 2$. For P not supporting pairwise uniformity, it is known [BGMT12, TW13] 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 [BCK15], except that the CSP(P) instances are not quite uniformly random; they are "slightly pruned" random instances.

Now suppose P is a predicate that does not support a t-wise uniform distribution, where $t > 2$. The second and third authors recently essentially showed [OW14] 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 [OW14] are also in the "Goldreich [Gol00] style"; i.e., there are no literals, but the "right-hand sides" are random. However it is not hard to show the proofs in [OW14] go through in the standard random model of this paper.) Future work [MWW15] 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 [FPV15]. They show that their lower bounds against statistical algorithms for solving planted random CSPs also imply lower bounds against statistical algorithms for refuting uniformly random CSPs. Specifically, they show that for random instances of CSP(P) such that P supports a t-wise uniform distribution over satisfying assignments, any statistical algorithm using queries that take $\widetilde{O}(n^{t/2})$ possible values can only refute instances with at least $\widetilde{\Omega}(n^{t/2})$ constraints. As an application of this result, they also show that any convex program refuting an instance of CSP(P) must have dimension at least $\widetilde{\Omega}(n^{t/2})$.

2.3 Certifying independence number and chromatic number of random hypergraphs

Coja-Oghlan, Goerdt, and Lanka [COGL07] also use their CSP refutation techniques to certify that random 3- and 4-uniform hypergraphs have small independence number and large chromatic number. We extend these results to random k-uniform hypergraphs.

Theorem 2.9. *For a random k-uniform hypergraph H, there is a polynomial time algorithm certifying that the independence number of H is at most β with high probability when H has at least $\widetilde{O}\left(\frac{n^{5k/2}}{\beta^{2k}}\right)$ hyperedges.*

Theorem 2.10. *For a random k-uniform hypergraph H, there is a polynomial time algorithm certifying that the chromatic number of H is at least ξ with high probability when H has at least $\widetilde{O}\left(\xi^{2k}n^{k/2}\right)$ hyperedges.*

The proofs of these theorems follow the outline of [COGL07]. We show Theorem 2.9 using a slightly more general form of Theorem 2.1. Theorem 2.10 follows almost directly from Theorem 2.9 using the fact that every color class of a valid coloring is an independent set. Details are given in Appendix C.

3 Preliminaries and notation

3.1 Constraint satisfaction problems

We review some basic definitions and facts related to constraint satisfaction problems (CSPs). In this section we discuss only the Boolean domain, which we prefer to write as $\{-1, 1\}$ rather than $\{0, 1\}$. The straightforward extensions to larger domains appear in Section B. We will need the following notation: 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 3.1. Given a predicate $P : \{-1, 1\}^k \to \{0, 1\}$, an instance $\mathcal{I}$ of the CSP(P) problem over variables $x_1, \ldots, x_n$ 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 $\text{Opt}(\mathcal{I}) \geq \overline{P}$.[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 a symmetric function. We use $\overline{m}$ to denote the expected number of constraints, namely $2^k n^k p$. As noted in Fact 3.6 below, the number of constraints m in a draw from $\mathcal{F}_P(n,p)$ is very tightly concentrated around $\overline{m}$, and we often blur the distinction between these parameters. Appendix D explicitly describes a method for simulating an instance drawn from $\mathcal{F}_P(n,p)$ when the number of constraints is fixed.

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 [GL03] (under the name "discrepancy"), Khot [Kho06] ("quasi-randomness"), Austrin and Håstad [AH13] ("adaptive uselessness"), and Chan [Cha13] ("low correlation"), among other places. To define it, we first need to define the induced distribution of an instance and an assignment.

Definition 3.2. 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 3.3. 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 $d_{\text{TV}}(\mathcal{D}_{\mathcal{I},x}, U^k) \leq \epsilon$ 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 3.4. If $\mathcal{D}$ and $\mathcal{D}'$ are probability distributions on the same finite set A then $d_{\text{TV}}(\mathcal{D}, \mathcal{D}')$ denotes their *total variation distance*, $\frac{1}{2} \sum_{\alpha \in A} |\mathcal{D}(\alpha) - \mathcal{D}'(\alpha)|$. If $d_{\text{TV}}(\mathcal{D}, \mathcal{D}') \leq \epsilon$ we say that $\mathcal{D}$ and $\mathcal{D}'$ are ϵ-*close*. If $d_{\text{TV}}(\mathcal{D}, \mathcal{D}') \geq \epsilon$ 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 3.5. *If $\mathcal{I}$ is ϵ-quasirandom, then $\text{Opt}(\mathcal{I}) \leq \overline{P} + \epsilon$ (and in fact, $|\text{Opt}(\mathcal{I}) - \overline{P}| \leq \epsilon$).*

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

Fact 3.6. *Let $\mathcal{I} \sim \mathcal{F}_P(n,p)$. Then the following statements hold with high probability.*

1. $m = |\mathcal{I}| \in \overline{m} \cdot \left(1 \pm O\left(\sqrt{\frac{\log n}{\overline{m}}}\right)\right)$.

---

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

2. $\mathrm{Opt}(\mathcal{I}) \leq \overline{P} \cdot \left(1 + O\left(\sqrt{\frac{1}{\overline{P}} \cdot \frac{n}{m}}\right)\right)$.

3. $\mathcal{I}$ is $O\left(\sqrt{2^k \cdot \frac{n}{m}}\right)$-quasirandom.

3.2 Algorithms and refutations on random CSPs

Definition 3.7. 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

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

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 3.8. In Section 5 we will encounter a "two-sided error" variant of this definition. This is the slightly 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 just over the random coins of $\mathcal{A}$.

Remark 3.9. 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".

3.3 t-wise uniformity

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

Definition 3.10. 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 will also consider the more general notion of (ϵ, t)-wise uniformity. This is typically defined using *Fourier coefficients*:

Definition 3.11. 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 [O'D14]. 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 $x_S \in \mathbb{R}^{|S|}$).

Remark 3.12. It is a simple fact (and it follows from Lemma 3.13 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 [AAK+07] for some quantitative improvement):

Lemma 3.13. (Alon–Goldreich-Mansour [AGM03, 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*

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

In particular if $t = k$ we have the bound $2^k \cdot \epsilon$ (and this can also be improved [Gol11] to $2^{k/2-1} \cdot \epsilon$).

Finally, we make a crucial definition:

Definition 3.14. 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 $P^{-1}(0)$.

3.4 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 [AH09, Theorem 3.1]. Herein we will extend this fact by giving a dual characterization of being far from t-wise supporting.

Definition 3.15. 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 [AH09, Theorem 3.1]:

Lemma 3.16. *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 $z \in \{-1, 1\}^k$.

$$
\begin{array}{lll}
\text{minimize} & \displaystyle\sum_{z\in\{-1,1\}^k}(1-P(z))\mathcal{D}(z) & (1) \\[3ex]
\text{s.t.} & \displaystyle\sum_{z\in\{-1,1\}^k}\mathcal{D}(z)z^S = 2^k\cdot\widehat{\mathcal{D}}(S) = 0 \qquad \forall S\subseteq[k] \quad 0<|S|\le t & (2) \\[3ex]
& \displaystyle\sum_{z\in\{-1,1\}^k}\mathcal{D}(z) = 1 & (3) \\[3ex]
& \mathcal{D}(z)\ge 0 \qquad \forall z\in\{-1,1\}^k &
\end{array}
$$

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 (see Remark 3.12). 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 $0<|S|\le t$ as well as a variable ξ corresponding to constraint (3).

$$
\begin{array}{lll}
\text{maximize} & \xi & (4) \\[3ex]
\text{s.t.} & \displaystyle\sum_{\substack{S\subseteq[k]\\0<|S|\le t}} c(S)z^S \le 1-P(z)-\xi \qquad \forall z\in\{-1,1\}^k. & (5)
\end{array}
$$

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. $\qquad\square$

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 3.17. *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 3.16, so $K\le k^t+1$ in general, with $K\le 2^k$ 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. $\qquad\square$

4 Quasirandomness and its implications for refutation

4.1 Strong refutation of k-XOR

In this section, we state our result on strong refutation of random k-XOR instances with $m = \widetilde{O}\left(n^{k/2}\right)$ constraints. (Recall that essentially this result was very recently obtained by Barak

and Moitra [BM15].) We actually have a slightly more general result, allowing variables and coefficients to take values in $[-1, 1]$ and not just in $\{-1, 1\}$. We will use this additional freedom to prove refutation results for CSPs over larger alphabets in Appendix B and refutation results for independence number and chromatic number of random hypergraphs in Appendix C.

Theorem 4.1. *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.

We give the proof in Appendix A. It follows techniques from [COGL07] fairly closely and is essentially the same as the proof of [BM15]. We will use this theorem to prove our results in subsequent sections.

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

Corollary 4.2. *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 4.1, 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 $\overline{m} = \tilde{O}_k(n^{k/2})$.

13

4.2 Quasirandomness and strong refutation of any k-CSP

Next, we will use the algorithm of Theorem 4.1 to certify that an instance of $\mathrm{CSP}(P)$ is quasirandom. This will immediately give us a strong refutation algorithm.

In order to certify quasirandomness, Lemma 3.13 implies that it suffices to certify each Fourier coefficient of $D_{\mathcal{I},x}$ has small magnitude.

Lemma 4.3. *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 4.4. *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 4.4.

Proof of Lemma 4.3. Without loss of generality, assume $1 \in S$. Applying definitions, we see that

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

(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 4.4. 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 $-w_S(T, c')$. $\qquad\square$

The existence of an algorithm for certifying quasirandomness follows from Lemmas 3.13 and 4.3.

Theorem 4.5. *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 4.6. *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 $\overline{m} \geq \frac{2^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.*

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

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

Definition 4.7. An instance $\mathcal{I}$ of 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 3.6 shows that random instances with $\widetilde{O}(n)$ constraints are $(o(1), t)$-quasirandom for all $t \leq k$ with high probability. Lemma 4.3 directly implies that we can certify (ϵ, t)-quasirandomness when $m \geq \widetilde{O}(n^{t/2})$.

Theorem 4.8. *There is an efficient algorithm that certifies that an instance* $\mathcal{I} \sim \mathcal{F}_P(n,p)$ *of* 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 CSP(P).

Theorem 4.9. *Let P be δ-far from t-wise supporting. Then there is an efficient algorithm that, given an instance* $\mathcal{I} \sim \mathcal{F}_P(n,p)$ *of* CSP(P)*, certifies that* $\mathrm{Opt}(\mathcal{I}) \leq 1 - \delta + \gamma$ *with high probability 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 3.13.

Proof 1. Run the algorithm of Theorem 4.8 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 3.13 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. $\square$

Proof 2 gives a slightly weaker version of Theorem 4.9, 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 [AGM03]'s connection between (ϵ, t)-wise uniformity and ϵ-closeness to a t-wise uniform distribution. We were able to convert Proof 2 into an SOS proof (see Section 6.4), but we did not see how to translate Proof 1 into an SOS version. Proof 2 requires Plancherel's Theorem, a fundamental result in Fourier analysis.

Theorem 4.10 (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 $x \in \{-1, 1\}^n$,

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

15

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

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

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 4.3 to certify that $\left|\widehat{D_{\mathcal{I},x}}(S)\right| \leq \frac{\gamma}{2^{2k}}$ for all S. $\qquad\square$

With Corollary 3.17, Theorem 4.9 implies that we can $\Omega_k(1)$-refute instances of $\mathrm{CSP}(P)$ with $\tilde{O}_k(n^{t/2})$ constraints when P is not t-wise supporting.

Corollary 4.11. *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 $\mathrm{Opt}(\mathcal{I}) \leq 1 - 2^{-\tilde{O}(k^t)}$ with high probability when $\overline{m} \geq 2^{\tilde{O}(k^t)}n^{t/2}\log^5 n$ and $t \geq 2$.*

4.4 Proof of Lemma 4.4

Recall the statement of the lemma.

Lemma 4.4. *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 4.12 (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 4.4. 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 4.1 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.

16

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 4.1 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 can then divide the v_U's by $4s\sqrt{\tau p} \log n$ and apply Theorem 4.1. If $s = 1$, we get a bound of $4\sqrt{\tau p} \cdot n \log n$ in the same way as (10). $\qquad\square$

5 Hardness of learning implications

Recent work by Daniely et al. [DLSS14] 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 [Val84]. In this model, the learner is given m labeled training examples $(x_1, \ell(x_1)), \ldots, (x_m, \ell(x_m))$, where each $x_i \in \{-1, 1\}^n$, each $\ell(x_i) \in \{0, 1\}$, and the examples are drawn from some unknown distribution $\mathcal{D}$ on $\{-1, 1\}^n \times \{0, 1\}$. For some hypothesis class $\mathcal{H} \subseteq \{0, 1\}^{\{-1,1\}^n}$ we say that $\mathcal{D}$ can be *realized* by $\mathcal{H}$ if there exists some $h \in \mathcal{H}$ such that $\mathbf{Pr}_{(x, \ell(x)) \sim \mathcal{D}}[h(x) \neq \ell(x)] = 0$. In improper PAC learning, on an input of m training examples drawn from $\mathcal{D}$ such that $\mathcal{D}$ can be realized by some $h \in \mathcal{H}$, and an error parameter ϵ, the algorithm outputs some hypothesis function $f_h : \{-1, 1\}^n \to \{0, 1\}$ (not necessarily in $\mathcal{H}$) such that $\mathbf{Pr}_{(x, \ell(x)) \sim \mathcal{D}}[f_h(x) \neq \ell(x)] \leq \epsilon$. In improper *agnostic* PAC learning, the assumption that $\mathcal{D}$ can be realized by some $h \in \mathcal{H}$ is removed and the algorithm must output a hypothesis that performs almost as well as the best hypothesis in $\mathcal{H}$. More formally, the hypothesis f_h must satisfy the following: $\mathbf{Pr}_{(x, \ell(x)) \sim \mathcal{D}}[f_h(x) \neq \ell(x)] \leq \min_{h \in \mathcal{H}} \mathbf{Pr}_{(x, \ell(x)) \sim \mathcal{D}}[h(x) \neq \ell(x)] + \epsilon$. In improper approximate agnostic PAC learning, the learner is also given an approximation factor $a \geq 1$ and must output a hypothesis f_h such that $\mathbf{Pr}_{(x, \ell(x)) \sim \mathcal{D}}[f_h(x) \neq \ell(x)] \leq a \cdot \min_{h \in \mathcal{H}} \mathbf{Pr}_{(x, \ell(x)) \sim \mathcal{D}}[h(x) \neq \ell(x)] + \epsilon$.

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.

5.1 Hardness assumptions

The hardness assumptions made in [DLSS14] are the same as those presented in Section 2.1, except for a few minor differences. First, their model fixes the number of constraints rather than the probability with which each constraint is included in the instance. It is well-known that results in one model easily translate to the other. We include a proof in Appendix D for completeness.

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 presented in [DLSS14], we falsify the appropriate SRCSP assumption using the following approach. For each predicate P and corresponding $\delta > 0$, we define a degree-t polynomial that δ-separates P. Using the refutation techniques presented in the preceding sections, we deduce that $\tilde{O}(n^{t/2})$ 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 5.1. *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, $\frac{Q(z)}{\theta_1 - \theta_0} \geq \frac{\theta_0}{\theta_1 - \theta_0} = -\frac{\theta_1 - \theta_0}{\theta_1 - \theta_0} + \frac{\theta_1}{\theta_1 - \theta_0} = -1 + \frac{\theta_1}{\theta_1 - \theta_0}$. $\square$

We now demonstrate that the above can be applied to the predicates suggested in [DLSS14] by defining separating polynomials and applying Theorem 4.9

5.2 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 [Hua13]. Huang showed that it is hereditarily approximation resistant; Daniely et al. also observed that its 0-variability is $\Omega(k^{1/3})$ [DLSS14].

Definition 5.2. 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 $CSP(H_\kappa)$ 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 $CSP(H_\kappa)$ by demonstrating that H_k does not support a 4-wise uniform distribution and applying Theorem 4.9.

Theorem 5.3. *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:

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

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.

$$Q(z) = \underset{I \in \mathcal{Z}_6}{\text{avg}} \, \zeta(I, z).$$

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,

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

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

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

For $\kappa \geq 9$, (12) is at least $5 - \frac{30}{7}$. Applying Lemma 5.1, there exists $\mathcal{Q} : \{-1, 1\}^k \to \mathbb{R}$ that $\frac{1}{8}$-separates H_κ. $\qquad \square$

From this and the fact that $\overline{H_\kappa} = 2^{\tilde{O}(k^{1/3}) - k}$ (see [Hua13]), we obtain the following corollary.

19

Corollary 5.4. *For sufficiently large n and $k \geq 93$, there exists an efficient algorithm that refutes random instances of $\mathrm{CSP}(H_\kappa)$ with $\tilde{O}(n^2)$ constraints with high probability. This falsifies Assumption 2.1 in the case of the Huang predicate.*

Remark 5.5. If we instead choose to scale Q by a factor of $\frac{1}{5} \cdot \frac{\kappa^2 - 3\kappa + 2}{2\kappa^2 - 6\kappa - 44}$ rather than substituting $\kappa = 9$ into (12), we can achieve a better separation of $\delta = \frac{\kappa^2 - 3\kappa - 46}{2\kappa^2 - 6\kappa - 44}$. For $\kappa \geq 9$, this expression is strictly increasing and it approaches $\frac{1}{2}$ as κ grows.

5.3 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 $0 \leq \delta \leq 1$. 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 5.6. For $z \in \{-1, 1\}^k$ where $z = z_1, \ldots, z_k$, define $S_z = \sum_{i=1}^{k} z_i$ and call S_z the Hamming weight of z.

Note that this is analogous to the notion of a Hamming weight of a vector in $\{0, 1\}^k$, but differs in that it is not simply the count of the number of 1's. We define a general predicate that is satisfied when S_z is at least some fixed threshold value θ.

Definition 5.7. 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 5.8. *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:

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

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

Proof. We can write (13) as follows:

$$Q(z) = a\mathscr{K}_1 \left(\frac{k - S_z}{2}; k \right) + b\mathscr{K}_2 \left(\frac{k - S_z}{2}; k \right) + c\mathscr{K}_3 \left(\frac{k - S_z}{2}; k \right) + d\mathscr{K}_4 \left(\frac{k - S_z}{2}; k \right), \tag{14}$$

20

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 [Kra29, KL96]. Substituting $\nu = \frac{k-S_z}{2}$, yields the following expressions. In [KL96] the first three expressions are given explicitly and the fourth can be easily obtained by applying their recursive formula.

$$\mathscr{K}_1\left(\frac{k-S_z}{2}; k\right) = S_z, \qquad \mathscr{K}_3\left(\frac{k-S_z}{2}; k\right) = \frac{S_z^3 - (3k-2)S_z}{6},$$

$$\mathscr{K}_2\left(\frac{k-S_z}{2}; k\right) = \frac{S_z^2 - k}{2}, \qquad \mathscr{K}_4\left(\frac{k-S_z}{2}; k\right) = \frac{S_z^4 + (8-6k)S_z^2 + 3k^2 - 6k}{24}.$$

Finally, substituting these expressions into (14) and by some algebra,

$$Q(z) = \frac{d}{24}S_z^4 + \frac{c}{6}S_z^3 + \left(\frac{b}{2} + \frac{d}{3} - \frac{dk}{4}\right)S_z^2 + \left(a + \frac{c}{6}\cdot(-3k+2)\right)S_z - \frac{bk}{2} + \frac{dk}{24}(3k-6). \qquad (15)$$

$\square$

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 $d = 0$).

5.3.1 Almost-Majority and hardness of learning intersections of halfspaces

Definition 5.9. Daniely et al. define the following predicate in order to show hardness of improperly learning intersections of four halfspaces.

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

The reduction relies on the assumption that for all $d > 0$, it is hard to distinguish random instances of $\mathrm{CSP}(I_{8k})$ 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 4.9 we deduce that there exists an efficient algorithm that refutes random instances of $\mathrm{CSP}(I_{8k})$ with $\tilde{O}(n^{3/2})$ constraints with high probability. Daniely et al. define a pairwise uniform distribution supported on I_{8k} as well as a pairwise uniform distribution supported on Thr_k^{-1}, so $t = 3$ is optimal.

Theorem 5.10. *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} does not support a 3-wise uniform distribution.*

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:

$$Q_u(s) = \frac{1+k}{6}s^3 + \left(\frac{1-k}{2}\right)s^2 + \left(k^2 - k - 1 + \frac{1+k}{6}\cdot(-3k+2)\right)s - \frac{(1-k)k}{2}$$

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

21

Then by Lemma 5.8, for all $z \in \{-1, 1\}^k$, $Q(z) = Q_u(S_z)$. It therefore suffices to lower bound $Q_u(s)$ both when $s \geq -1$ and for all $s \in [-k, k]$.

First we show that Q_u is monotonically increasing in s.

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

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

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

$$Q_u(-k) = \frac{-k-1}{6}k^3 + \left(\frac{1-k}{2}\right)k^2 - \left(\frac{3k^2 - 7k - 4}{6}\right)k - \frac{(1-k)k}{2}$$
$$= -\frac{1}{6}\left[k^4 + 7k^3 - 13k^2 - k\right], \tag{16}$$
$$= -\frac{1}{6}\left[k(k-2)(k^2 + 9k + 5) + 9k\right], \tag{17}$$

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 $Q_u(-1)$:

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

By applying Lemma 5.1, 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}. $\qquad\square$

Because $\mathrm{VAR}_0(I_{8k})$ is evidently $\Omega(k)$ and $\overline{I_{8k}} < \frac{1}{7}$ for all $k \geq 5$, we have the following Corollary.

Corollary 5.11. *For odd $k \geq 5$ and sufficiently large n, there exists an efficient algorithm that distinguishes between random instances of $CSP(I_{8k})$ with $\tilde{O}(n^{3/2})$ constraints and satisfiable instances with high probability.*

Remark 5.12. Thr_3^{-1} is the same as 3-OR and Thr_5^{-1} is the same as is the same as 2-out-of-5-SAT, so this approach can be used to $\Omega_k(1)$-refute 3-SAT instances and 2-out-of-5-SAT instances with $\tilde{O}_k(n^{3/2})$ constraints, which improves upon the $O(n^{3/2+\epsilon})$ constraints required for refutation of 2-out-of-5-SAT in [GJ02, GJ03].

5.4 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 $CSP(\mathrm{Thr}_k^1)$ with n^d constraints and instances with value at least $1 - \frac{1}{10\phi}$. This is based on the fact that $\max_{\mathcal{D}} \mathbf{E}_{z \sim \mathcal{D}}[\mathrm{Thr}_1(z)] = 1 - \frac{1}{k+1}$, 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 5.13. *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

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

and let

$$Q_u(s) = \frac{1}{18k^2} s^4 - \frac{5}{54k^{3/2}} s^3 + \left(-\frac{1}{3k} + \frac{4}{9k^2}\right) s^2 + \left(\frac{31}{54\sqrt{k}} - \frac{5}{27k^{3/2}}\right) s + \frac{1}{6} - \frac{1}{3k} \tag{18}$$
$$= \frac{1}{54}\left[\frac{3}{k^2} s^4 - \frac{5}{k^{3/2}} s^3 + \left(-\frac{18}{k} + \frac{24}{k^2}\right) s^2 + \left(\frac{31}{\sqrt{k}} - \frac{10}{k^{3/2}}\right) s + 9 - \frac{18}{k}\right].$$

Then by Lemma 5.8, 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 (18) as follows:

$$Q_u(s) = \frac{1}{54}\left[3\sigma^4 - 5\sigma^3 + \left(-18 + \frac{24}{k}\right)\sigma^2 + \left(31 - \frac{10}{k}\right)\sigma + 9 - \frac{18}{k}\right]. \tag{19}$$

First we lower-bound $Q_u(s)$ for all $\sigma \in \mathbb{R}$ using the following expression, which is equivalent to (19) by some algebra.

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

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 $s > 0$.

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

Applying Lemma 5.1, 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. □

Corollary 5.14. *For sufficiently large n and k, there exists an efficient algorithm that distinguishes between random instances of $CSP(\mathrm{Thr}_k^1)$ with $\widetilde{O}(n^2)$ constraints and instances with value at least 0.9 with high probability.*

5.5 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 [DLSS14], 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.

Theorem 5.15. *Assume $k \geq 99$ and k is odd. Then there exists a degree-4 polynomial 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.*

Proof. Define $Q : \{-1,1\}^k \to \mathbb{R}$ and $\mathcal{Q}_u : \mathbb{R} \to \mathbb{R}$ as follows:

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

$$Q_u(s) = \tfrac{s^4}{3k^2} + \tfrac{s^3}{3k^{3/2}} + \left(-\tfrac{7}{4k} + \tfrac{8}{3k^2}\right)s^2 + \left(\tfrac{1}{2k^{1/2}} + \tfrac{2}{3k^{3/2}}\right)s + \tfrac{3}{4} - \tfrac{2}{k}$$

Again, for simplicity we set $\sigma = sk^{-1/2}$ and obtain the following expression:

$$Q_u(s) = \tfrac{1}{3}\sigma^4 + \tfrac{1}{3}\sigma^3 - \tfrac{7}{4}\sigma^2 + \tfrac{1}{2}\sigma + \tfrac{3}{4} + \tfrac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right). \tag{20}$$

Observe that for $k \geq 99$, $\frac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right) = \tfrac{2}{3k}\left(\left(2\sigma - \tfrac{1}{4}\right)^2 - \tfrac{49}{16}\right) > -\tfrac{1}{48}$. We now lower-bound the value of Q_u for $s \geq -\tfrac{1}{2}k^{1/2}$, or equivalently, $\sigma \geq -\tfrac{1}{2}$:

$$Q_u(s) = \tfrac{1}{3}\sigma^4 + \tfrac{1}{3}\sigma^3 - \tfrac{7}{4}\sigma^2 + \tfrac{1}{2}\sigma + \tfrac{3}{4} + \tfrac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right)$$
$$= \tfrac{1}{3}\left(\sigma - \tfrac{35}{29}\right)^2\left(\sigma + \tfrac{1}{2}\right)\left(\sigma + \tfrac{200}{69}\right) + \tfrac{61}{12006}\left(\sigma + \tfrac{1}{2}\right)\left(\left(\sigma + \tfrac{4631}{3538}\right)^2 + \tfrac{1526073}{12517444}\right) + \tfrac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right) + \tfrac{1}{24}.$$

The first two terms are clearly nonnegative when $\sigma \geq -\tfrac{1}{2}$, so

$$> \tfrac{1}{24} - \tfrac{1}{48} = \tfrac{1}{48}.$$

We also show that $Q_u(s) \geq -\tfrac{14}{3}$ for all $s \in \mathbb{R}$.

$$Q_u(s) = \tfrac{1}{3}\sigma^4 + \tfrac{1}{3}\sigma^3 - \tfrac{7}{4}\sigma^2 + \tfrac{1}{2}\sigma + \tfrac{3}{4} + \tfrac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right)$$
$$= \left(\sigma + \tfrac{19}{9}\right)^2\left(\sigma - \tfrac{29}{18}\right)^2 + \tfrac{211}{486}\left(\sigma + \tfrac{397}{211}\right)^2 + \tfrac{195823}{8306226} + \tfrac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right) - \tfrac{14}{3}$$
$$\geq \left(\sigma + \tfrac{19}{9}\right)^2\left(\sigma - \tfrac{29}{18}\right)^2 + \tfrac{211}{486}\left(\sigma + \tfrac{397}{211}\right)^2 + \tfrac{1079081}{365473944} - \tfrac{14}{3}.$$

The first three terms are always nonnegative, so $Q_u(s) \geq -\tfrac{14}{3}$.

Applying Lemma 5.1, $T_k^{-\frac{1}{2}\sqrt{k}}$ is $\frac{1}{255}$-far from supporting a 4-wise uniform distribution. $\qquad\square$

Now we demonstrate that there exists a 4-wise uniform distribution supported on $\mathrm{Thr}_k^{1-2\sqrt{k+1}}$ when $k = 2^m - 1$ for some integer $m \geq 3$.

Claim 5.16. *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 [ABI86, MS77]; see also [AS04, 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 [MS77, page 280] states that for all $c \in \mathcal{C}^{\perp}$,

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

24

Observe that the quantity $\frac{1}{2}(1 - c_i)$ simply maps c_i from $\{-1, 1\}$ to $\{0, 1\}$ so that we can write the bound to match the presentation in [MS77]. Therefore,

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

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 $-1 - 2\sqrt{k + 1}$. $\qquad\square$

Remark 5.17. In order to construct a 4-wise uniform distribution for any value of k, one could simply express k as a sum of powers of 2, construct separate distributions on disjoint variables as described above for each power of 2 (down to the minimum length for which we can achieve distance at least 5, after which point we use the uniform distribution, and obtain a 4-wise uniform distribution. The total Hamming weight of a vector supported by this distribution would then be at least $-O(\sqrt{k})$.

6 SOS refutation proofs

6.1 The SOS proof system

We first define the SOS proof system introduced in [GV01]. Call a polynomial $q \in \mathbb{R}[X_1, \ldots, X_n]$ sum-of-squares (SOS) if there exist $q_1, \ldots, q_\ell \in \mathbb{R}[X_1, \ldots, X_n]$ such that $q = q_1^2 + \cdots + q_\ell^2$.

Definition 6.1. Let $X = (X_1, \ldots, X_n)$ be indeterminates. Let $q_1, \ldots, q_m, r_1, \ldots, r_{m'} \in \mathbb{R}[X]$ and let $A = \{q_1 \geq 0, \ldots, q_m \geq 0\} \cup \{r_1 = 0, \ldots, r_{m'}\}$. There is a degree-$d$ SOS proof that A implies $s \geq 0$, written as

$$A \vdash_d s \geq 0,$$

if there exist SOS $u_0, u_1, \ldots, u_m \in \mathbb{R}[X]$ and $v_1, \ldots, v_{m'} \in \mathbb{R}[X]$ such that

$$
s = u_0 + \sum_{i=1}^{m} u_i q_i + \sum_{i=1}^{m'} v_i r_i
$$

with $\deg(u_0) \leq d$, $\deg(u_i q_i) \leq d$ for all $i \in [m]$, and $\deg(v_i r_i) \leq d$ for all $i \in [m']$. If it also holds that $u_0, u_1, \ldots, u_m = 0$, we will write $A \vdash_d s = 0$.

It is well-known that a degree-d SOS proof can be found using an SDP of size $n^{O(d)}$ if it exists [Sho87, Par00, Las00, Las01].

In this section, we will take the set A to be $\{x_i^2 = 1\}_{i \in [n]}$, enforcing that variables are ± 1-valued. We show that with high probability there exists a low-degree SOS proof that a polynomial representing the value of a CSP instance is close to its expectation.

For more information on the SOS proof system and its applications to approximation algorithms, see, e.g., [OZ13, Lau09].

6.2 SOS certification of quasirandomness

All of our SOS results rely on the following theorem, which is the SOS version of Theorem 4.1.

Theorem 6.2. *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 $T \in [n]^k$,*

$$\mathbf{E}[w(T)] = 0 \tag{21}$$

$$\mathbf{Pr}[w(T) \neq 0] \leq p \tag{22}$$

$$|w(T)| \leq 1. \tag{23}$$

Then, with high probability,

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_{2k} \sum_{T \in [n]^k} w(T) x^T \leq 2^{O(k)} \sqrt{p} n^{3k/4} \log^{3/2} n.$$

This theorem was essentially proven by Barak and Moitra [BM15]. We give a proof in Appendix A.3. We first use this theorem to show that an SOS version of Lemma 4.4 holds.

Lemma 6.3. *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, with high probability,*

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

Proof. We sketch the differences from the proof of Lemma 4.4 given in Section 4.4. For $s \geq 2$, the lemma follows by using Theorem 6.2 instead of Theorem 4.1. If $s = 1$, it suffices to show that

$$\{x_i^2 \leq 1\} \vdash_2 v(i) x_i \leq |v(i)|.$$

for any v since summing over all i as in (10) finishes the proof. If $v_i \geq 0$, observe that

$$|v_i| - v(i) x_i = \frac{|v(i)|}{2}(x_i - 1)^2 + \frac{|v(i)|}{2}(1 - x_i^2).$$

If $v(i) < 0$, we use $(x_i + 1)^2$ instead of $(x_i - 1)^2$. $\qquad\square$

The lemma implies an SOS version of Lemma 4.3. To make this precise, we define a specific polynomial representation of $\widehat{D_{\mathcal{I},x}}(S)$:

$$\widehat{D_{\mathcal{I},x}}(S)^{\text{poly}} = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} c^S x_T^S,$$

where $x_T^S = \prod_{i \in S} x_{T_i}$. Note that this is a polynomial in the x_i's.

We can show these polynomials are not too large.

Lemma 6.4. *Let $\emptyset \neq S \subseteq [k]$ with $|S| = s$. Then*

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_{2s} \widehat{D_{\mathcal{I},x}}(S)^{\text{poly}} \leq \frac{2^{O(k)} \max\{n^{s/4}, \sqrt{n}\} \log^{5/2} n}{\overline{m}^{1/2}}$$

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_{2s} \widehat{D_{\mathcal{I},x}}(S)^{\text{poly}} \geq -\frac{2^{O(k)} \max\{n^{s/4}, \sqrt{n}\} \log^{5/2} n}{\overline{m}^{1/2}}.$$

with high probability, assuming also that $\overline{m} \geq \max\{n^{s/2}, n\}$.

Proof. The proof is essentially the same as that of Lemma 4.3. The expression we bound in that proof is exactly $\widehat{D_{\mathcal{I},x}}(S)^{\mathrm{poly}}$. We use Lemma 6.3 instead of Lemma 4.4 to show that this can be done in degree-$2s$ SOS. $\qquad\square$

Based on Lemma 3.13, we will think of Lemma 6.4 as giving an SOS proof of quasirandomness. Below, we use it to prove SOS versions of Theorems 4.6 and 4.9.

6.3 Strong refutation of any k-CSP

We now define the natural polynomial representation of $\mathrm{Val}_{\mathcal{I}}(x)$:

$$\mathrm{Val}_{\mathcal{I}}^{\mathrm{poly}}(x) = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} \left(\sum_{S \subseteq [k]} \widehat{P}(S) c^S x_T^S \right),$$

where x_T^S is as above.

We can then give an SOS proof strongly refuting CSP(P).

Theorem 6.5. *Given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P),

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_{2k} \mathrm{Val}_{\mathcal{I}}^{\mathrm{poly}}(x) \leq \overline{P} + \gamma$$

with high probability when $\overline{m} \geq \frac{2^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.

Proof. By rearranging terms, we see that

$$\mathrm{Val}_{\mathcal{I}}^{\mathrm{poly}}(x) = \overline{P} + \sum_{\emptyset \neq S \subseteq [k]} \widehat{P}(S) \widehat{D_{\mathcal{I},x}}(S)^{\mathrm{poly}}.$$

Note that this is just Plancherel's Theorem in SOS. The theorem then follows from Lemma 6.4 and the observation that $\sum_{S \subseteq [k]} |\widehat{P}(S)| \leq 2^{O(k)}$. $\qquad\square$

6.4 $\Omega(1)$-refutation of non-t-wise supporting CSPs

Theorem 6.6. *Let P be δ-far from being t-wise supporting. Given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P),
$$\{x_i^2 = 1\}_{i \in [n]} \vdash_{\max\{k,2t\}} \mathrm{Val}_{\mathcal{I}}^{\mathrm{poly}}(x) \leq 1 - \delta + \gamma.$$
with high probability when $\overline{m} \geq \frac{2^{O(k)} n^{t/2} \log^5 n}{\gamma^2}$ and $t \geq 2$.

To prove this theorem, we will need to following claim, which says that any true inequality in k variables over $\{-1,1\}^k$ can be proved in degree-k SOS. Recall that the multilinearization of a monomial $z_1^{s_1} z_2^{s_2} \cdots z_k^{s_k} \in \mathbb{R}[z_1, \ldots, z_k]$ is defined to be $z_1^{s_1 \bmod 2} z_2^{s_2 \bmod 2} \cdots z_k^{s_k \bmod 2}$, i.e., we replace all z_i^2 factors by 1. We extend this definition to all polynomials in $\mathbb{R}[z_1, \ldots, z_k]$ by linearity.

Claim 6.7. *Let $f : \{-1,1\}^k \to \mathbb{R}$ such that $f(z) \geq 0$ for all $z \in \{-1,1\}^k$ and let f^{poly} be the unique multilinear polynomial such that $f(z) = f^{\mathrm{poly}}(z)$ for all $z \in \{-1,1\}^k$. Then*

$$\{z_i^2 = 1\}_{i \in [k]} \vdash_k f^{\mathrm{poly}}(z) \geq 0.$$

Proof. Since $f(x) \geq 0$, there exists a Boolean function $g : \{-1, 1\}^k \to \mathbb{R}$ such that $g(z)^2 = f(z)$ for all $z \in \{-1, 1\}^k$. Let g^{poly} be the unique multilinear polynomial such that $g(z) = g^{\mathrm{poly}}(z)$ for all $z \in \{-1, 1\}^k$. Since $g^{\mathrm{poly}}(z)^2 = f^{\mathrm{poly}}(z)$ for all $z \in \{-1, 1\}^k$, uniqueness of the multilinear polynomial representation of f implies that the multilinearization of $(g^{\mathrm{poly}})^2$ is equal to f^{poly}. Written another way, we have that $\{z_i^2 = 1\}_{i \in [k]} \vdash_k f^{\mathrm{poly}}(z) = g^{\mathrm{poly}}(z)^2$. This implies that $\{z_i^2 = 1\}_{i \in [k]} \vdash_k f^{\mathrm{poly}}(z) \geq 0$. $\qquad\square$

Proof of Theorem 6.6. The proof is an SOS version of Proof 2 of Theorem 4.9 above. Claim 6.7 implies that for Q of degree at most t that δ-separates P,

$$\{z_i^2 = 1\}_{i \in [k]} \vdash_k Q(z) - P(z) + 1 - \delta \geq 0.$$

Summing over all constraints, we get

$$\{x_i^2 = 1\}_{i \in [n]} \vdash_k \mathrm{Val}_{\mathcal{I}}^{\mathrm{poly}}(x) - (1 - \delta) \leq \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} \left(\sum_{S \subseteq [k]} \widehat{Q}(S) c^S x_T^S \right),$$

Rearranging terms as in the proof of Theorem 6.5, we see that the right hand side is equal to

$$\sum_{S \subseteq [k]} \widehat{Q}(S) \widehat{D_{\mathcal{I},x}}(S)^{\mathrm{poly}}.$$

Since Q has mean 0, $|Q| \leq 2^k$ and $\sum_{S \subseteq [k]} |\widehat{P}(S)| \leq 2^{O(k)}$. The theorem then follows from Lemma 6.4. $\qquad\square$

With Corollary 3.17, the theorem implies that we can $\Omega_k(1)$-refute any $\mathrm{CSP}(P)$ in SOS when P is not t-wise supporting.

Corollary 6.8. *Let P be a predicate that does not support any t-wise uniform distribution. Given an instance $\mathcal{I} \sim \mathcal{F}_P(n, p)$ of $\mathrm{CSP}(P)$,*

$$\{x_i^2 = 1\}_{i \in [n]} \vdash_{\max\{k, 2t\}} \mathrm{Val}_{\mathcal{I}}(x) \leq 1 - 2^{-\widetilde{O}(k^t)} + \gamma$$

with high probability when $\overline{m} \geq 2^{\widetilde{O}(k^t)} n^{t/2} \log^5 n$ and $t \geq 2$.

7 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. We give a simple result on refuting q-colorability of random hypergraphs in Appendix C, but it follows from refutation of binary CSPs and perhaps a stronger result could be proven by studying CSPs with larger alphabets. 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 [Gol00]-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 [Fei07], 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 $m \geq \tilde{O}(n^{3/2})$ constraints even in this model.

Another valuable open direction would be to shore up the known proof-complexity evidence suggesting that $\tilde{\Theta}(n^{t/2})$ 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 [BCK15] 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 $m \geq \tilde{O}(n^{t/2})$ constraints.

Followup work on the very interesting paper [FKO06] 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 [DLSS14], 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 [DSS14] 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 $f(k) = \omega(1)$. It is a good question as to whether their methods can still obtain hardness-of-learning results for, say, halfspace-type problems, based on random CSP assumptions consistent with our theorems. A concrete direction along these lines would involve determining whether or not there are Boolean halfspaces that are $o(1)$-close to supporting 4-wise (or higher) uniformity.

Acknowledgments

References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 496–505, 2007. 3.3

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986. 5.5

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 171–180, 2010. 1

[Ach09] Dimitris Achlioptas. *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*, chapter 8: Random Satisfiability, pages 243–268. IOS Press, 2009. 1

[AGM03]   Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Inform. Process. Lett.*, 88(3):107–110, 2003. 3.13, 4.3

[AH09]   Per Austrin and Johan Håstad. Randomly supported independence and resistance. In *STOC'09—Proceedings of the 2009 ACM International Symposium on Theory of Computing*, pages 483–492. ACM, New York, 2009. 3.4, 3.4

[AH13]   Per Austrin and Johan Håstad. On the usefulness of predicates. *Transactions on Computation Theory*, 5(1):1, 2013. 3.1

[Ale03]   M. Alekhnovich. More on average case vs approximation complexity. In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 298–307, 2003. 1

[AM09]   Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009. 2

[App13]   Benny Applebaum. Personal Communication, 2013. 1

[AS04]   Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004. 5.5

[Aus08]   Per Austrin. *Conditional Inapproximability and Limited Independence*. PhD thesis, KTH - Royal Institute of Technology, 2008. B.1, B.1, B.1

[BB02]   Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, (003), 2002. 1, 2.5

[BCK15]   Boaz Barak, Siu On Chan, and Pravesh Kothari. Sum of Squares Lower Bounds from Pairwise Independence. *ArXiv e-prints*, January 2015. 2.2, 7

[BGMT12]   Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8:269–289, 2012. 2.2

[BJK05]   Andrei Bulatov, Peter Jeavons, and Andrei Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM Journal on Computing*, 34(3):720–742, 2005. 1

[BKPS99]   Paul Beame, Richard Karp, Toniann Pitassi, and Michael Saks. On the complexity of unsatisfiability proofs for random k-CNF formulas. In *Proceedings of the 30th ACM Symposium on Theory of Computing*, pages 561–571. ACM, New York, 1999. 1

[BKPS02]   Paul Beame, Richard Karp, Toniann Pitassi, and Michael Saks. The efficiency of resolution and Davis-Putnam procedures. *SIAM J. Comput.*, 31(4):1048–1075, 2002. 1

[BKS13]   Boaz Barak, Guy Kindler, and David 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. 1, 2.1

[BM15]   Boaz Barak and Ankur Moitra. Tensor prediction, rademacher complexity and random 3-xor. *CoRR*, abs/1501.06521, 2015. 1, 4, 2, 4.1, 4.1, 6.2, A.4

[BS14]   Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of International Congress of Mathematicians*, 2014. 2.2

[BSW99]   Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. In *Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999)*, pages 517–526 (electronic). ACM, New York, 1999. 2.2

[Cha13]   Siu On Chan. Approximation resistance from pairwise independent subgroups. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 447–456, 2013. 2.2, 3.1

[CLP02]   Andrea Crisanti, Luca Leuzzi, and Giorgio Parisi. The 3-SAT problem with large number of clauses in the ∞-replica symmetry breaking scheme. *Journal of Physics A: Mathematical and General*, 35(3):481, 2002. 1

[COCF10]   Amin Coja-Oghlan, Colin Cooper, and Alan Frieze. An efficient sparse regularity concept. *SIAM J. Discrete Math.*, 23(4):2000–2034, 2009/10. 1, 2, 2

[COGL04]   Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong Refutation Heuristics for Random k-SAT. In Klaus Jansen, Sanjeev Khanna, José D.P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 3122 of *Lecture Notes in Computer Science*, pages 310–321. Springer Berlin Heidelberg, 2004. 1, A.4

[COGL07]   Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong refutation heuristics for random k-SAT. *Combin. Probab. Comput.*, 16(1):5–28, 2007. 1, 2, 2.3, 2.3, 4.1, C, C.1, C.2, C.3, C

[COGLS04] Amin Coja-Oghlan, Andreas Goerdt, André Lanka, and Frank Schädlich. Techniques from combinatorial approximation algorithms yield efficient algorithms for random $2k$-SAT. *Theoretical Computer Science*, 329(1):1–45, 2004. 1

[CS88]   Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *J. Assoc. Comput. Mach.*, 35(4):759–768, 1988. 1, 2.2

[CW04]   Moses Charikar and Anthony Wirth. Maximizing quadratic programs: extending Grothendieck's inequality. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60, 2004. A.1, A.1

[DLSS13]   Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. More data speeds up training time in learning halfspaces over sparse vectors. In *Advances in Neural Information Processing Systems*, volume 26, pages 145–153. 2013. 1, 2.1

[DLSS14]   Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 441–448, 2014. 1, 2.1, 2.1, 2.1, 5, 5.1, 5.1, 5.2, 5.5, 7, D

[DSS14]   Amit Daniely and Shai Shalev-Shwartz. Complexity theoretic limitations on learning DNFs. *CoRR*, abs/1404.3378, 2014. 1, 2.1, 2.1, 7

[DSS15]   Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, 2015. 1

[Fei02]   Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 534–543, 2002. 1

[Fei07]   Uriel Feige. Refuting smoothed 3CNF formulas. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 407–417, 2007. 7

[FG01]   Joel Friedman and Andreas Goerdt. Recognizing more unsatisfiable random 3-SAT instances efficiently. In *Automata, languages and programming*, volume 2076 of *Lecture Notes in Comput. Sci.*, pages 310–321. Springer, Berlin, 2001. 1, 3, 2

[FGK05]   Joel Friedman, Andreas Goerdt, and Michael Krivelevich. Recognizing more unsatisfiable random k-SAT instances efficiently. *SIAM J. Comput.*, 35(2):408–430, 2005. 1

[FK81]   Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981. 2

[FKO06]   Uriel Feige, Jeong Han Kim, and Eran Ofek. Witnesses for non-satisfiability of dense random 3CNF formulas. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 497–508. IEEE, 2006. 1, 7

[FO04]   Uriel Feige and Eran Ofek. Easily refutable subformulas of large random 3CNF formulas. In *Automata, languages and programming*, volume 3142 of *Lecture Notes in Comput. Sci.*, pages 519–530. Springer, Berlin, 2004. 1

[FO05]   Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005. 1

[FP83]   John Franco and Marvin Paull. Probabilistic analysis of the Davis Putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematics*, 5(1):77–87, 1983. 2.2

[FPV14]   Vitaly Feldman, Will Perkins, and Santosh Vempala. Subsampled Power Iteration: a Unified Algorithm for Block Models and Planted CSP's. 07 2014. 2

[FPV15]   Vitaly Feldman, Will Perkins, and Santosh Vempala. On the complexity of random satisfiability problems with planted solutions. In *Proceedings of the 47th ACM Symposium on Theory of Computing*, 2015. 2, 2.2

[Fu96]   Xudong Fu. *On the complexity of proof systems*. PhD thesis, University of Toronto, 1996. 1

[GJ02]   Andreas Goerdt and Tomasz Jurdziński. Some results on random unsatisfiable k-Sat instances and approximation algorithms applied to random structures. In *Mathematical Foundations of Computer Science 2002*, volume 2420 of *Lecture Notes in Comput. Sci.*, pages 280–291. Springer, Berlin, 2002. 3, 1, 5.12

[GJ03]   Andreas Goerdt and Tomasz Jurdziński. Some results on random unsatisfiable k-Sat instances and approximation algorithms applied to random structures. *Combin. Probab. Comput.*, 12(3):245–267, 2003. Combinatorics, probability and computing (Oberwolfach, 2001). 1, 2.5, 5.12

[GK01]   Andreas Goerdt and Michael Krivelevich. Efficient recognition of random unsatisfiable k-SAT instances by spectral methods. In *STACS 2001 (Dresden)*, volume 2010 of *Lecture Notes in Comput. Sci.*, pages 294–304. Springer, Berlin, 2001. 1

[GL03]   Andreas Goerdt and André Lanka. Recognizing more random unsatisfiable 3-SAT instances efficiently. *Electronic Notes in Discrete Mathematics*, 16:21–46, 2003. 1, 3.1

[Gol00]   Oded Goldreich. Candidate One-Way Functions Based on Expander Graphs. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 7, 2000. 1, 2.2, 7

[Gol11]   Oded Goldreich. Three XOR-lemmas—an exposition. In *Studies in complexity and cryptography*, volume 6650 of *Lecture Notes in Comput. Sci.*, pages 248–272. Springer, Heidelberg, 2011. 3.13, B.3

[Gri01]   Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613 – 622, 2001. 1

[GV01]   Dima Grigoriev and Nicolai Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1–3):153 – 160, 2001. First St. Petersburg Conference on Days of Logic and Computability. 2.2, 6.1

[Hua13]   Sangxia Huang. Approximation resistance on satisfiable instances for predicates with few accepting inputs. In *Proceedings of the 45th ACM Symposium on Theory of Computing*, pages 457–466. ACM, 2013. 2.1, 5.2, 5.2

[Kho06]   Subhash Khot. Ruling out PTAS for Graph Min-Bisection, Dense k-Subgraph, and Bipartite Clique. *SIAM Journal on Computing*, 36(4):1025–1071, 2006. 3.1

[KL96]   Ilia Krasikov and Simon Litsyn. On integral zeros of Krawtchouk polynomials. *Journal of Combinatorial Theory, Series A*, 74(1):71–99, 1996. 5.3

[KLP96]   Viggo Kann, Jens Lagergren, and Alessandro Panconesi. Approximability of Maximum Splitting of k-sets and some other APX-complete problems. *Inf. Proc. Lett.*, 58(3):105–110, 1996. 1

[Kra29]   Mikhail Krawtchouk. Sur une généralisation des polynomes d'Hermite. *Comptes Rendus*, 189:620–622, 1929. 5.3

[Las00]   Jean B. Lasserre. Optimisation globale et théorie des moments. *C. R. Acad. Sci. Paris Sér. I Math.*, 331(11):929–934, 2000. 2.2, 6.1

[Las01]   Jean B. Lasserre. Global Optimization with Polynomials and the Problem of Moments. *SIAM J. Optim.*, 11(3):796–817, 2001. 6.1

[Lau09]   Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 157–270. Springer, New York, 2009. 6.1

[MS77]   F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I*. North-Holland Publishing Co., Amsterdam, 1977. North-Holland Mathematical Library, Vol. 16. 5.5

[MWW15]   Ryuhei Mori, Osamu Watanabe, and David Witmer. Improved SDP integrality gaps for random CSPs. In progress, 2015. 2.2

[O'D14]   Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 3.3, B.1

[OW14]   Ryan O'Donnell and David Witmer. Goldreich's PRG: Evidence for near-optimal polynomial stretch. In *Proceedings of the 29th IEEE Conference on Computational Complexity*, 2014. 2.2

[OZ13]   Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1537–1556. SIAM, Philadelphia, PA, 2013. 2.2, 6.1, A.3, A.3

[Par00]   Pablo Parillo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000. 2.2, 6.1

[Rag09]   Prasad Raghavendra. *Approximating NP-hard problems: efficient algorithms and their limits*. PhD thesis, University of Washington, 2009. 1

[Rao07]   Anup Rao. An Exposition of Bourgain's 2-Source Extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2007. B.3

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

[Sch08]   Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 593–602, 2008. 1, 2.2

[Sho87]   N.Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987. 6.1

[Tao12]   Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. A.1, A.4

[Tul09]   Madhur Tulsiani. Csp gaps and reductions in the lasserre hierarchy. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, pages 303–312, 2009. 2.2

[TW13]   Madhur Tulsiani and Pratik Worah. LS_+ lower bounds from pairwise independence. In *Proceedings of 28th IEEE Conference on Computational Complexity*, 2013. 2.2

[Val84]   Leslie Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, November 1984. 5

[Vaz86]   Umesh Vazirani. *Randomness, adversaries and computation (random polynomial time)*. PhD thesis, University of California, Berkeley, 1986. B.3

[Wig55]   Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955. 2

A Proof of Theorem 4.1

We restate Theorem 4.1:

Theorem 4.1. *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 $T \in [n]^k$,*

$$\mathbf{E}[w(T)] = 0 \tag{24}$$

$$\mathbf{Pr}[w(T) \neq 0] \leq p \tag{25}$$

$$|w(T)| \leq 1. \tag{26}$$

Then there is an efficient algorithm certifying that

$$\sum_{T \in [n]^k} w(T) x^T \leq 2^{O(k)} \sqrt{p} n^{3k/4} \log^{3/2} n. \tag{27}$$

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

The proof of this theorem constitutes the remainder of this section. It will often be convenient to consider $T \in [n]^k$ to be $(T_1, T_2) \in [n]^{k_1} \times [n]^{k_2}$ with $k_1 + k_2 = k$. In such situations, we will write $w(T) = w(T_1, T_2)$. For intuition, the reader can think of the special case of $w(T) \in \{-1, 0, 1\}$ for all T and $y \in \{-1, 1\}^n$. Under these additional constraints, $\sum_{T \in [n]^k} w(T) x^T$ is $\mathrm{Opt}(\mathcal{I}) - \frac{1}{2}$ for a random k-XOR instance $\mathcal{I}$ so we are certifying that a random k-XOR instance does not have value much bigger than $\frac{1}{2}$.

A.1 The even arity case

When k is even, we can think of $\sum_{T \in [n]^k} w(T) x^T$ as a quadratic form:

$$\sum_{T \in [n]^k} w(T) x^T = \sum_{T_1, T_2 \in [n]^{k/2}} w(T_1, T_2) y_{T_1} y_{T_2}, \tag{28}$$

where $y_U = x^U$. We give two methods to certify that the value of this quadratic form is at most $O_k(\sqrt{p} n^{3k/4} \log n)$. The first method uses an SDP-based approximation algorithm and works only for $x \in \{-1, 1\}^n$. The second method uses ideas from random matrix theory and works for any x with $\|x\|_\infty \leq 1$.

Approximation algorithms approach If $x \in \{-1, 1\}^n$, we can apply an approximation algorithm of Charikar and Wirth [CW04] for quadratic programming. They prove the following theorem:

Theorem A.1. [CW04, Theorem 1] *Let M be any $n \times n$ matrix with all diagonal elements 0. There exists an efficient randomized algorithm that finds $y \in \{-1, 1\}^n$ such that*

$$\mathbf{E}[y^\top M y] \geq \Omega\left(\frac{1}{\log n}\right) \max_{x \in \{-1,1\}^n} x^\top M x.$$

By Markov's Inequality, this statement holds with probability at least $1/2$. We can run the algorithm $O(\log n)$ times to get a high probability result. To apply Theorem A.1, we separate out the diagonal terms of (28), rewriting it as

$$\sum_{T_1 \neq T_2 \in [n]^{k/2}} w(T_1, T_2) y_{T_1} y_{T_2} + \sum_{U \in [n]^{k/2}} w(U, U) y_U^2. \tag{29}$$

35

We can certify that each of the two terms in this expression is at most $O(\sqrt{p}n^{3k/4}\log n)$. For the first term, we will need the following claim.

Claim A.2. *With high probability, it holds that*

$$\sum_{T_1,T_2\in[n]^{k/2}} w(T_1,T_2)y_{T_1}y_{T_2} \leq O(\sqrt{p}n^{3k/4}).$$

This follows from applying Bernstein's Inequality (Theorem 4.12) for fixed y and then taking a union bound over all $y \in \{-1,1\}^{n^{k/2}}$. Using the claim, we see that Theorem A.1 allows us to certify that the value of the first term in (29) is at most $O(\sqrt{p}n^{3k/4}\log n)$.

We will use the next claim to bound the second term of (29).

Claim A.3. *With high probability, it holds that*

$$\sum_{U\in[n]^{k/2}} |w_{U,U}| \leq O(\sqrt{p}n^{3k/4}).$$

Since the $|w_T| \leq 1$ and $\mathbf{Pr}[w_T \neq 0] \leq p$, the claim follows from the Chernoff Bound. The second term of (29) is upper bounded by $\sum_{U\in[n]^{k/2}} |w_{U,U}|$ and we can compute this quantity in polynomial time to certify that its value is at most $O(\sqrt{p}n^{3k/4})$.

Random matrix approach Observe that (28) is $y^\top By$ for a matrix B indexed by $U \in [n]^{k/2}$ so that $B_{U_1,U_2} = w(U_1,U_2)$. Then $y^\top By \leq \|B\| \|y\|^2$. To certify that $y^\top By$ is small, we compute $\|B\|$. We need to show that $\|B\|$ is small with high probability. First, note that $\|B\|$ is equal to the norm of the $2n^{k/2} \times 2n^{k/2}$ symmetric matrix

$$\tilde{B} = \begin{pmatrix} 0 & B \\ B^\top & 0 \end{pmatrix}$$

For example, this appears as (2.80) in [Tao12]. The upper triangular entries of $\tilde{B}$ are independent random variables with mean 0 and variance at most p by the properties of the w_S's. We can then apply a standard bound on the norm of random symmetric matrices [Tao12].

Proposition A.4. [Tao12, Proposition 2.3.13] *Let M be a random symmetric matrix $n \times n$ whose upper triangular entries M_{ij} with $i \geq j$ are independent random variables with mean 0, variance at most 1, and magnitude at most K. Then, with high probability,*

$$\|M\| = O(\sqrt{n}\log n \cdot \max\{1, K/\sqrt{n}\}).$$

Let $\tilde{B}' = \frac{1}{\sqrt{p}}\tilde{B}$. The upper triangular entries of $\tilde{B}'$ are independent random variables with mean 0, variance at most 1, and magnitude at most $1/\sqrt{p}$. Applying Proposition A.4 to $\tilde{B}'$ shows that $\|B\| = O\left(kn^{k/4}\sqrt{p}\log n \cdot \max\left\{1, \frac{1}{\sqrt{p}n^{k/4}}\right\}\right)$ with high probability. Since $\|y\|_\infty \leq 1$ by assumption, $\|y\|^2 \leq n^{k/2}$ and (28) is at most $O(k\sqrt{p}n^{3k/4}\log n)$ with high probability when $p \geq n^{-k/2}$.

A.2 The odd arity case

Fix an assignment $x \in [-1,1]^n$. For $i \in [n]$, the monomials containing x_i can contribute at most $W_i := \left| \sum_{T \in [n]^{k-1}} w(T,i) x^T \right|$ to the objective if x_i is set optimally. By Cauchy-Schwarz,

$$\sum_{T \in [n]^k} w(T) x^T \leq \sum_{i \in [n]} W_i \leq \sqrt{n} \sqrt{\sum_{i \in [n]} W_i^2}, \tag{30}$$

so it suffices to bound $\sum_{i \in [n]} W_i^2$. We will write this as a quadratic polynomial and then bound it using spectral methods:

$$\sum_{i \in [n]} W_i^2 = \sum_{T,U \in [n]^{k-1}} \sum_{i \in [n]} w(T,i) w(U,i) x^T x^U$$
$$= \sum_{T_1', T_2', U_1', U_2' \in [n]^{\frac{k-1}{2}}} \sum_{i \in [n]} w(T_1', U_1', i) w(T_2', U_2', i) x^{(T_1', T_2')} x^{(U_1', U_2')}. \tag{31}$$

Define the $n^{k-1} \times n^{k-1}$ matrix A indexed by $[n]^{k-1}$:

$$A_{(i_1,i_2),(j_1,j_2)} = \begin{cases} \sum_{\ell \in [n]} w(i_1, j_1, \ell) w(i_2, j_2, \ell) & \text{if } (i_1, j_1) \neq (i_2, j_2) \\ 0 & \text{otherwise,} \end{cases} \tag{32}$$

where we have divided the indices of A into 2 blocks of $\frac{k-1}{2}$ coordinates each. Define $x^{\otimes k-1} \in \mathbb{R}^{[n]^{k-1}}$ so that $x^{\otimes k-1}(T) = x^T$. Then (31) is equal to

$$(x^{\otimes k-1})^\top A x^{\otimes k-1} + \sum_{T,U \in [n]^{\frac{k-1}{2}}} (x^T)^2 (x^U)^2 \sum_{i \in [n]} w(T,U,i)^2. \tag{33}$$

The first term is at most $\|A\| n^{k-1}$ since the variables are bounded. We can compute $\|A\|$ to certify this. With high probability, $\|A\|$ is not too big.

Lemma A.5. *Let $k \geq 3$ and $p \geq n^{-k/2}$. Let $\{w(T)\}_{T \in [n]^k}$ be indepedent random variables satisfying conditions (6), (7), and (8) above. Let A be defined as in (32). With high probability,*

$$\|A\| \leq 2^{O(k)} p n^{k/2} \log^3 n.$$

We can therefore certify that the first term is $2^{O(k)} p n^{3k/2-1} \log^3 n$. We will prove the lemma in Appendix A.4.

The second term of (33) is at most $\sum_{T \in [n]^k} w(T)^2$. We can easily compute this and the Chernoff Bound implies that its value is at most $p n^{3k/2-1}$ with high probability.

So far, with high probability we can certify that $\sum_{i \in [n]} W_i^2 = 2^{O(k)} p n^{3k/2-1} \log^3 n$. Plugging this bound into (30) concludes the proof.

Remark A.6. It would have been more natural to have written $\sum_{i \in [n]} W_i^2 = (x^{\otimes k-1})^\top A' x^{\otimes k-1}$ for A' such that $A'_{T,U} = \sum_{i \in [n]} w(T,i) w(U,i)$. However, $\|A'\|$ could be too large because of the contribution of the second term in (33). We use the additional assumption that $\|x\|_\infty \leq 1$ to get around this issue.

Remark A.7. We have defined A so that $A_{(i_1,i_2),(j_1,j_2)} = \sum_{\ell \in [n]} w(i_1, j_1, \ell) w(i_2, j_2, \ell)$, not $A_{i,j} = \sum_{\ell \in [n]} w(i, \ell) w(j, \ell)$. The reduces the correlation among entries $w(b,c)$ and $w(b,c')$ for $c \neq c'$. Intuitively, A looks more like a random matrix with independent entries, so we can bound its norm using the trace method. See the proof of Lemma A.5 in Appendix A.4.

A.3 An SOS version

In this section, we will prove the SOS version of Theorem 4.1.

Theorem 6.2. *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 $T \in [n]^k$,*

$$\mathbf{E}[w(T)] = 0$$
$$\mathbf{Pr}[w(T) \neq 0] \leq p$$
$$|w(T)| \leq 1.$$

Then, with high probability,

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_{2k} \sum_{T \in [n]^k} w(T) x^T \leq 2^{O(k)} \sqrt{p} n^{3k/4} \log^{3/2} n.$$

Rather than writing out the full proof, we will indicate the small changes required to convert the above proof of Theorem 4.1 into SOS form.

Even arity The random matrix proof for the even case can easily be converted into an SOS proof with degree k. When $O(k\sqrt{p}n^{k/4}\log n)I - B \succeq \mathbf{0}$, there exists a matrix M such that $M^\top M = O(k\sqrt{p}n^{k/4}\log n)I - B$. Then

$$O(k\sqrt{p}n^{k/4}\log n)\|y\|^2 - y^\top B y = (My)^\top(My) = \sum_{T \in [n]^{k/2}} \left(\sum_{U \in [n]^{k/2}} M_{T,U} y_U \right)^2$$

so

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_k \sum_{T \in [n]^k} w(T) x^T \leq O(k\sqrt{p}n^{3k/4}\log n).$$

Odd arity A couple of additional issues arise in the odd case. First of all, the square root in (30) is not easily expressed in SOS, so we instead prove the squared version

$$\left(\sum_{T \in [n]^k} w(T) x^T \right)^2 \leq 2^{O(k)} n^{3k/2} \log^3 n. \tag{34}$$

By a simple extension of [OZ13, Fact 3.3], (34) implies (27) in SOS :

Fact A.8.

$$X^2 \leq b^2 \vdash_2 X \leq b.$$

Proof.

$$\frac{1}{2b}(b^2 - X^2) + \frac{1}{2b}(b - X)^2 = \frac{b}{2} - \frac{1}{2b}X^2 + \frac{b}{2} - X + \frac{1}{2b}X^2 = b - X. \qquad \square$$

Secondly, we do not know how to prove the Cauchy-Schwarz inequality (30) in SOS. However, O'Donnell and Zhou show that a very similar inequality can be proved in SOS [OZ13, Fact 3.8]:

Fact A.9.

$$\vdash_2 YZ \leq \frac{1}{2}Y^2 + \frac{1}{2}Z^2.$$

Using this fact instead of Cauchy-Schwarz to prove the squared version of (30), we can follow the argument above to show that

$$\{x_i^2 \leq 1\}_{i \in [n]} \vdash_{2k} \left(\sum_{T \in [n]^k} w(T) x^T \right)^2 \leq n \left(x^{\otimes k-1} \right)^\top A x^{\otimes k-1} + n \sum_{T \in [n]^k} w(T)^2.$$

The norm bound can be proven in SOS exactly as in the even case.

A.4 Proof of Lemma A.5

We restate the definition of the matrix A and the statement of the lemma.

Lemma A.5. *Let $k \geq 3$ and $p \geq n^{-k/2}$. Let $\{w(T)\}_{T \in [n]^k}$ be indepedent random variables satisfying conditions (6), (7), and (8) above. Let A be the $[n]^{k-1} \times [n]^{k-1}$ indexed by $[n]^{k-1}$ that is defined as follows:*

$$A_{(i_1,i_2),(j_1,j_2)} = \begin{cases} \sum_{\ell \in [n]} w(i_1,j_1,\ell) w(i_2,j_2,\ell) & \text{if } (i_1,j_1) \neq (i_2,j_2) \\ 0 & \text{otherwise.} \end{cases}$$

Then with high probability,

$$\|A\| \leq 2^{O(k)} p n^{k/2} \log^3 n.$$

The proof closely follows the arguments of [COGL04, Lemma 17] and [BM15, Section 5]. Both proofs use the trace method: To bound the norm of a symmetric random matrix M, it suffices to bound $\mathbf{E}[\text{tr}(M^r)]$ for large r. For non-symmetric matrices, we can instead work with $MM^\top$. In our particular case, we have the following.

Claim A.10. *If $\mathbf{E}[\text{tr}((AA^\top)^r)] \leq n^{O(k)} 2^{O(r)} r^{6r} p^{2r} n^{kr}$, then $\|A\| \leq 2^{O(k)} p n^{k/2} \log^3 n$ with high probability.*

Proof. Observe that $\|A\|^{2r} \leq \text{tr}((AA^\top)^r)$. By Markov's Inequality, $\mathbf{Pr}[\|A\| \geq B] \leq \frac{\mathbf{E}[\text{tr}((AA^\top)^r)]}{B^{2r}}$. We get the claim by plugging in $r = \Theta(\log n)$ and setting constants appropriately. $\square$

Remark A.11. We can get arbitrarily small $1/\text{poly}(n)$ probability of failure: This proof shows that $\|A\| \leq K 2^{O(k)} p n^{k/2} \log^3 n$ with probability at most $n^{-\log K}$.

In the the remainder of this section, we will bound $\mathbf{E}[\text{tr}((AA^\top)^r)]$.

Lemma A.12. *Under the conditions of Lemma A.5, $\mathbf{E}[\text{tr}((AA^\top)^r)] \leq n^{O(k)} 2^{O(r)} r^{6r} p^{2r} n^{kr}$ with high probability.*

Proof. Recall that we index A by elements of n^{k-1} divided into two blocks of $\frac{k-1}{2}$ coordinates each. First, note that

$$\text{tr}((AA^\top)^r) = \sum_{i_1,\ldots,i_{2r} \in [n]^{\frac{k-1}{2}}} (AA^\top)_{(i_1,i_2),(i_3,i_4)} (AA^\top)_{(i_3,i_4),(i_5,i_6)} \cdots (AA^\top)_{(i_{2r-1},i_{2r}),(i_1,i_2)}.$$

Expanding this out using the definition of A and setting $w_T = w(T)$, we get that

$$\text{tr}((AA^\top)^r) = \sum w_{i_1,j_1,\ell_1} w_{i_2,j_2,\ell_1} w_{i_3,j_1,\ell_2} w_{i_4,j_2,\ell_2} \cdots w_{i_{2r-1},j_{2r-1},\ell_{2r-1}} w_{i_{2r},j_{2r},\ell_{2r-1}} w_{i_1,j_{2r-1},\ell_{2r}} w_{i_2,j_{2r},\ell_{2r}},$$

where the sum is over $\ell_1, \ldots, \ell_{2t} \in [n]$ and $i_1, \ldots, i_{2t}, j_1, \ldots, j_{2t} \in [n]^{k-1}$ satisfying

$$
\begin{align}
(i_s, j_s) &\neq (i_{s+1}, j_{s+1}) && \text{for } 1 \leq s \leq 2r - 1 \tag{35}\\
(i_{s+2}, j_s) &\neq (i_{s+3}, j_{s+1}) && \text{for } 1 \leq s \leq 2r - 3 \tag{36}\\
(i_1, j_{2r-1}) &\neq (i_2, j_{2r}). && \tag{37}
\end{align}
$$

Let Ω be the set of all $(i_1, \ldots, i_{2r}, j_1, \ldots, j_{2r}) \in ([n]^{\frac{k-1}{2}})^{4r}$ satisfying (35), (36), and (37). Then for $J \in \Omega$ and $L = (\ell_1, \ldots, \ell_{2r}) \in [n]^{2r}$, define

$$
P_{J,L} = w_{i_1,j_1,\ell_1} w_{i_2,j_2,\ell_1} w_{i_3,j_1,\ell_2} w_{i_4,j_2,\ell_2} \cdots w_{i_{2r-1},j_{2r-1},\ell_{2r-1}} w_{i_{2r},j_{2r},\ell_{2r-1}} w_{i_1,j_{2r-1},\ell_{2r}} w_{i_2,j_{2r},\ell_{2r}}. \tag{38}
$$

Let $|J| = |\{i_1, \ldots, i_{2r}, j_1, \ldots, j_{2r}\}|$ be the number of distinct elements of $[n]^{\frac{k-1}{2}}$ in J and define $|L| = |\{\ell_1, \ldots, \ell_{2r}\}|$ similarly. We then have

$$
\mathbf{E}[\operatorname{tr}((AA^\top)^r)] = \sum_{J \in \Omega} \sum_{L \in [n]^{2r}} \mathbf{E}[P_{J,L}] = \sum_{a=1}^{4r} \sum_{b=1}^{2r} \sum_{\substack{J \in \Omega \\ |J|=a}} \sum_{\substack{L \in [n]^{2r} \\ |L|=b}} \mathbf{E}[P_{J,L}].
$$

To bound this sum, we will start by bounding $\mathbf{E}[P_{J,L}]$. We will need two claims.

Claim A.13. *The number of distinct $w_{i,j,\ell}$ factors in $P_{J,L}$ is at least $2|L|$.*

Proof. For each $\ell \in L$, (38) shows that $P_{J,L}$ contains a pair of the form $w_{i_s,j_s,\ell} w_{i_{s+1},j_{s+1},\ell}$ or $w_{i_{s+2}j_s,\ell} w_{i_{s+3},j_{s+1},\ell}$. Since $J \in \Omega$, we know that $(i_s, j_s) \neq (i_{s+1}, j_{s+1})$ or $(i_{s+2}, j_s) \neq (i_{s+3}, j_{s+1})$, so each of these pairs must have two distinct $w_{i,j,\ell}$ factors. We then have at least $2|L|$ distinct $w_{i,j,\ell}$ factors. $\square$

Claim A.14. *The number of distinct $w_{i,j,\ell}$ factors in $P_{J,L}$ is at least $|J| - 2$.*

Proof. Consider looking over the factors of $P_{J,L}$ from left to right in the order of (38) until we have seen all elements of J. The first pair $w_{i_1,j_1,\ell_1} w_{i_2,j_2,\ell_1}$ contains at most four previously-unseen elements of J. Every subsequent pair of factors $w_{i_s,j_s,\ell_s} w_{i_{s+1},j_{s+1}}$ or $w_{i_{s+2}j_s,\ell_{s+1}} w_{i_{s+3},j_{s+1},\ell_{s+1}}$ in $P_{J,L}$ shares two variables of J with its preceding pair. Each such pair can then contain at most two new elements of J. After seeing u $w_{i,j,\ell}$'s, we have therefore seen at most $4 + 2\left(\frac{u-2}{2}\right)$ distinct elements of J. To get all $|J|$ elements of J, we must have seen at least $|J| - 2$ $w_{i,j,\ell}$'s and these must be distinct. $\square$

Since $\mathbf{Pr}[w_{i,j,\ell} \neq 0] \leq p$, $\mathbf{E}[P_{J,L}] \leq p^{\#\{\text{distinct } w_{i,j,\ell} \text{ factors in } P_{J,L}\}}$. It then follows that

$$
\mathbf{E}[P_{J,L}] \leq p^{\max\{2|L|,|J|-2\}}.
$$

The two claims also imply two other facts we will need below.

Claim A.15. *If $|L| > r$, then $\mathbf{E}[P_{J,L}] = 0$.*

Proof. We will show that if $|L| > r$, there is an $w_{i,j,\ell}$ factor in $P_{J,L}$ that occurs exactly once. Since $\mathbf{E}[w_{i,j,\ell}] = 0$, this proves the claim.

Assume for a contradiction that $|L| > r$ and every $w_{i,j,\ell}$ factor occurs at least twice. Since there are at least $2|L|$ distinct $w_{i,j,\ell}$'s, there must be at least $4|L| > 4r$ total $w_{i,j,\ell}$'s. However, looking at (38), $P_{J,L}$ has at most $4r$ $w_{i,j,\ell}$ factors. $\square$

Claim A.16. *If $|J| > 2r + 2$, then $\mathbf{E}[P_{J,L}] = 0$.*

This can be proved in exactly the same manner.

Next, observe that the number of choices of J with $|J| = a$ is at most $n^{\frac{a(k-1)}{2}}a^{4r} \leq n^{\frac{a(k-1)}{2}}(4r)^{4r}$. The number of choices of L with $|L| = b$ is at most $n^b b^{2r} \leq n^b(2r)^{2r}$. All together, we can write

$$\mathbf{E}[\operatorname{tr}((AA^\top)^r)] \leq \sum_{a=1}^{2r+2}\sum_{b=1}^{r} 2^{10r}r^{6r}n^{\frac{a(k-1)}{2}+b}p^{\max\{2b,a-2\}}.$$

We bound each term of the sum.

Claim A.17.

$$n^{\frac{a(k-1)}{2}+b}p^{\max\{2b,a-2\}} \leq n^{kr+k-1}p^{2r}.$$

Proof. If $2b > a - 2$,

$$n^{\frac{a(k-1)}{2}+b}p^{\max\{2b,a-2\}} \leq n^{\frac{(2b+2)(k-1)}{2}+b}p^{\max\{2b,a-2\}} = n^{k-1}(n^k p^2)^b.$$

If $2b \leq a - 2$,

$$n^{\frac{a(k-1)}{2}+b}p^{\max\{2b,a-2\}} \leq n^{\frac{a(k-1)}{2}+\frac{a}{2}-1}p^{a-2} = n^{k-1}(n^k p^2)^{a/2-1}.$$

Recall that we assumed $n^k p^2 \geq 1$. Since $a \leq 2r + 2$ and $b \leq r$, the claim follows. $\qquad\square$

To conclude, observe that

$$\mathbf{E}[\operatorname{tr}[(AA^\top)^r]] \leq \sum_{a=1}^{2r+2}\sum_{b=1}^{r} 2^{10r}r^{6r}n^{kr+k-1}p^{2r} \leq n^{O(k)}2^{O(r)}r^{6r}p^{2r}n^r$$

$\qquad\square$

Remark A.18. If we did not have conditions (35), (36), and (37), we would only have been able to show that $|L| \leq 2r$. This would have led to a weaker bound of $O(\sqrt{n})$.

B Extension to larger alphabets

B.1 Preliminaries

CSPs over larger domains We begin by discussing CSPs over domains of size $q > 2$. We prefer to identify such domains with $\mathbb{Z}_q$, so our predicates are $P : \mathbb{Z}_q^k \to \{0,1\}$. The extensions of the definitions and facts from Section 3.1 are straightforward; the only slightly nonobvious notion is that of a literal. We take the fairly standard [Aus08] definition that a literal for variable x_i is any $x_i + c$ for $c \in \mathbb{Z}_q$. Thus there are now q^k possible "negation patterns" c for a P-constraint. We denote by $\mathcal{F}_{q,P}(n,p)$ the distribution over instances of $\operatorname{CSP}(P)$ in which each of the $q^k n^k$ constraints is included with probability p; the expected number of constraints is therefore $\overline{m} = q^k n^k p$. We have the following slight variant of Fact 3.6.

Fact B.1. *Let $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$. 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. $\operatorname{Opt}(\mathcal{I}) \leq \overline{P} \cdot \left(1 + O\left(\sqrt{\frac{\log q}{\overline{P}} \cdot \frac{n}{\overline{m}}}\right)\right).$

3. $\mathcal{I}$ is $O\left(\sqrt{q^k \log q \cdot \frac{n}{\overline{m}}}\right)$-quasirandom.

Fourier analysis over larger domains Let $\mathcal{U}_q$ is the uniform distribution over $\mathbb{Z}_q$. We consider the space $L^2(\mathbb{Z}_q, \mathcal{U}_q)$ of functions $f : \mathbb{Z}_q \to \mathbb{R}$ equipped with the inner product $\langle f, g \rangle = \mathbf{E}_{z \sim \mathcal{U}_q}[f(z)g(z)]$ and its induced norm $\|f\|_2 = \mathbf{E}_{z \sim \mathcal{U}_q}[f(z)^2]^{1/2}$. Fix an orthonormal basis $\chi_0, \ldots, \chi_{q-1}$ such that $\chi_0 = 1$.

Now let $L^2(\mathbb{Z}_q^k, \mathcal{U}_q^k)$ be the space of functions $f : \mathbb{Z}_q^k \to \mathbb{R}$, where $\mathcal{U}_q^k$ is the uniform distribution over $\mathbb{Z}_q^k$ and we have the analogous inner product and norm. Then, for $\sigma \in \mathbb{Z}_q^k$, define $\chi_\sigma : \mathbb{Z}_q^k \to \mathbb{R}$ such that

$$\chi_\sigma(x) = \prod_{i \in [k]} \chi_{\sigma_i}(x_i).$$

The set $\{\chi_\sigma\}_{\sigma \in \mathbb{Z}_q^k}$ forms an orthonormal basis for $L^2(\mathbb{Z}_q^k, \mathcal{U}_q^k)$ [Aus08, Fact 2.3.1] and we can write any function $f : \mathbb{Z}_q^k \to \mathbb{R}$ in terms of this basis:

$$f(x) = \sum_{\sigma \in \mathbb{Z}_q^k} \widehat{f}(\sigma) \chi_\sigma(x).$$

Orthonormality once again gives us Plancherel's Theorem in this setting:

Theorem B.2.
$$\langle f, g \rangle = \sum_{\sigma \in \mathbb{Z}_q^k} \widehat{f}(\sigma) \widehat{g}(\sigma).$$

For $\sigma \in \mathbb{Z}_q^k$, define $\mathrm{supp}(\sigma) = \{i \in [k] \mid \sigma_i \neq 0\}$ and $|\sigma| = |\mathrm{supp}(\sigma)|$. Then we define the degree of f to be $\max\{|\sigma| \mid \widehat{f}(\sigma) \neq 0\}$. Note that this is the degree of f when it is written as a polynomial in the χ_a's for $a \in \mathbb{Z}_q$.

Given a k-tuple T and $\sigma \in \mathbb{Z}_q^k$, we use $T(\sigma)$ to denote the $|\sigma|$-tuple formed by taking the projection of T onto the coordinates in $\mathrm{supp}(\sigma)$. Similarly, use $T(\overline{\sigma})$ to denote the $(k - |\sigma|)$-tuple formed by taking the projection of T onto coordinates in $[k] \setminus \mathrm{supp}(\sigma)$.

See [O'D14, Aus08] for more background on Fourier analysis over larger domains.

B.2 Conversion to Boolean functions

To more easily apply our above results, we would like to rewrite a function $f : \mathbb{Z}_q^k \to \mathbb{R}$ as a Boolean function $f^b : \{0,1\}^{k'} \to \mathbb{R}$ for some k'. It will actually be more convenient to define f^b on a subset of $\{0,1\}^{k'}$. In particular, consider the set $\Omega_k = \{v \in \{0,1\}^{[k] \times \mathbb{Z}_q} \mid \sum_{a \in \mathbb{Z}_q} v(i,a) = 1 \ \forall i \in [k]\}$. Note there is a bijection ϕ between $\mathbb{Z}_q^k$ and Ω_k: For $z \in \mathbb{Z}_q^k$, $(\phi(x))(i,a) = 1_{\{z_i = a\}}$. In the other direction, given $v \in \Omega_k$ set $\phi^{-1}(v)_i = \sum_{a \in \mathbb{Z}_q} a \cdot v(i,a)$.

For a function $f : \mathbb{Z}_q^k \to \mathbb{R}$, we can then define its Boolean version $f^b : \Omega_k \to \mathbb{R}$ as

$$f^b(v) = \sum_{\alpha \in \mathbb{Z}_q^k} f(\alpha) \prod_{i \in [k]} v(i, \alpha_i),$$

Observe that $f(z) = f^b(\phi(z))$ for $z \in \mathbb{Z}_q^k$. Also, note that if $f(z) = g(z)$ for all $z \in \mathbb{Z}_q^k$, $f^b = g^b$ over all $\mathbb{R}^k$ by construction. f^b is a multilinear polynomial and its degree is defined in the standard way. The degree of f is defined as in the previous section.

Claim B.3. *The degree of f^b is equal to the degree of f.*

Proof. Abbreviate $\text{supp}(\sigma)$ as $s(\sigma)$ and denote $\text{supp}(\sigma)$'s complement with respect to $[k]$ as $s(\overline{\sigma})$. Applying the definition and writing f's Fourier expansion, we see that $f^b(v)$ is equal to

$$\sum_{\alpha \in \mathbb{Z}_q^k} \sum_{\sigma \in \mathbb{Z}_q^k} \widehat{f}(\sigma) \chi_\sigma(\alpha) \prod_{i \in [k]} v(i, \alpha_i) = \sum_{\sigma \in \mathbb{Z}_q^k} \widehat{f}(\sigma) \sum_{\alpha' \in \mathbb{Z}_q^{|\sigma|}} \chi_\sigma(\alpha') \prod_{i=1}^{|\sigma|} v(s(\sigma)_i, \alpha_i') \sum_{\alpha'' \in \mathbb{Z}_q^{k-|\sigma|}} \prod_{i=1}^{k-|\sigma|} v(s(\overline{\sigma})_i, \alpha_i'').$$

Now observe that

$$\sum_{\alpha'' \in \mathbb{Z}_q^{k-|\sigma|}} \prod_{i=1}^{k-|\sigma|} v(s(\overline{\sigma})_i, \alpha_i'') = \prod_{i=1}^{k-|\sigma|} \sum_{a \in \mathbb{Z}_q} v(s(\overline{\sigma})_i, a) = 1$$

by the assumption that $v \in \Omega_k$. The degree of f^b is therefore $|\sigma|$. $\qquad\square$

B.3 Quasirandomness and strong refutation

To prove quasiandomness and strong refutation results for CSPs over larger alphabets, we proceed exactly as in the binary case. We used the $t = k$ case of Lemma 3.13 (the Vazirani XOR Lemma [Vaz86, Gol11]) to certify quasirandomness for binary CSPs. A generalization of this case holds for Abelian groups [Rao07, Lemma 4.2].

Lemma B.4. *Let G be an Abelian group and let $\mathcal{U}_G$ be the uniform distribution over G. Also, let $\{\chi_\sigma\}_{\sigma \in G}$ be an orthonormal basis for $L^2(G, \mathcal{U}_G)$ and let $D : G \to \mathbb{R}$ be a distribution over G. If $\widehat{D}(\sigma) \leq \epsilon$ for all $\sigma \in G$, then $d_{\text{TV}}(D, \mathcal{U}_G) \leq \frac{1}{2}|G|^{3/2}\epsilon$.*

Viewing the induced distribution density $D_{\mathcal{I},x}(\sigma)$ as a function of $x \in \mathbb{Z}_q^n$ for fixed $\sigma \in \mathbb{Z}_q^k$, we will consider $D_{\mathcal{I},y}^b(\sigma) : \Omega_n \to \mathbb{R}$. As before, we can certify that $D_{\mathcal{I},y}^b$ has small Fourier coefficients.

Lemma B.5. *Let $\sigma \in \mathbb{Z}_q^k$ such that $\sigma \neq \mathbf{0}$ and $|\sigma| = s$. There is an algorithm that with high probability certifies that*

$$\left| \widehat{D_{\mathcal{I},y}^b}(\sigma) \right| \leq \frac{q^{O(k)} \max\{n^{s/4}, \sqrt{n}\} \log^{5/2} n}{\sqrt{\overline{m}}}$$

for all $y \in \{0,1\}^{[n] \times \mathbb{Z}_q}$ when $\overline{m} \geq \max\{n^{s/2}, n\}$.

Proof. The proof is essentially identical to the proof of Lemma 4.3. We highlight the differences. First of all, we can write

$$\widehat{D_{\mathcal{I},y}^b}(\sigma) = \sum_{x \in \mathbb{Z}_q^n} \widehat{D_{\mathcal{I},x}}(\sigma) \prod_{i \in [n]} y(i, x_i) = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \sum_{x \in \mathbb{Z}_q^n} \chi_\sigma(x_T + c) \prod_{i \in [n]} y(i, x_i).$$

Since χ_σ only depends on coordinates in $\text{supp}(\sigma)$, we can rearrange and use the fact that $\sum_{a \in \mathbb{Z}_q} y_{i,a} = 1$ to get

$$\widehat{D_{\mathcal{I},y}^b}(\sigma) = \frac{1}{m} \sum_{\alpha \in \mathbb{Z}_q^{|\sigma|}} \sum_{U \in [n]^{|\sigma|}} \prod_{i=1}^{|\sigma|} y(U_i, \alpha_i) \sum_{\substack{T \in [n]^k \\ T(\sigma) = U}} w_{\sigma,\alpha}(T),$$

where $w_{\sigma,\alpha}(T) = \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \chi_\sigma(\alpha + c(\sigma))$. Observe that $\mathbf{E}[w_{\sigma,\alpha}(T)] = 0$ and $\mathbf{Pr}[w_{\sigma,\alpha}(T) \neq 0] \leq q^k p$. Since $\|\chi_\sigma\| = 1$, observe that the Cauchy-Schwarz Inequality implies that $|\chi_\sigma| \leq q^{k/2}$ for all σ. Then $|w_{\sigma,\alpha}(T)| \leq q^{3k/2}$ for all α and σ. For every α, we can then apply Lemma 4.4 just as in the proof of Lemma 4.3. $\qquad\square$

These two lemmas then imply the larger alphabet versions of the quasirandomness certification and strong refutation results above.

Theorem B.6. *There is an efficient algorithm that certifies that an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of* $\mathrm{CSP}(P)$ *is γ-quasirandom with high probability when* $\overline{m} \geq \frac{q^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.

Theorem B.7. *There is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of* $\mathrm{CSP}(P)$, *certifies that* $\mathrm{Opt}(\mathcal{I}) \leq \overline{P} + \gamma$ *with high probability when* $\overline{m} \geq \frac{q^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.

B.4 Refutation of non-t-wise supporting CSPs

We will show that the dual polynomial characterization of being far from t-wise supporting described in Section 3.4 generalizes to larger alphabets. We extend the definitions of t-wise supporting and δ-separating polynomials to the $\mathbb{Z}_q$ case in the natural way.

Lemma B.8. *For $P : \mathbb{Z}_q^k \to \{0, 1\}$ and $0 \leq \delta < 1$, there exists a polynomial $Q : \mathbb{Z}_q^k \to \mathbb{R}$ of degree at most t that δ-separates P if and only if P is δ-far from supporting a t-wise uniform distribution.*

Proof. The proof uses the following dual linear programs exactly as in the proof of Lemma 3.16.

$$
\begin{aligned}
&\text{minimize} && \sum_{z \in \mathbb{Z}_q^k} (1 - P(z))\mathcal{D}(z) && (39)\\
&\text{s.t.} && \sum_{z \in \mathbb{Z}_q^k} \mathcal{D}(z)\chi_\sigma(z) = q^k \widehat{\mathcal{D}}(\sigma) = 0 && \forall \sigma \in \mathbb{Z}_q^k \quad 0 < |\sigma| \leq t && (40)\\
&&& \sum_{z \in \mathbb{Z}_q^k} \mathcal{D}(z) = 1 &&\\
&&& \mathcal{D}(z) \geq 0 && \forall z \in \mathbb{Z}_q^k &&
\end{aligned}
$$

$$
\begin{aligned}
&\text{maximize} && \xi\\
&\text{s.t.} && \sum_{\substack{\sigma \in \mathbb{Z}_q^k \\ 0 < |\sigma| \leq t}} c(S)\chi_\sigma(z) \leq 1 - P(z) - \zeta && \forall z \in \mathbb{Z}_q^k.
\end{aligned}
$$

To prove Lemma 3.16, we needed to show in the binary case that feasible solutions to the primal LP (1) were t-wise uniform. We now argue that the constraint (40) is a sufficient condition for t-wise uniformity of $\mathcal{D}$ in the q-ary case. For a distribution $\mathcal{D}$ over $\mathbb{Z}_q^k$ and $S \subseteq [k]$, define $\mathcal{D}_S$ to be the marginal distribution of $\mathcal{D}$ on $(\mathbb{Z}_q^k)_S$, i.e., $\mathcal{D}_S(z) = \sum_{z' \in \mathbb{Z}_q^k, z'_S = z} \mathcal{D}(z')$. We need to show that (40) implies that $\mathcal{D}_S = \mathcal{U}_q^{|S|}$ for all $S \subseteq [k]$ with $1 \leq |S| \leq t$.

Fix such an S and let $|S| = s$. Consider the basis $\{\chi_\alpha\}_{\alpha \in \mathbb{Z}_q^s}$. Lemma B.4 implies that it suffices to show that $\mathbf{E}_{z \sim \mathcal{U}_q^s}[D_S(z)\chi_\alpha(z)] = 0$ for all $\alpha \in \mathbb{Z}_q^s$. Observe that $\mathbf{E}_{z \sim \mathcal{U}_q^s}[D_S(z)\chi_\alpha(z)] = \mathbf{E}_{z' \sim \mathcal{U}_q^k}[\mathcal{D}(z')\chi_\sigma(z')]$ for $\sigma \in \mathbb{Z}_q^k$ such that $\sigma_i = \alpha_i$ for $i \in S$ and $\sigma_i = 0$ otherwise. Since $|S| \leq t$, we know that $|\sigma| \leq t$ and (40) implies $\mathbf{E}_{z' \sim \mathcal{U}_q^k}[\mathcal{D}(z')\chi_\sigma(z')] = 0$.

The rest of the proof is exactly as in the binary case. $\qquad\square$

We can again use these separating polynomials to obtain almost δ-refutation for predicates that are δ-far from t-wise supporting.

Theorem B.9. *Let P be δ-far from being t-wise supporting. There exists an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq 1 - \delta + \gamma$ with high probability when $\overline{m} \geq \frac{q^{O(k)} n^{t/2} \log^5 n}{\gamma^2}$ and $t \geq 2$.*

The proof is essentially identical to Proof 2 of Theorem 4.9.
Corollary 4.11 also extends to larger alphabets.

Corollary B.10. *Let P be a predicate that does not support any t-wise uniform distribution. Then there is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq 1 - 2^{-\tilde{O}(q^t k^t)}$ with high probability when $\overline{m} \geq 2^{\tilde{O}(q^t k^t)} n^{t/2} \log^5 n$ and $t \geq 2$.*

This follows directly from Theorem B.9 and the following extension of Corollary 3.17 to larger alphabets.

Corollary B.11. *Suppose $P : \mathbb{Z}_q^k \to \{0,1\}$ is not t-wise supporting. Then it is in fact δ-far from t-wise supporting for $\delta = 2^{-\tilde{O}(q^t k^t)}$.*

The proof is essentially identical to the proof of Corollary 3.17: Observe that the LP (39) has at most $q^t k^t$ variables and proceed exactly as before.

B.5   SOS proofs

Here we give SOS versions of our refutation results for larger alphabets.

Certifying Fourier coefficients are small   To give an SOS proof that Fourier coefficients of $D_{\mathcal{I},y}^b$ are small, we again need to define a specific polynomial representation of $\widehat{D_{\mathcal{I},y}^b}(\sigma)$.

$$\widehat{D_{\mathcal{I},y}}(\sigma)^{\mathrm{poly}} = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \sum_{\alpha \in \mathbb{Z}_q^{|\sigma|}} \chi_\sigma(\alpha + c(\sigma)) \prod_{i=1}^{|\sigma|} y(T(\sigma)_i, \alpha_i).$$

Lemma B.12. *Let $0 \neq \sigma \in \mathbb{Z}_q^k$ with $|\sigma| = s$. Then*

$$\{y(i,a)^2 \leq 1\}_{\substack{i \in [n] \\ a \in \mathbb{Z}_q}} \vdash_{\max\{2s,k\}} \widehat{D_{\mathcal{I},y}}(\sigma)^{\mathrm{poly}} \leq \frac{q^{O(k)} \max\{n^{s/4}, \sqrt{n}\} \log^{5/2} n}{\overline{m}^{1/2}}$$

$$\{y(i,a)^2 \leq 1\}_{\substack{i \in [n] \\ a \in \mathbb{Z}_q}} \vdash_{\max\{2s,k\}} \widehat{D_{\mathcal{I},y}}(\sigma)^{\mathrm{poly}} \geq -\frac{q^{O(k)} \max\{n^{s/4}, \sqrt{n}\} \log^{5/2} n}{\overline{m}^{1/2}}.$$

with high probability, assuming also that $\overline{m} \geq \max\{n^{s/2}, n\}$.

Proof. In the proof of Lemma B.5, we certify that $|\widehat{D_{\mathcal{I},y}}(\sigma)|$ is small by certifying that $\left|\widehat{D_{\mathcal{I},y}}(\sigma)^{\mathrm{poly}}\right|$ is small. The proof of Lemma B.5 relies only on Lemma 4.4; we can replace this with its SOS version Lemma 6.3. $\square$

Remark B.13. We stated the lemma with the weaker set of axioms $\{y(i,a)^2 \leq 1\}_{i \in [n], a \in \mathbb{Z}_q}$. Since $y(i,a)^2 = y(i,a)$ implies $y(i,a)^2 \leq 1$ in degree-2 SOS, the lemma holds with the axioms $\{y(i,a)^2 = y(i,a)\}_{i \in [n], a \in \mathbb{Z}_q}$ as well.

Strong refutation of any k-CSP From our SOS proof that the Fourier coefficients $\widehat{D^b_{\mathcal{I},y}}(\sigma)$ are small, we can get SOS proofs of strong refutation for any k-CSP. To do this, we need to define a specific polynomial representation of $\mathrm{Val}^b_{\mathcal{I}}(y)$ for an instance $\mathcal{I}$ of $\mathrm{CSP}(P)$:

$$\mathrm{Val}_{\mathcal{I}}(y)^{\mathrm{poly}} = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \sum_{\alpha \in \mathbb{Z}_q^k} P(\alpha + c) \prod_{i \in [k]} y(T_i, \alpha_i).$$

Theorem B.14. *Given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$,*

$$\{y(i,a)^2 = y(i,a)\}_{\substack{i \in [n] \\ a \in \mathbb{Z}_q}} \cup \left\{ \sum_{a \in \mathbb{Z}_q} y(i,a) = 1 \right\}_{i \in [n]} \vdash_{2k} \mathrm{Val}_{\mathcal{I}}(y)^{\mathrm{poly}} \leq \overline{P} + \gamma$$

with high probability when $\overline{m} \geq \frac{q^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.

Proof. First, use the Fourier expansion of P to write

$$\mathrm{Val}_{\mathcal{I}}(y)^{\mathrm{poly}} = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \sum_{\alpha \in \mathbb{Z}_q^k} \sum_{\sigma \in \mathbb{Z}_q^k} \widehat{P}(\sigma) \chi_\sigma(\alpha + c) \prod_{i \in [k]} y(T_i, \alpha_i).$$

For each $T \in [n]^k$, $c \in \mathbb{Z}_q^k$, and $\sigma \in \mathbb{Z}_q^k$, we have a term of the form $\sum_{\alpha \in \mathbb{Z}_q^k} \chi_\sigma(\alpha + c) \prod_{i \in [k]} y(T_i, \alpha_i)$.

Note that χ_σ only depends on the coordinates in $\mathrm{supp}(\sigma)$. We can then write this as

$$\left(\sum_{\alpha \in \mathbb{Z}_q^{|\sigma|}} \chi_\sigma(\alpha + c(\sigma)) \prod_{i=1}^{|\sigma|} y(T(\sigma)_i, \alpha_i) \right) \left(\prod_{i=1}^{k - |\sigma|} \sum_{a \in \mathbb{Z}_q} y(T(\overline{\sigma})_i, a) \right)$$

Using the axioms $\sum_{a \in \mathbb{Z}_q} y_{i,a} = 1$, the second term is equal to 1 and we have

$$\left\{ \sum_{a \in \mathbb{Z}_q} y(i,a) = 1 \right\}_{i \in [n]} \vdash_k \sum_{\alpha \in \mathbb{Z}_q^k} \chi_\sigma(\alpha + c) \prod_{i \in [k]} y(T_i, \alpha_i) = \sum_{\alpha \in \mathbb{Z}_q^{|\sigma|}} \chi_\sigma(\alpha + c(\sigma)) \prod_{i=1}^{|\sigma|} y(T(\sigma)_i, \alpha_i).$$

Summing over all T, c, and σ, we obtain the following.

$$\left\{ \sum_{a \in \mathbb{Z}_q} y(i,a) = 1 \right\}_{i \in [n]} \vdash_k \mathrm{Val}_{\mathcal{I}}(y)^{\mathrm{poly}} = \frac{1}{m} \sum_{T \in [n]^k} \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \sum_{\sigma \in \mathbb{Z}_q^k} \widehat{P}(\sigma) \sum_{\alpha \in \mathbb{Z}_q^{|\sigma|}} \chi_\sigma(\alpha + c(\sigma)) \prod_{i=1}^{|\sigma|} y(T(\sigma)_i, \alpha_i).$$

This is equal to

$$\overline{P} + \sum_{\mathbf{0} \neq \sigma \in \mathbb{Z}_q^k} \widehat{P}(\sigma) \widehat{D_{\mathcal{I},y}}(\sigma)^{\mathrm{poly}}.$$

Since $|P(z)| \leq 1$ and $|\chi_\sigma(z)| \leq q^{O(k)}$, $\sum_{\sigma \in \mathbb{Z}_q^k} |\widehat{P}(\sigma)| \leq q^{O(k)}$. We can then apply Lemma B.12 for each σ to complete the proof. $\square$

46

SOS refutation of non-t-wise supporting CSPs

Theorem B.15. *Let P be δ-far from being t-wise supporting. Then, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$,*

$$\left\{y(i,a)^2 = y(i,a)\right\}_{\substack{i\in[n]\\a\in\mathbb{Z}_q}} \cup \{y(i,a)y(i,b) = 0\}_{\substack{i\in[n]\\a\neq b\in\mathbb{Z}_q}} \cup \left\{\sum_{a\in\mathbb{Z}_q} y(i,a) = 1\right\}_{i\in[n]} \vdash_{\max\{k,2t\}} \mathrm{Val}_{\mathcal{I}}(y)^{\mathrm{poly}} \leq 1-\delta+\gamma.$$

with high probability when $\overline{m} \geq \frac{q^{O(k)}n^{t/2}\log^5 n}{\gamma^2}$ and $t \geq 2$.

To prove this theorem, we need a version of Claim 6.7 for larger alphabets.

Claim B.16. *Let $f : \mathbb{Z}_q^k \to \mathbb{R}$ such that $f(z) \geq 0$ for all $z \in \mathbb{Z}_q^k$ and let $f^b(v) = \sum_{\alpha\in\mathbb{Z}_q^k} f(\alpha)\prod_{i\in[k]} v(i,\alpha_i)$. Then*

$$\left\{v(i,a)^2 = v(i,a)\right\}_{\substack{i\in[k]\\a\in\mathbb{Z}_q}} \cup \{v_{i,a}v_{i,b} = 0\}_{\substack{i\in[k]\\a\neq b\in\mathbb{Z}_q}} \vdash_k f^b(v) \geq 0.$$

Proof. Since $f(z) \geq 0$ for all $z \in \mathbb{Z}_q^k$, there exists a function $g : \mathbb{Z}_q^k \to \mathbb{R}$ such that $g^2(z) = f(z)$ for all $z \in \mathbb{Z}_q^k$. We then write $g^b(v) = \sum_{\alpha\in\mathbb{Z}_q^k} g(\alpha)\prod_{i\in[k]} v(i,\alpha_i)$. Using $v(i,a)^2 = v(i,a)$, it follows that

$$g^b(v)^2 = \sum_{\alpha\in\mathbb{Z}_q^k} g(\alpha)^2 \prod_{i\in[k]} v(i,\alpha_i) + \sum_{\alpha'\neq\alpha''\in\mathbb{Z}_q^k} g(\alpha')g(\alpha'')\prod_{i\in[k]} v(i,\alpha_i')v(i,\alpha_i'')$$

The first term is equal to $f^b(\sigma)$. For the second term, note that each of the products $\prod_{i\in[k]} v(i,\alpha_i')v(i,\alpha_i'')$ must contain factors $v(i,a)v(i,b)$ with $a \neq b$ since $\alpha' \neq \alpha''$. We have the axiom $v(i,a)v(i,b) = 0$, so the second term is 0. Then $f^b = (g^b)^2$ and the claim follows. $\square$

With this claim, the proof of the theorem exactly follows that of Theorem 4.9.

Proof of Theorem B.15. Claim B.16 implies that

$$\left\{v(i,a)^2 = v(i,a)\right\}_{\substack{i\in[k]\\a\in\mathbb{Z}_q}} \cup \{v(i,a)v(i,b) = 0\}_{\substack{i\in[k]\\a\neq b\in\mathbb{Z}_q}} \cup \left\{\sum_{a\in\mathbb{Z}_q} v(i,a) = 1\right\}_{i\in[k]} \vdash_k P^b(v)-(1-\delta) \leq Q^b(v)$$

Summing over all constraints, we get that

$$A \vdash_k m\mathrm{Val}_{\mathcal{I}}(y)^{\mathrm{poly}} - m(1-\delta) \leq \sum_{T\in[n]^k}\sum_{c\in\mathbb{Z}_q^k} \mathbb{1}_{\{(T,c)\in\mathcal{I}\}} \sum_{\alpha\in\mathbb{Z}_q^k} Q(\alpha+c)\prod_{i=1}^k y(i,\alpha_i)$$

where $A = \left\{y(i,a)^2 = y(i,a)\right\}_{\substack{i\in[n]\\a\in\mathbb{Z}_q}} \cup \{y(i,a)y(i,b) = 0\}_{\substack{i\in[n]\\a\neq b\in\mathbb{Z}_q}} \cup \left\{\sum_{a\in\mathbb{Z}_q} y(i,a) = 1\right\}_{i\in[n]}$. Using the Fourier expansion of Q, we see that the right-hand side of the inequality is

$$\sum_{T\in[n]^k}\sum_{c\in\mathbb{Z}_q^k} \mathbb{1}_{\{(T,c)\in\mathcal{I}\}} \sum_{\alpha\in\mathbb{Z}_q^k}\sum_{\sigma\in\mathbb{Z}_q^k} \widehat{Q}(\sigma)\chi_\sigma(\alpha+c)\prod_{i=1}^k y(T_i,\alpha_i)$$

Just as in the proof of Theorem B.14, we can rewrite this in degree-k SOS as

$$\sum_{T\in[n]^k}\sum_{c\in\mathbb{Z}_q^k} \mathbb{1}_{\{(T,c)\in\mathcal{I}\}} \sum_{\sigma\in\mathbb{Z}_q^k} \widehat{Q}(\sigma) \sum_{\alpha\in\mathbb{Z}_q^{|\sigma|}} \chi_\sigma(\alpha+c(\sigma))\prod_{i=1}^{|\sigma|} y(T(\sigma)_i,\alpha_i).$$

47

We then rearrange to get

$$\sum_{\mathbf{0} \neq \sigma \in \mathbb{Z}_q^k} \widehat{Q}(\sigma) \widehat{D_{\mathcal{I},y}}(\sigma)^{\text{poly}}.$$

Since $\mathbf{E}[Q] = 0$ and $Q \geq -1$, we know that $|Q| \leq q^{O(k)}$ and therefore $|\widehat{Q}(\sigma)| \leq q^{O(k)}$. We can then apply Lemma B.12 for each σ to complete the proof. $\qquad\square$

C Certifying that random hypergraphs have small independence number and large chromatic number

First, we recall some standard definitions. Let $H = (V, E)$ be a hypergraph. We say that S is an independent set of H if for all $e \in E$, it holds that $e \notin S$. The independence number $\alpha(H)$ is then the size of the largest independent set of H. A q-coloring of H is a function $f : V \to [q]$ such that $f^{-1}(i)$ is an independent set for every $i \in [q]$. The chromatic number $\chi(H)$ is the the smallest $q \in \mathbb{N}$ for which there exists a q-coloring of H.

We define $\mathcal{H}(n, p, k)$ to be the distribution over n-vertex, k-uniform (unordered) hypergraphs in which each of the $\binom{n}{k}$ possible hyperedges is included independently with probability p. Let $\overline{m}$ be the expected number of hyperedges $p\binom{n}{k}$.

Coja-Oghlan, Goerdt, and Lanka used CSP refutation techniques to show the following results [COGL07]:

Theorem C.1. (Coja-Oghlan–Goerdt–Lanka [COGL07, Theorem 3]). *For $H \sim \mathcal{H}(n, p, 3)$, there is a polynomial time algorithm certifying that $\alpha(H) < \epsilon n$ with high probability for any constant $\epsilon > 0$ when $\overline{m} > n^{3/2} \ln^6 n$ and $\overline{m} = o(n^2)$.*

Theorem C.2. (Coja-Oghlan–Goerdt–Lanka [COGL07, implicit in Section 4]). *For $H \sim \mathcal{H}(n, p, 4)$, there is a polynomial time algorithm certifying that $\alpha(H) < \epsilon n$ with high probability for any constant $\epsilon > 0$ when $\overline{m} \geq O\left(\frac{n^2}{\epsilon^4}\right)$.*

Theorem C.3. (Coja-Oghlan–Goerdt–Lanka [COGL07, Theorem 4]). *For $H \sim \mathcal{H}(n, p, 4)$, there is a polynomial time algorithm certifying that $\chi(H) > \xi$ with high probability for constant ξ when $\overline{m} \geq O(\xi^4 n^2)$.*

We generalize these results to k-uniform hypergraphs:

Theorem C.4. *For $H \sim \mathcal{H}(n, p, k)$, there is a polynomial time algorithm certifying that $\alpha(H) < \beta$ with high probability when $\overline{m} \geq O_k\left(\frac{n^{5k/2} \log^3 n}{\beta^{2k}}\right)$, assuming that $\beta \geq n^{3/4} \log n$.*

Theorem C.5. *For $H \sim \mathcal{H}(n, p, k)$, there is a polynomial time algorithm certifying that $\chi(H) > \xi$ with high probability when $\overline{m} \geq O_k\left(\xi^{2k} n^{k/2} \log^3 n\right)$, assuming that $\xi \leq \frac{n^{1/4}}{\log n}$.*

The proofs are simple extensions of the $k = 3$ and $k = 4$ cases from [COGL07]. We will first prove Theorem C.4 using Theorem 4.1 and this will almost immediately imply Theorem C.5.

Proof of Theorem C.4. Recall that Theorem 4.1 deals with k-tuples, not sets of size k. It is easy to express a hypergraph in terms of k-tuples rather than sets of size k. For a set S and $t \in \mathbb{Z}_{\geq 0}$, recall the notation $\binom{S}{t} = \{T \subseteq S \mid |T| = t\}$. For each possible hyperedge $e \in \binom{[n]}{k}$, we associate an arbitrary tuple T_e from among the $k!$ tuples in $[n]^k$ containing the same k elements. To draw from $\mathcal{H}(n, p, k)$, we include each T_e independently with probability p and include all other $T \in [n]^k$ with probability 0.

For $T \in [n]^k$, we define the random variable $w(T)$ as follows:

$$w(T) = \begin{cases} p - 1_{\{e \in E\}} & \text{if } T = T_e \text{ for some } e \in \binom{[n]}{k} \\ 0 & \text{otherwise.} \end{cases}$$

Let $x \in \{0,1\}^n$ be the indicator vector of an independent set I so that $x^T = 1$ if $T \subseteq I$ and $x^T = 0$ otherwise. First, observe that

$$\sum_{T \in [n]^k} w(T) x^T = p \sum_{S \in \binom{[n]}{k}} x^S - \sum_{e \in \binom{[n]}{k}} 1_{\{e \in E\}} x^e = p \binom{|I|}{k},$$

where the second term is 0 because I is an independent set. The $w(T)$'s satisfy conditions (6), (7), and (8) and $\|x\|_\infty \leq 1$, so Theorem 4.1 implies we can certify that

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

with high probability. Simplifying, we see that we can certify

$$|I| \leq O_k \left(\frac{n^{5/4} \log^{\frac{3}{2k}} n}{\overline{m}^{\frac{1}{2k}}} \right)$$

and plugging in the value of $\overline{m}$ from the statement of the theorem completes the proof. $\qquad\square$

Proof of Theorem C.5. For a coloring of a hypergraph H, each color class is an independent set of H. If $\chi(H) \leq \xi$, then there exists a color class of size at least $\frac{n}{\xi}$ and therefore $\alpha(H) \geq \frac{n}{\xi}$. We can then certify that $\alpha(H) < \frac{n}{\xi}$ using Theorem C.4. $\qquad\square$

D Simulating $\mathcal{F}_P(n,p)$ with a fixed number of constraints

The setting of [DLSS14] fixes the number of constraints in a CSP instance, whereas the model described in Section 3 includes each possible constraint in the instance with some probability p. Here we show that results from our setting easily extend to that of [DLSS14] by giving an algorithm that simulates the behavior of our model when the number of constraints is fixed.

Recall that an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ is generated as follows. For each $S \in [n]^k$ and each $c \in \{-1,1\}^k$, constraint (c, S) is included with probability p, so the expected number of constraints is $p \cdot (2n)^k$.

In the model where the number of constraints is fixed, the instance is guaranteed to have m distinct constraints for some value of m. The instance $\mathcal{J}$ is chosen uniformly from all subsets of $\{-1,1\}^k \times [n]^k$ with size exactly m

Theorem D.1. *Suppose there exists an efficient algorithm R that, on a given CSP instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$, for all $p \geq p_{\min}$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq \eta$ for some $0 \leq \eta < 1$ with high probability. Then there exists an efficient algorithm $\mathcal{A}$ that certifies that a random instance $\mathcal{J}$ of $\mathrm{CSP}(P)$ with μ constraints has $\mathrm{Opt}(\mathcal{J}) \leq \eta + 2\left(\mu^{-1} \ln \mu\right)^{1/2}$ with high probability when $\mu \left(1 - \left(\mu^{-1} \ln \mu\right)^{1/2}\right) \geq (2n)^k p_{\min}$.*

Algorithm 1

Algorithm $\mathcal{A}$

1: $p \leftarrow \mu(1-d)(2n)^{-k}$.
2: draw $m \sim \text{Binomial}\left(p, (2n)^k\right)$
3: **if** $m > \mu$ or $m < \mu(1-2d)$ **then**
4:      **return** "fail."
5: $\mathcal{I} \leftarrow \mathcal{J}$
6: **for** $i = m+1 \ldots \mu$ **do**
7:      Remove a random constraint from $\mathcal{I}$ chosen uniformly
8: Run R on $\mathcal{I}$
9: **if** R certifies that $\text{Opt}(\mathcal{I}) \leq \eta$ **then**
10:      **return** "$\text{Opt}(\mathcal{J}) \leq \eta + 2d$."
11: **else**
12:      **return** "fail."

Proof. On a random instance with μ constraints, we can generate an instance $\mathcal{I}$ that simulates this behavior by choosing an appropriate value for p, drawing $m \sim \text{Binomial}\left(p, (2n)^k\right)$ and then discarding $\mu - m$ of the constraints. For brevity, let $d = \left(\mu^{-1} \ln \mu\right)^{1/2}$. Algorithm 1 describes the behavior of $\mathcal{A}$.

The fraction of removed constraints is at most $2d$, so even if all of the removed constraints would have been satisfied, their contribution to $\text{Opt}(\mathcal{J})$ is at most $2d$. Consequently, $\mathcal{A}$ will never incorrectly output "$\text{Opt}(\mathcal{J}) \leq \eta + 2d$."

Furthermore, the probability of failing to refute an instance with value at most $1 - \eta + 2d$ due to exiting at step 2 is $o_{k,t}(1)$. We treat m as a sum of $(2n)^k$ independent Bernoulli variables with probability p and denote $\mathbf{E}[m]$ by $\overline{m}$. Applying a Chernoff bound yields the following.

$$
\begin{aligned}
\mathbf{Pr}[m > \mu] &= \mathbf{Pr}[m > \overline{m}/(1-d)] \\
&= \mathbf{Pr}[m > \overline{m}(1 + \tfrac{d}{1-d})] \\
&\leq \exp\left(-\frac{\mu^{-1}\mu \ln \mu (1-(\mu \ln \mu)^{-1/2})}{3}\right) \\
&< \exp\left(-\Theta(\ln \mu)\right) = 1/\text{poly}(\mu).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathbf{Pr}[m < \mu(1-2d)] &= \mathbf{Pr}[m < \overline{m}(1 - \tfrac{d}{1-d})] \\
&\leq \exp\left(-\Theta(\ln \mu)\right) = 1/\text{poly}(\mu).
\end{aligned}
$$

If $\mu(1 - \left(\mu^{-1} \ln \mu\right)^{1/2}) \geq (2n)^k p_{\min}$, then $p \geq p_{\min}$ and R will be able to certify $\text{Opt}(\mathcal{I}) \leq \eta$ with high probability. $\qquad\square$