SOS Lower Bounds with Hard Constraints: Think Global, Act Local

Pravesh K. Kothari
Department of Computer Science, Princeton University and Institute for Advanced Study, Princeton, USA
kothari@cs.princeton.edu

Ryan O'Donnell[1]
Department of Computer Science, Carnegie Mellon University, Pittsburgh, USA
odonnell@cs.cmu.edu

Tselil Schramm[2]
Department of Computer Science, Harvard and MIT, Cambridge, USA
tselil@mit.edu

Abstract

Many previous Sum-of-Squares (SOS) lower bounds for CSPs had two deficiencies related to global constraints. First, they were not able to support a "cardinality constraint", as in, say, the Min-Bisection problem. Second, while the pseudoexpectation of the objective function was shown to have some value β, it did not necessarily actually "satisfy" the constraint "objective $= \beta$". In this paper we show how to remedy both deficiencies in the case of random CSPs, by translating *global* constraints into *local* constraints. Using these ideas, we also show that degree-$\Omega(\sqrt{n})$ SOS does not provide a $(\frac{4}{3} - \varepsilon)$-approximation for Min-Bisection, and degree-$\Omega(n)$ SOS does not provide a $(\frac{11}{12} + \varepsilon)$-approximation for Max-Bisection or a $(\frac{5}{4} - \varepsilon)$-approximation for Min-Bisection. No prior SOS lower bounds for these problems were known.

2012 ACM Subject Classification Theory of computation → Semidefinite programming, Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases sum-of-squares hierarchy, random constraint satisfaction problems

Digital Object Identifier 10.4230/LIPIcs.ITCS.2019.49

Related Version `https://arxiv.org/abs/1809.01207`

Acknowledgements The authors very much thank Sangxia Huang and David Witmer for their contributions to the early stages of this research. Thanks also to Svante Janson for discussions concerning contiguity of random graph models. We also greatfully acknowledge comments on the manuscript from Johan Håstad as well as several anonymous reviewers.

[1] Some work performed at the Boğaziçi University Computer Engineering Department, supported by Marie Curie International Incoming Fellowship project number 626373. Also supported by NSF grants CCF-1618679, CCF-1717606. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

[2] This work was partly supported by an NSF Graduate Research Fellowship (1106400), and also by a Simons Institute Fellowship.

1 Introduction

Consider the task of refuting a random 3SAT instance with n variables and $50n$ clauses; i.e., certifying that it's unsatisfiable (which it is, with very high probability). There is no known $2^{o(n)}$-time algorithm for this problem. An oft-cited piece of evidence for the exponential difficulty is the fact [9, 19] that the very powerful Sum-of-Squares (SOS) SDP hierarchy fails to refute such random 3SAT instances in $2^{o(n)}$ time. Colloquially, degree-$\Omega(n)$ SOS "thinks" that the random 3SAT instance is satisfiable (with high probability).

But consider the following method of refuting satisfiability of a random $50n$-clause CNF ϕ:

> For all $k \in \{0, 1, 2, \ldots, n\}$,
> refute "ϕ is satisfiable by an assignment of Hamming weight k".

Could it be that $O(1)$-degree SOS succeeds in refuting random 3SAT instances in this manner? It seems highly unlikely, but prior to this work the possibility could not be ruled out.

SOS lower bounds with Hamming weight constraints

Recall that the known SOS lower bounds for random 3SAT are actually stronger: they show degree-$\Omega(n)$ SOS thinks that random 3SAT instances are satisfiable even as *as 3XOR* (i.e., with every clause having an odd number of true literals). Hamming weight calculations are quite natural in the context of random 3XOR; indeed Grigoriev, Hirsch, and Pasechnik [10] showed that the *dynamic* degree-5 SOS proof system *can* refute random 3XOR instances by using integer counting techniques. Thus the above "refute solutions at each Hamming weight" strategy seems quite natural in the context of random CSPs.

In 2012, Yuan Zhou raised the question of proving strong SOS lower bounds for random 3XOR instances together with a global cardinality constraint such as $\sum_i x_i = \frac{n}{2}$. This would rule out the above refutation strategy. It is also a natural SOS challenge, seemingly combining the two strong SOS results known prior to 2012 – the bound for random 3XOR due to Grigoriev and Schoenebeck [9, 19] and the bound for Knapsack due to Grigoriev [8].

One may ask why the Grigoriev–Schoenebeck SOS lower bound doesn't already satisfy $\sum_i x_i = \frac{n}{2}$. The difficulty is connected to the meaning of the word "satisfy". One should think of the SOS Method as trying to find not just a satisfying assignment to a CSP, but more generally a *distribution* on satisfying assignments. The SOS algorithm finds a "degree-d pseudodistribution" on satisfying assignments in $n^{O(d)}$ time, provided one exists; roughly speaking, this means an object that "looks like" a distribution on satisfying assignment to all tests that are squared polynomials of degree at most d. For a random 3XOR instance with n variables and $O(n)$ constraints, the Grigoriev–Schoenebeck degree-$\Omega(n)$ pseudodistribution indeed claims to have 100% of its probability mass on satisfying assignments. Furthermore, its assignments claim to give probability 50% to each of $x_i = 0$ and $x_i = 1$ for all i; in other words, the "pseudoexpectation" of x_i is $\frac{1}{2}$, so the pseudoexpectation of $\sum_i x_i$ is $\frac{n}{2}$. However, this *doesn't* mean that the pseudodistribution "satisfies" the hard constraint $\sum_i x_i = \frac{n}{2}$. To actually "satisfy" this constraint, the expression $\sum_i x_i$ must have *pseudovariance zero*; i.e., SOS must not only "think" it knows a distribution on 3XOR-satisfying assignments which has $\sum_i x_i = \frac{n}{2}$ on average, it must think that *all* of these satisfying assignments have $\sum_i x_i$ exactly $\frac{n}{2}$.

In this work we show how to upgrade any SOS lower bound for random CSPs based on t-wise uniformity so as to include the hard cardinality constraint $\sum_i x_i = \frac{n}{2}$ (or indeed

$\sum_i x_i = \frac{n}{2} + k$ for any $|k| = O(\sqrt{n})$).[3] The idea is conceptually simple: just add a matching of 2XOR constraints, $x_{2i-1} \neq x_{2i}$ for all $1 \leqslant i \leqslant \frac{n}{2}$.

SOS lower bounds with exact objective constraints

A random 3AND CSP with n variables and $m = \alpha n$ constraints (each an AND of 3 random literals) will have objective value $\frac{1}{8} + \varepsilon$ with high probability, for ε arbitrarily small as a function of α; i.e., the best assignment will satisfy at most $(\frac{1}{8} + \varepsilon)m$ constraints. On the other hand, it's not too hard to show that the Grigoriev–Schoenebeck degree-$\Omega(n)$ pseudodistribution will give the objective function a pseudoexpectation of $\frac{1}{4} \pm o(1)$. (Roughly speaking, for almost all 3AND constraints, the SOS pseudodistribution will think it can obtain probability $\frac{1}{4}$ on each of the 3XOR-satisfying assignments, and one of these, namely $(1, 1, 1)$, satisfies 3AND.) Thus it would appear that degree-$\Omega(n)$ SOS has an integrality gap of factor $2 - \varepsilon$ on random 3AND instances.

But is this misleading? Suppose we solved the SOS SDP and it reported a solution with pseudoexpectation $\frac{1}{4}$. We might then "double-check" by re-running the SDP, together with an additional "equality constraint" specifying that the number of satisfied 3AND constraints is indeed $\frac{1}{4}m$.[4] As far as we know now, this run could return "infeasible", actually *refuting* the possibility of $\frac{1}{4}m$ constraints being satisfiable! Again, the issue is that under the Grigoriev–Schoenebeck SOS pseudodistribution, the objective function will have a pseudoexpectation like $\frac{1}{4}$, but will also have *nonzero pseudovariance*.

We show how to fix this issue – i.e., have the objective constraint be *exactly* SOS-satisfied – in the context of any SOS lower bound for random CSPs based on t-wise uniformity. Here we briefly express the idea of our solution, in the specific case of 3AND: We show that one can design a probability distribution θ on $r \times 3$ Boolean matrices such that two properties hold: (i) θ is 2-wise uniform; (ii) for *every* outcome in the support of θ, *exactly* a $\frac{1}{4} - \varepsilon_r$ fraction of the r rows satisfy 3AND, where ε_r is an explicit positive constant depending on r that tends to 0 as r grows. We then use recent work [14] on constructing SOS lower bounds from t-wise uniform distributions to show that degree-$\Omega(n)$ SOS thinks it can "weakly satisfy" a random "distributional CSP" in which each constraint specifies that a random $4r$-tuple of variables should be distributed according to θ. By "weak satisfaction", we mean that SOS will at least think it can get a local distribution on each $4r$-tuple whose support is contained within θ's support (and therefore always having exactly a $\frac{1}{4} - \varepsilon_r$ fraction of rows satisfying 3AND). Now viewing each such tuple as the conjunction of r (random) 3AND constraints, we get that the SOS solution thinks it satisfies exactly a $\frac{1}{4} - \varepsilon_r$ fraction of these constraints.

Further consequences

Via our first result – satisfying global cardinality constraints – we open up the possibility of establishing SOS lower bounds for natural problems like Min- and Max-Bisection (by performing reductions within SOS, as in [21]). Previously, no such SOS integrality gaps were known (Guruswami, Sinop, and Zhou [11] had given an SOS integrality gap approaching $\frac{11}{10}$ for the Balanced-Separator problem, which is like Min-Bisection but without a hard bisection constraint.) Under assumptions like $\mathsf{NP} \not\subseteq \bigcap_{\varepsilon>0} \mathsf{TIME}(2^{n^\varepsilon})$, some hardness results were

[3] We also show in the full version that this is not too far from tight, in the sense that it is easier to refute XOR with Hamming weight constraints that are too imbalanced (if $k = \omega(n^{1/4})$).

[4] Actually, it was recently observed that it is not clear we can definitely solve the associated SDP exactly [16, 18]. This does not affect the status of our lower bounds.

previously known: no PTAS for Min-Bisection (due to Khot [13]) and factor $\frac{15}{16}+\varepsilon$ hardness for Max-Bisection (due to Holmerin and Khot [13], improving on the factor $\frac{16}{17}+\varepsilon$ NP-hardness known for Max-Cut). However, in the context of SOS lower bounds, it makes sense to shoot for more: namely, hardness factors that are known subject to Feige's R3SAT Hypothesis [7] (and similar hypotheses for random CSPs).

Feige himself [7] showed factor $\frac{4}{3}-\varepsilon$ hardness for Min-Bisection under his hypothesis (with a quadratic size blowup). Also, it's possible to show factor $\frac{11}{12}+\varepsilon$ hardness for Max-Bisection (with linear size blowup) under Feige's Hypothesis for 4XOR; this is arguably "folklore", via the gadget techniques of Trevisan et al. [20] (see also [12, 17]). We are able to convert both of these results to SOS lower bounds, showing that degree-$\Omega(\sqrt{n})$ SOS fails to $(\frac{4}{3}-\varepsilon)$-approximate Min-Bisection, and degree-$\Omega(n)$ SOS fails to $(\frac{5}{4}+\varepsilon)$-approximate Min-Bisection. Our proof of the latter can also be modified to show that degree-$\Omega(n)$ SOS fails to $(\frac{11}{12}+\varepsilon)$-approximate Max-Bisection.

It is worth pointing out that the benefit of our second main result, the ability to enforce objective equality constraints exactly, also arises in these SOS Bisection lower bounds. For example, the $(\frac{4}{3}-\varepsilon)$-hardness for Min-Bisection is a kind of gadget reduction from random 3AND CSPs; showing that the "good cut" in the completeness case is an exact bisection relies on the "good assignment" in the 3AND instance satisfying exactly a $\frac{1}{4}$ fraction of constraints.

1.1 Statement of main theorems

Recent work [3, 14] has established a general framework for showing lower bounds for SOS on random CSPs, using the idea of t-wise uniformity. The following is a fairly general example of what's known:

▶ **Theorem 0** ([14]). *Let $P:\{0,1\}^k\to\{0,1\}$ be a predicate, and suppose there is a $(t-1)$-wise uniform distribution ν on $\{0,1\}^k$ with $\mathbb{E}_\nu[P]=\beta$. Consider a random n-variable, $m=\Delta n$-constraint instance of CSP($P^{\pm}$), meaning that each constraint is P applied to k randomly chosen literals. Then with high probability, there is a degree-$\Omega\left(\frac{n}{\Delta^{2/(t-2)}\log\Delta}\right)$ SOS pseudoexpectation $\widetilde{\mathbb{E}}[\cdot]$ with the following property:*
Case 1: *$\beta=1$. In this case, $\widetilde{\mathbb{E}}[\cdot]$ satisfies all the CSP constraints as identities.*
Case 2: *$\beta<1$. In this case, $\widetilde{\mathbb{E}}[\mathrm{OBJ}(x)]=\beta\pm o(1)$, where $\mathrm{OBJ}(x)$ denotes the objective value of the CSP.*

For example, the case of random 3SAT described in the previous section corresponds to $P=\mathrm{OR}_3$, $t=3$, ν being the uniform distribution on triples satisfying XOR_3, $\beta=1$, and $\Delta=50$; the case of random 3AND has the same t, ν, and Δ, but $P=\mathrm{AND}_3$ and $\beta=\frac{1}{4}$.

Our main theorems are now as follows:

▶ **Theorem 1.** *In the $\beta=1$ case of Theorem 0, one can additionally get the pseudodistribution $\widetilde{\mathbb{E}}$ to satisfy (with pseudovariance zero) the global bisection constraint $\sum_{i=1}^n x_i=\frac{n}{2}$ (assuming n even). More generally, for any integer $B\in[\frac{n}{2}-O(\sqrt{n}),\frac{n}{2}+O(\sqrt{n})]$, we can ensure the pseudodistribution satisfies the global Hamming weight constraint $\sum_{i=1}^n x_i=B$.*

▶ **Theorem 2.** *In the $\beta<1$ case of Theorem 0, there exists a sequence of positive constants ε_r with $\varepsilon_r\to 0$ such that for a random* n-variable, $m=\Delta n$-constraint instance of CSP($P^{\pm}$), with high probability there is a degree-$\Omega_r\left(\frac{n}{\Delta^{2/(t-2)}\log\Delta}\right)$ SOS pseudodistribution $\widetilde{\mathbb{E}}$ which satisfies (with pseudovariance zero) the hard constraint "$\mathrm{OBJ}(x)=\beta-\varepsilon_r$". Furthermore, we can also obtain cardinality constraints as in Theorem 1.*

In the full version, we show that Theorem 1 is not too far from tight, by demonstrating that random k-XOR instances become easier to refute when one imposes an imbalanced Hamming weight constraint $\sum_{i=1}^{n} x_i = \frac{n}{2} \pm \omega(n^{1/4})$.[5]

▶ Remark. In the above theorem we have written "random*" with an asterisk because the random instance is not drawn precisely in the standard way. Rather, it is obtained by choosing m/r groups of random constraints, where in each group we fix a literal pattern and then choose r nonoverlapping constraints with this pattern. This technicality is an artifact of our proof; it seems likely that it is unnecessary. Indeed, it is possible that these two distributions on random hypergraphs are simply $o(1)$-close in total variation distance, at least when $m = O(n)$.[6] In any case, by alternate means (including the techniques from Theorem 1) we are able to show the following alternative result in Section 5.2: When $m = o(n^{1.5})$, with high probability a purely random instance of CSP($P^{\pm}$) has an SOS pseudodistribution of the stated degree that exactly satisfies $\mathrm{OBJ}(x) = \beta - \varepsilon$ for some $\varepsilon > 0$ that can be made arbitrarily small.

▶ Remark. Our proof of Theorem 2 only relies on the "Case 1, $\beta = 1$" part of [14]'s Theorem 0. In fact, our Theorem 2 can actually be used to effectively *deduce* "Case 2, $\beta < 1$" from "Case 1, $\beta = 1$" in Theorem 0. This is of interest because [14]'s argument for Case 2 was not a black-box reduction from Case 1, but instead involved verifying a more technical expansion property in random graphs, as well as slightly reworking the proof of Case 2.

Finally, we obtain the following theorems concerning Bisection problems:

▶ **Theorem 3.** *For the Max-Bisection problem in a graph on n vertices, for $d = \Omega(n)$, the degree-d Sum-of-Squares Method cannot obtain an approximation factor better than $\frac{11}{12} - \varepsilon$ for any constant $\varepsilon > 0$.*

For the Min-Bisection problem, for $d' = \Omega(\sqrt{n})$, the degree-d' SOS Method cannot obtain an approximation factor better than $\frac{4}{3} - \varepsilon$, and for $d = \Omega(n)$ the degree-d SOS Method cannot obtain an approximation factor better than $\frac{5}{4} - \varepsilon$.

The proofs are included in the full version.

Organization of this paper

In Section 2, we provide some preliminaries and technical context for the study of CSPs and SOS. In Section 3, we extend the results of [14] to obtain lower bounds for CSPs with global cardinality constraints, proving Theorem 1. Section 4 shows how to construct local distributions over assignments to groups of disjoint predicates so that the number of satisfied constraints is always exactly the same, and Section 5 shows how to use such distributions to prove Theorem 2. In Section 5, one can also find a discussion of random vs. random* CSPs. We wrap up with some concluding remarks and future directions in Section 6.

2 Preliminaries

CSPs

A constraint satisfaction problem (CSP) is defined by an *alphabet* Ω (usually $\{0,1\}$ or $\{\pm 1\}$ in this paper) and a collection $\mathcal{P}$ of *predicates*, each predicate being some $P : \Omega^k \to \{0,1\}$ (with different P's possibly having different *arities*, k). An *instance* $\mathcal{H}$ consists of a set V of

[5] We conjecture that Theorem 1 is tight, and that Hamming weight constraints with imbalance $\omega(n^{1/2})$ already make k-XOR easier to refute.

[6] Thanks to Svante Janson for some observations in the direction of showing this.

n variables, as well as m constraints. Each constraint h consists of a *scope* S and a predicate $P \in \mathcal{P}$, where S is a tuple of k distinct variables, k being the arity of P. An *assignment* gives a value $x_i \in \Omega$ to the ith variable; it *satisfies* constraint $h = (S, P)$ if $P(x_{S_1}, \ldots, x_{S_k}) = 1$. We may sometimes write this as $P(x_S) = 1$ for brevity. The associated *objective value* is the fraction of satisfied constraints,

$$\mathrm{OBJ}(x) = \operatorname*{avg}_{h=(S,P)\in\mathcal{H}} \{P(x_S)\}.$$

Sometimes we are concerned with CSPs of the following type: the alphabet $\Omega = \{\pm 1\}$ is Boolean, there is a single predicate $P : \Omega^k \to \{0, 1\}$ (e.g., $P = \mathrm{OR}_3$, the 3-ary Boolean OR predicate), and the predicate set $\mathcal{P}$ consists of all 2^k versions of P in which inputs may be negated. We refer to this scenario as P-CSP with *literals*, denoted $\mathrm{CSP}(P^{\pm})$. For example, the case of $P = \mathrm{OR}_3$ is the classic "3SAT" CSP.

Distributional CSPs

A *distributional CSP* is one where, rather than having a predicate associated with each scope, we have a probability distribution. More precisely, each *distributional constraint* $h = (S, \nu)$ now consists of a scope S of some arity k, as well as a probability distribution ν on Ω^k. The optimization task involves finding a *global probability distribution* μ on assignments. We say that μ *satisfies* constraint $h = (S, \nu)$ if the marginal $\mu|_S$ of μ on S is equal to ν; we say the distributional CSP is *satisfiable* if there is a μ satisfying all constraints.

We may also say that μ *weakly satisfies* $h = (S, \nu)$ if $\mathrm{supp}(\mu|_S) \subseteq \mathrm{supp}(\nu)$. A "usual" (*predicative*, i.e., non-distributional) CSP can be viewed as a distributional CSP as follows: For each predicate P, select any distribution ν_P whose support is exactly the satisfying assignments to P; then the existence of a global assignment in the predicative CSP of objective value β is equivalent to the existence of a global probability distribution μ that *weakly* satisfies a β fraction of constraints.

Random CSPs

We are frequently concerned with CSPs chosen uniformly at random. Given a predicate set $\mathcal{P}$, a random CSP with n variables and m constraints is chosen as follows: For each constraint we first choose a random $P \in \mathcal{P}$. Supposing it has arity k, we then choose a uniformly random length-k scope S from the n variables, and impose the constraint (S, P). We can similarly define a random distributional CSP given a collection $\mathcal{D}$ of distributions ν. We remark that our choice of having exactly m constraints is not really essential, and not much would change if we had, e.g., a Poisson(m) number of random constraints, or if we chose each possible constraint independently with probability such that m constraints are expected.

SOS

The SOS Method [5] can be thought of as an algorithmic technique for finding upper bounds on the best objective value achievable in a predicative or distributional CSP. For example, in a random 3SAT instance with $m = 50n$, it is very likely that every assignment x has $\mathrm{OBJ}(x) \leqslant \frac{7}{8} + o(1)$; ideally, the SOS Method could certify this, or could at least certify unsatisfiability, meaning an upper bound of $\mathrm{OBJ}(x) < 1$ for all assignments. The SOS Method has a tunable *degree* parameter d; increasing d increases the effectiveness of the method, but also its run-time, which is essentially $n^{O(d)}$ (though see [16, 18] for a more

precise discussion). In this work we are only concerned with showing negative results for the power of SOS. Showing that degree-d SOS fails to certify a good upper bound on the maximum objective value is equivalent to showing that a *degree-d pseudodistribution* exists under which the objective function has a large *pseudoexpectation*. We define these terms now.

For simplicity we restrict attention to CSPs with Boolean alphabet (either $\Omega = \{0,1\}$ or $\Omega = \{\pm 1\}$), although it straightforward to extend the definitions for larger alphabets.[7] The SOS method introduces *indeterminates* $X_1, \dots, X_n$ associated to the CSP variables; intuitively, one thinks of them as standing for the outcome of a global assignment chosen from a supposed probability distribution on assignments. An associated *degree-d pseudoexpectation* is a real-valued linear map $\widetilde{\mathbb{E}}$ on $\mathbb{R}_{\leqslant d}[X_1, \dots, X_n]$ (the space of formal polynomials in $X_1, \dots, X_n$ of degree at most d) satisfying three properties:

1. $\widetilde{\mathbb{E}}[\text{multilin}(Q(X))] = \widetilde{\mathbb{E}}[Q(X)]$; here multilin$(Q(X))$ refers to the *multilinearization* of $Q(X)$, meaning the reduction mod $X_i^2 = X_i$ (in case $\Omega = \{0,1\}$) or mod $X_i^2 = 1$ (in case $\Omega = \{\pm 1\}$).
2. $\widetilde{\mathbb{E}}[1] = 1$;
3. $\widetilde{\mathbb{E}}[Q(X)^2] \geqslant 0$ whenever $\deg(Q) \leqslant d/2$.

We tend to think of the first condition (as well as the linearity of $\widetilde{\mathbb{E}}$) as being "syntactically" enforced; i.e., given $\widetilde{\mathbb{E}}$'s values on the multilinear monomials, its value on all polynomials is determined through multilinearization and linearity. It is not hard to show that every pseudoexpectation $\widetilde{\mathbb{E}}$ arises from a *signed probability distribution* μ; i.e., a (possibly negative) function $\mu : \Omega^n \to \mathbb{R}$ with $\sum_x \mu(x) = 1$. We call this the associated *pseudodistribution*. Intuitively, we think of a degree-d pseudodistribution as a "supposed" distribution on global assignments, which at least passes the tests in Item 3 above.

Given a CSP instance $\mathcal{H}$, if there is a degree-d pseudodistribution with $\widetilde{\mathbb{E}}[\text{OBJ}(X)] \geqslant \beta$, this means that the degree-d SOS Method *fails* to certify an upper bound of $\text{OBJ}(x) < \beta$ for the CSP. Informally, we say that degree-d SOS "thinks" that there is a distribution on assignments under which the average objective value is at least β. Similarly, given a distributional CSP $\mathcal{H}$, if there is a degree-d pseudodistribution in which $\widetilde{\mathbb{P}}[X_S = (a_1, \dots, a_k)] = \nu(a_1, \dots, a_k)$ for all constraints $h = (S, \nu)$, we say that degree-d SOS "thinks" that $\mathcal{H}$ is fully satisfiable. Here $\widetilde{\mathbb{P}}[X_S = (a_1, \dots, a_k)]$ means $\widetilde{\mathbb{E}}[1_{X_S = (a_1, \dots, a_k)}]$, where $1_{X_S = (a_1, \dots, a_k)}$ denotes the natural arithmetization of the 0-1 indicator as a degree-k multilinear polynomial.

Satisfaction of identities in SOS

Formally speaking, one says that a degree-d pseudodistribution *satisfies an identity* $Q(X) = b$ if $\widetilde{\mathbb{E}}[(Q(X) - b)R(X)] = 0$ for all polynomials $R(X)$ of degree at most $d - \deg(Q)$. Note that this is *stronger* than simply requiring $\widetilde{\mathbb{E}}[Q(X)] = b$ (the $R \equiv 1$ case). A great deal of this paper is concerned with precisely this distinction; it may be relatively easy to come up with a degree-d pseudodistribution over $\{0,1\}^n$ satisfying, say, $\widetilde{\mathbb{E}}[\sum_i X_i] = \frac{n}{2}$, but much harder to find one that "satisfies the identity $\sum_i X_i = \frac{n}{2}$". The terminology here is a little unfortunate; we will try to ameliorate things by introducing the following stronger phrase:

▶ **Definition 4.** We say that a degree-d pseudodistribution satisfies identity $Q(X) = b$ *with pseudovariance zero* if we have both $\widetilde{\mathbb{E}}[Q(X)] = b$ and *also*

$$\widetilde{\mathbb{VAR}}[Q(X) - b] = \widetilde{\mathbb{E}}[Q(X)^2] - b^2 = 0.$$

[7] Specifically, for each variable x and each alphabet element $a \in \Omega$, one introduces an indeterminate called $1_{x=a}$ that is constrained as a $\{0,1\}$ value and is interpreted as the indicator of whether x is assigned a.

As is shown in [2, Lemma 3.5 (SOS Cauchy-Schwarz)], this condition is equivalent to the pseudodistribution "satisfying the identity $Q(X) = b$" for Q of degree up to $d/2$.[8]

Intuitively, in this situation degree-d SOS not only "thinks" that it knows a distribution on assignments x under which $Q(x)$ has expectation b, it further thinks that *every* outcome x in the support of its supposed assignment has $Q(x) = b$.

3 Random CSPs with Hamming weight constraints

In this section we will prove Theorem 1, which extends the known random CSP lower bounds (as in Theorem 0) to CSPs with a hard Hamming weight constraint on the variable assignment.

3.1 Hypergraph expansion and prior SOS lower bounds for random CSPs

The paper [14] works in the general setting of distributional CSPs with an upper bound of K on all constraint arities. An instance is thought of as a "factor graph" G: a bipartite graph with n variable-vertices, m constraint-vertices, and edges joining a constraint-vertex to the variable-vertices in its scope. More precisely, the neighborhood $N(h)$ of each constraint-vertex h is defined to be an *ordered* tuple of k_h variable-vertices. We write ν_h for the local probability distribution on $\Omega^{N(h)}$ associated to constraint h. In [14], each ν_h is assumed to be a $(\tau - 1)$-wise uniform distribution, where τ is a global integer parameter satisfying $3 \leqslant \tau \leqslant K$. Finally, the graph G is assumed to satisfy a certain high-expansion condition (discussed in the full version) called the "Plausibility Assumption" involving two parameters $0 < \zeta < 1$ and $1 \leqslant \mathtt{SMALL} \leqslant n/2$, assumed to satisfy $K \leqslant \zeta \cdot \mathtt{SMALL}$. In this case, the main theorem of [14] is that there is a SOS-pseudodistribution of degree $\frac{1}{3}\zeta \cdot \mathtt{SMALL}$ that weakly satisfies all constraints.

In [14] it is assumed that all constraint distributions ν_h have the *same* level of uniformity, namely $(\tau - 1)$-wise uniformity, $\tau \geqslant 3$. In this work, in order to incorporate Hamming weight constraints on the assignment, we would like to consider the possibility that different constraint distributions have different levels of uniformity. To that end, suppose that each ν_h is $(t_h - 1)$-wise uniform, where the t_h's are various integers. Slightly more broadly than [14], we allow $1 \leqslant t_h \leqslant k_h + 1$ for all h, and we allow the constraints to have arity k_h as low as 1.

In the full version, we examine how these assumptions affect the proofs in [14]. The upshot is Theorem 5 below. Before we give the theorem, we briefly introduce some notation and comments: A "constraint-induced" subgraph H is a subgraph of the factor graph G given by choosing some set of constraints C, as well as all edges and constraint-vertices adjacent to C. We write $c(H)$ for the number of constraints in H, $e(H)$ for the number of edges, $v(H)$ for the number of variable-vertices, and $T(H) = \sum_{h \in \mathrm{cons}(H)} t_h$. To reduce to (3.1) in the following theorem, we use the observation in the full version that adding edges to a subgraph to make it constraint-induced can only decrease "income".

For notational simplicity we have also adjusted the parameters ζ and $\mathtt{SMALL}$ by factors of 2.

[8] In this paper we are flexible when it comes to constant factors in the degree. For this reason we need not worry about this factor-2 loss in the degree, as a degree-$2d$ pseudoexpectation which satisfies an identity with pseudovariance 0 automatically gives a degree-d pseudoexpectation which satisfies the identity exactly.

▶ **Theorem 5** (Essentially from [14]). *Let $0 < \zeta < 1$, $\mathrm{SMALL} \leqslant n$ and assume all constraint-vertices in G have arity at most $\zeta \cdot \mathrm{SMALL}$.*

Suppose that for every set of nonempty constraint-induced subgraph H with $c(H) \leqslant \mathrm{SMALL}$, it holds that

$$v(H) \geqslant e(H) - \frac{T(H)}{2} + \zeta c(H). \tag{3.1}$$

Then there is an SOS-pseudodistribution of degree $\frac{1}{3}\zeta \cdot \mathrm{SMALL}$ that weakly satisfies all constraints.

There are a lot of parameters in the above theorem, and our goal is not to derive the most general possible quantitative result. Instead we'll simply work out some of the basic consequences.

A basic setting treated in [14], relevant for Theorem 0, is the following. For a fixed small t we choose a random CSP with n variables and Δn constraints, with each constraint supporting a $(t-1)$-wise uniform distribution. E.g., in random 3SAT, $t = 3$. Then if

$$\Delta = \mathrm{const} \cdot \left(\frac{n}{\mathrm{SMALL}}\right)^{\frac{t}{2}-1-\zeta}$$

for a sufficiently small positive constant, it is shown in [14] that the main condition (3.1) indeed holds with high probability. Choosing, say, $\zeta = \frac{1}{\log n}$ and $\mathrm{SMALL} = \mathrm{polylog}(n)$, we see that, with high probability, we will have weakly satisfying pseudodistributions of degree $\mathrm{polylog}(n)$ even when $\Delta = \widetilde{\Theta}(n^{t/2-1})$.

In fact, it's possible to show that we have such pseudoexpectations when there are, simultaneously, $n^{1.5}/\mathrm{polylog}(n)$ 2-wise-supporting constraints, *and* $n^2/\mathrm{polylog}(n)$ 3-wise-supporting constraints, *and* $n^{2.5}/\mathrm{polylog}(n)$ 4-wise-supporting constraints, *... and also* $n/\mathrm{polylog}(n)$ 1-wise-supporting constraints, and $n^{.5}/\mathrm{polylog}(n)$ 0-wise-supporting constraints.

3.2 Expansion in the presence of matching and unary constraints

However, if we want to impose a cardinality constraint by way of adding 1-wise independent 2-ary $\neq$ constraints, then $n/\mathrm{polylog}(n)$ such constraints will not suffice. Indeed, what we would like to now show is that if the 1-wise-supporting constraints are carefully chosen to not overlap, we can add a full, linear-sized "matching" of them without compromising the lower bound. Then, when $\Omega = \{0, 1\}$, we can impose the 1-wise-uniform constraints $x_1 \oplus x_2 = 1$, $x_3 \oplus x_4 = 1, \ldots, x_{n-1} \oplus x_n = 1$ and thereby force the pseudoexpectation to satisfy the global constraint $\sum_i x_i = \frac{n}{2}$.

▶ **Theorem 6.** *Fix a uniformity parameter $3 \leqslant t \leqslant O(1)$, an arity $k \leqslant O(1)$, a number $U = O(\sqrt{n})$ of "unary" constraints, and a small failure probability $0 < p < 1/2$. Assume also that $\zeta \leqslant 1/2$.*

Suppose we form a random factor graph with n variable-vertices and Δn constraint-vertices $\mathcal{C}$ of arity k; assume each constraint-vertex is equipped with an associated $(t-1)$-wise uniform distribution.

Furthermore, suppose we add in two sets $\mathcal{M}_1$, $\mathcal{M}_2$ of nonrandom, nonoverlapping constraints, whose associated variable vertices partition $[n]$. The "unary" constraints of $\mathcal{M}_1$ should satisfy $|\mathcal{M}_1| \leqslant U$ and have an associated 0-wise uniform distribution; the "matching" constraints of $\mathcal{M}_2$ should be of constant arity and have an associated 1-wise uniform distribution.

Then provided

$$\Delta \leqslant const \cdot p \cdot \left(\frac{n}{\text{SMALL}}\right)^{\frac{t}{2}-1-\zeta'}$$

for a sufficiently small universal constant, and $\zeta' = (k+1)\zeta$, the expansion condition in Theorem 5 holds except with probability at most p.

The proof of Theorem 6, appearing in the full version, uses standard combinatorial techniques for verifying the expansion of random graphs. Due to the fact that the unary and matching constraints are deterministic, we must augment these standard techniques with some straightforward case analysis. In fact, we only prove the theorem under the assumption that the constraints of $\mathcal{M}_2$ have arity 2; the more general case of constant arity is a slight elaboration that we omit.

3.3 Lower bound for CSPs with Hamming weight constraints

As in [14], we observe that for a given $\Delta \geqslant 10$, a good choice for ζ is $\frac{1}{\log \Delta}$. This yields the following corollary, which we will show implies Theorem 1:

▶ **Theorem 7.** *Let ν be a $(t-1)$-wise uniform distribution on $\{0,1\}^k$. Consider a random n-variable, $m = \Delta n$-constraint k-ary distributional CSP, in which each constraint distribution is ν up to a negation pattern in the k inputs. (All such "reorientations" are still $(t-1)$-wise uniform.) Suppose we also impose the following nonrandom distributional constraints:*

- *The 0-wise uniform constraints $x_1 = b_1, x_2 = b_2, \ldots, x_U = b_U$, for some string $b \in \{0,1\}^U$ with $U = O(\sqrt{n})$;*
- *The 1-wise uniform constraint that (x_{U+1}, x_{U+2}) is uniform on $\{(0,1),(1,0)\}$, and similarly for the pairs (x_{U+3}, x_{U+4}), $\ldots$, (x_{n-1}, x_n).*

Then with high probability, there is an SOS-pseudodistribution of degree $D = \Omega\left(\frac{n}{\Delta^{2/(t-2)} \log \Delta}\right)$ that weakly satisfies all constraints.

Let us now see why this implies Theorem 1. In this "$\beta = 1$" scenario, we have a $(t-1)$-wise uniform ν supported on the satisfying assignments for P. Whenever the random CSP($P^{\pm}$) instance has a P-constraint with a particular literal pattern, we impose the analogous ν-constraint with equivalent negation pattern. Now the SOS-pseudodistribution $\widetilde{\mathbb{E}}[\cdot]$ promised by Theorem 6 weakly satisfies all these ν-constraints, and hence satisfies all the P-constraints. Furthermore, it also has $\widetilde{\mathbb{E}}[x_i] = b_i \in \{0,1\}$ for all $i \leqslant U$, and $\widetilde{\mathbb{E}}[x_i(1-x_{i+1})]+\widetilde{\mathbb{E}}[(1-x_i)x_{i+1}] = 1$ for all pairs $(i, i+1) = (U+1, U+2), \ldots, (n-1, n)$, by weak satisfaction. Notice that the latter implies

$$\widetilde{\mathbb{E}}[(x_i + x_{i+1} - 1)^2] = 1 - \widetilde{\mathbb{E}}[x_i] - \widetilde{\mathbb{E}}[y_i] + 2\,\widetilde{\mathbb{E}}[x_i x_{i+1}] = 0,$$

and hence the SOS solution satisfies $x_i + x_{i+1} = 1$ with pseudovariance zero. Similarly (and easier), it satisfies the identity $x_i = b_i$ for all $i \leqslant U$. It now follows that the pseudodistribution satisfies the identity $\sum_{i=1}^{n} x_i = \frac{n}{2} + (|b| - \frac{U}{2})$ with pseudovariance zero, and this completes the proof of Theorem 1, because we can take any $|b| \in \{0, \ldots, U\}$.

4 Exact Local Distributions on Composite Predicates

In this section and in Section 5 we will show how to satisfy the constraint $\text{OBJ}(x) = \beta$ exactly, with pseudovariance zero. Our strategy will be to group predicates together into "composite" predicates, and then prove that there is a local $(t-1)$-wise uniform distribution

which is moreover supported on variable assignments for which an exact β-fraction of the predicates within the composite predicate are satisfied. We'll then apply Theorem 6 to the composite predicates.

We begin with an easier proof for the case when there is a pairwise-uniform distribution over satisfying assignments to our predicate in Section 4.1, and later in Section 4.2 we handle t-wise uniform distributions for larger t. While the pairwise-uniform theorem is less general, the proof is simpler and it already suffices for all of our bisection applications.

4.1 Pairwise-uniform distributions over β-satisfying assignments

Recall our setting: we have a Boolean k-ary predicate $P : \{\pm 1\}^k \to \{0,1\}$ and a pairwise-uniform distribution ν over assignments $x \in \{\pm 1\}^k$ such that $\mathbb{E}[P(x)] = \beta$. The following theorem states that we can extend ν into a distribution θ over assignments to groups of r predicates at a time, $\{\pm 1\}^{r\times k}$ so that exactly $(\beta - \varepsilon)\cdot r$ of the r predicates are satisfied by any assignment $y \sim \theta$.

▶ **Theorem 8.** *Let $P : \{\pm 1\}^k \to \{0,1\}$ be a k-ary Boolean predicate, and let ν be a pairwise-uniform distribution over assigments $\{\pm 1\}^k$ with the property that $\nu(x)$ is rational for each $x \in \{\pm\}^k$; that is, there exist a multiset $S \subseteq \{\pm 1\}^k$ such that for each $x \in \{\pm 1\}^k$, $\nu(x) = \mathbb{P}_{s\sim S}(s = x)$. Suppose also that $\mathbb{E}_{x\sim\nu}[P(x)] = \beta$, and that this is more than the expectation under the uniform distribution, so $\beta > \mathbb{E}_{x\sim\{\pm 1\}^k}[P(x)]$.*

Then for any constant $\varepsilon > 0$, there exists an integer $r = O_{\varepsilon,k}(|S|^3)$ and a rational $\tilde{\varepsilon} \leqslant \varepsilon$ so that there is a pairwise-uniform distribution θ over assigments to groups of r predicates, $\{\pm 1\}^{r\times k}$ such that exactly $(\beta - \tilde{\varepsilon})r$ of the predicates are satisfied by any assigment $y \sim \theta$.

Throughout we'll refer to the assignments in the support of θ as *matrices*, with each row of the $r \times k$ matrix corresponding to the assignment for a single copy of the predicate.

Since we have assumed that the probability of seeing any string in the support of ν is rational, without loss of generality we can assume that ν is uniform over some multiset $S \subseteq \{\pm 1\}^k$. As a first guess at θ, one might try to take $r = c\cdot|S|$ for some positive integer c, make c copies of the multiset S, and use a random permutation of the elements of this multiset to fill the rows of an $r \times k$ matrix. But this distribution is not *quite* pairwise uniform. The issue is that because each individual bit is uniformly distributed, every column of the matrix will always be perfectly balanced between ± 1. Therefore the expected product of two distinct bits in a given column is

$$\frac{1}{\binom{c|S|}{2}}\left(\binom{\frac{1}{2}c|S|}{2}(-1)^2 + \binom{\frac{1}{2}c|S|}{2}(+1)^2 + \left(\frac{1}{2}c|S|\right)^2(+1)(-1)\right) = -\frac{1}{c|S|-1} \neq 0.$$

So, the bits within a particular column have a slight negative correlation.

We'll compensate for this shortcoming as follows: we will randomly choose an element s in the support of ν to repeat multiple times. This may in turn alter the number of predicates satisfied out of the r copies of P, whereas our express goal was to satisfy the exact same number of predicates under every assignment. To adjust for this, we'll mix in some rows from the uniform distribution over $\{\pm 1\}^k$, where the number of rows we mix in will depend on whether $P(s) = 1$ or 0.

Proof. Let S be a multiset of strings in $\{\pm 1\}^k$ such that $\mathbb{P}_{s\sim S}(s = a) = \nu(a)$. We will also require a multiset $T \subseteq \{\pm 1\}^k$ which is a well-chosen mixture of ν and the uniform distribution; the following claim shows that we can choose such a set. Here, we take some care in choosing this combination; the exact choice of parameters will not matter until later.

▶ **Claim 9.** *For any constant $\varepsilon > 0$, there is a constant $L = O(1/\varepsilon)$ and a constant $R \geqslant 1$ so that there are multisets $S', T \subseteq \{\pm 1\}^k$ with the following properties: S' is $RL2^k$ copies of S, T has size $|T| = |S'| = LR2^k|S|$, and $\mathbb{P}_{x\sim T}[P(x) = 1] = \beta - \varepsilon'$, where $\varepsilon > \varepsilon' \stackrel{\text{def}}{=} \frac{1}{L}(\beta - \mathbb{E}_{x\sim\{\pm1\}^k}[P(x)])$.*

Proof. Let $s = |S|$, and let $U = \{\pm 1\}^k$. Suppose that $\eta 2^k$ of U's assignments satisfy P. Define T to be the multiset given by $2^k(L-1)R$ copies of S and sR copies of U. We have

$$\mathop{\mathbb{P}}_{x\sim T}[P(x) = 1] = \frac{1}{2^k LRs}\left(\beta s \cdot 2^k(L-1)R + \eta 2^k \cdot sR\right) = \beta - \frac{\beta - \eta}{L}.$$

By choosing L large as a function of ε, we can make this probability as small as we want. ◀

For convenience, let $\ell = 2^k LR|S|$. Set the number of rows $r = d\ell$ for an integer $d = O_\varepsilon(\ell^2)$ to be specified later. We also let a, b_1, b_0, c_1, c_0 be integers which will specify the number of rows from S', T, and the repeated assignment set; we'll set the integers later, but we will require the property that

$$d = a + b_1 + c_1 = a + b_0 + c_0. \tag{4.1}$$

We generate a sample from θ in the following fashion:

1. Sample $s \sim S$, and fill the first $a\ell$ rows with copies of s. Call these the A rows.
2. Set $i = P(s)$, that is $i = 1$ if s satisfies P and $i = 0$ otherwise.
3. Fill the next $b_i\ell$ rows with b_i copies of each string in S'. Call these the B rows.
4. Fill the last $c_i\ell$ rows with c_i copies of each string in T. Call these the C rows.
5. Randomly permute the rows of the matrix.

If $\delta\ell$ assignments in T are satisfying and $\beta\ell$ assignments in S' are satisfying, to ensure that the number of satisfying rows are always the same we enforce the constraint

$$a\ell + b_1\beta\ell + c_1\delta\ell = b_0\beta\ell + c_0\delta\ell, \tag{4.2}$$

Now we handle uniformity. We will prove that all of the degree-1 and degree-2 moments of the bits in the matrix are uniform under θ. First, we argue that the degree-1 moments are zero, and that the correlation of any two bits in the same row is zero.

▶ **Claim 10.** *The bits in a single row of M are pairwise uniform.*

Proof. We can condition on the row type, A, B, or C. For each type of row, there is a multiset U such that $\mathbb{E}[M_{ij}] = \mathbb{E}_{x\sim U}[x_j] = 0$ by the pairwise uniformity of the uniform distribution over U. The same argument proves the statement for the product of two bits in a fixed row. ◀

Thus, it suffices to prove that the bits in each column are pairwise-uniform; this is because the pairwise uniformity of rows implies that we can fix the values of any entire column, and the remaining individual bits in other columns will remain uniformly distributed. So we turn to proving that the columns are pairwise-uniform.

▶ **Claim 11.** *If we choose a, b_0, b_1, c_0, c_1 so that*

$$a(\ell - 1) - \beta(b_1 + c_1) - (1 - \beta)(b_0 + c_0) = 0,$$

then the bits in a single row of M are pairwise uniform.

Proof. We'll prove this by computing the expected product of two distinct bits, x and y, which both come from the ith column of M. We will compute the conditional expectation of xy given the group of rows that x, y were sampled from.

We first notice that conditioned on x coming from one type of row and y coming from another, x and y are independent of each other, and by the uniformity of individual bits in each group, $\mathbb{E}[xy \mid x, y \sim \text{ different groups}] = 0$.

Restricting our attention now to pairs of bits from within the same group, we compute the conditional expectations. If both bits come from the A rows, they are perfectly correlated. On the other hand, if both bits come from the B or C rows their correlation is as we computed above, $-\frac{1}{b\ell-1}$ or $-\frac{1}{c\ell-1}$ respectively. Therefore we can simplify

$$\begin{aligned}
&\mathbb{E}[xy] \\
&= \beta \cdot \mathbb{E}[xy \mid P(a) = 1] + (1-\beta) \cdot \mathbb{E}[xy \mid P(a) = 0] \\
&= +\beta\left(\frac{\binom{a\ell}{2}}{\binom{d\ell}{2}} + \mathbb{E}[xy|x, y \sim B, P(a) = 1] \cdot \frac{\binom{b_1\ell}{2}}{\binom{d\ell}{2}} + \mathbb{E}[xy|x, y \sim C, P(a) = 1] \cdot \frac{\binom{c_1\ell}{2}}{\binom{d\ell}{2}}\right) \\
&\quad + (1-\beta)\left(\frac{\binom{a\ell}{2}}{\binom{d\ell}{2}} + \mathbb{E}[xy|x, y \sim B, P(a) = 0] \cdot \frac{\binom{b_0\ell}{2}}{\binom{d\ell}{2}} + \mathbb{E}[xy|x, y \sim C, P(a) = 0] \cdot \frac{\binom{c_0\ell}{2}}{\binom{d\ell}{2}}\right) \\
&= \frac{\ell}{2\binom{d\ell}{2}}\Big(a(a\ell - 1) - \beta(b_1 + c_1) - (1-\beta)(b_0 + c_0)\Big), \qquad (4.3)
\end{aligned}$$

where (4.3) gives us the condition of the claim. ◀

Finally, we are done given that we can find positive integers satisfying the constraints

$$\begin{aligned}
d - a = b_1 + c_1 &= b_0 + c_0 && \text{(from (4.1))} \\
0 &= a + \beta(b_1 - b_0) + \delta(c_1 - c_0) && \text{(from (4.2))} \\
0 &= a(\ell - 1) - \beta(b_1 + c_1) - (1-\beta)(b_0 + c_0) && \text{(from (4.3))}
\end{aligned}$$

The following can be verified to satisfy the constraints above:

$$\begin{aligned}
a &:= 2(\beta - \delta)\ell; & b_1, c_0 &:= ((\beta - \delta)(\ell - 1) - 1)\ell; \\
b_0, c_1 &:= ((\beta - \delta)(\ell - 1) + 1)\ell; & d &:= 2(\beta - \delta)\ell^2.
\end{aligned}$$

By our choice of $\ell = 2^k|S|LR$ and since $\beta - \delta = \frac{\beta - \mathbb{E}[P(x)]}{L}$, we can choose R large enough so that $(\beta - \delta)(\ell - 1) > 1$, and because $\beta\ell$ and $\delta\ell$ are integers, these are all also positive integers, as required.

We compute the number of satisfied assigments as a function of the total, which is

$$\beta\frac{b_0}{d} + \delta\frac{c_0}{d} = \beta - \frac{\varepsilon'}{2} + \frac{1}{\ell}\left(\frac{1+\varepsilon'}{2} - \beta\right).$$

The conclusion thus holds, with $\tilde{\varepsilon} \stackrel{\text{def}}{=} \frac{\varepsilon'}{2} - \frac{1}{\ell}\left(\frac{1+\varepsilon'}{2} - \beta\right)$. ◀

4.2 $t-1$-wise uniform distributions over β-satisfying assignments

We now prove the generalization of the statement in the previous section to t-wise uniform distributions over β-satisfying assignments.

▶ **Theorem 12.** *Let $P : \{\pm 1\}^k \to \{0,1\}$ be a k-ary Boolean predicate, let $t \geqslant 2$ be an integer, and let ν be a $(t-1)$-wise uniform distribution over assigments $\{\pm 1\}^k$ so that there exist a multiset $S \subseteq \{\pm 1\}^k$ such that for each $x \in \{\pm 1\}^k$, $\nu(x) = \mathbb{P}_{s\sim S}(s = x)$. Suppose also that $\mathbb{E}_{x\sim\nu}[P(x)] = \beta > \mathbb{E}_{x\sim\{\pm 1\}^k}[P(x)]$.*

Then for any constant $\varepsilon > 0$, there exists an integer $r = O_{\varepsilon,k}(|S|^4)$ and a rational $\tilde{\varepsilon} \leqslant \varepsilon$ so that there is a $(t-1)$-wise uniform distribution θ over assigments to groups of r predicates, $\{\pm 1\}^{r\times k}$ such that exactly $(\beta - \tilde{\varepsilon})r$ of the predicates are satisfied by any assigment $y \sim \theta$.

The proof will use a similar, though slightly more involved, construction of θ than in the pairwise case. It may be helpful to note that the choice of $t = 3$ in Theorem 12 will not give the same construction as in Theorem 8 (although of course one could set $t = 3$ and obtain a result for pairwise-uniform ν). In particular, it will not be enough to choose one string to repeat many times in order to improve the column-wise correlations. Instead, we will repair the correlations in one column at a time, by sampling some subset of the bits in each column from a bespoke distribution, designed to make the columns $(t-1)$-wise independent. We will have to be careful with the choice of distribution, so that we can still control the number of satisfying assigmnets in M as a whole.

Proof. As in the proof of Theorem 8, we will require a well-chosen convex combination of ν and the uniform distribution to ensure that the number of satisfying assingments is always the same. We appeal to Claim 9, taking S' and T to be as described there, with $L = O(1/eps)$ (to be set more precisely later) and $R = 1$. For convenience let's let $\ell \stackrel{\text{def}}{=} 2^k L|S|$ and let's let $\delta = \beta - \varepsilon'$.

We also call $S'_{i=1}$ and $S'_{i=-1}$ to be the sub-multisets of S' which have the ith bit set to 1 and -1 respectively. We notice that a uniform sample from $S'_{i=1}$ is equivalent to a uniform sample from ν conditioned on the ith bit being 1. Also by the $(t-1)$-wise uniformity we have $|S'_{i=1}| = \ell/2$. Notice that since S' is made up of $2^k L$ copies of S, the discrepancy in the number of satisfying assignments between $S_{i=1}$ and $S_{i=-1}$ is always an integer multiple of $2^k L$.

Set r, the number of rows, be an integer which we will specify later. We also choose the integer a to represent the size of the correction rows, and b_n, c_n, the number of copies of S' and T for each $n \in [ak\ell/(2^k L)]$ (where $n2^k L$ is the number of satisfying assignments in the correction rows). To make sure the number of rows always adds up to r, we'll need the constraint,

$$r = ak\ell + b_n\ell + c_n\ell \quad \forall z \in \{\pm 1\}^k \tag{4.4}$$

In order to make sure that the columns are $(t-1)$-wise independent, we require a "column repair" distribution κ over $\{\pm 1\}^{a\ell}$. We will specify this distribution later; for now, we need only that κ is symmetric and that the number of 1s in any $z \sim \kappa$ is a multiple of $\ell/2$. The latter property is because, when we choose some part of column i according to κ, we will want to fix the rows with copies of $S_{i=1}$ and $S_{i=-1}$.

We generate a sample $M \in \{\pm 1\}^{r\times k}$ from θ in the following fashion:

1. For each $i \in [k]$, independently sample a string $z_i \sim \kappa$. Add $a\ell$ rows to M, where in the ith column we put the bits of z_i, and we set the remaining row bits so that if z_i has $(a - a')\ell/2$ entries of value 1 and $a'\ell/2$ entries of value -1, then we end up with a' copies of $S'_{i=-1}$ and $a - a'$ copies of $S'_{i=1}$. Call these rows A_i.
2. Compute the integer n such that $n \cdot 2^k L$ is the number of rows in $\cup_{i=1}^k A_i$ containing satisfying assignments to P, given our choices of z_i $\forall i \in [k]$.

3. Add $b_n\ell$ rows to M which contain b_n copies of each string from S. Call these rows B.
4. Add $c_n\ell$ rows to M which contain c_n copies of each string from T. Call these rows C.
5. Randomly permute the rows of M.

So that the number of satisfying assignments Λ is always the same, we require that

$$\Lambda = n2^k L + b_n \cdot \beta\ell + c_n \cdot \delta\ell \quad \forall n \in [ak\ell/(2^k L)] \tag{4.5}$$

Now, we will derive the conditions under which $(t-1)$-wise independence holds. As above, we first consider bits that are all contained in a fixed row.

▶ **Claim 13.** *The bits in a single row of M are $(t-1)$-wise uniform.*

Proof. We can condition on the row type, $A_1, \ldots, A_k, B$, or C. Sampling a uniform row from B or C is equivalent to sampling from ν or γ, which are $(t-1)$-wise uniform. Since κ is symmetric, sampling a row from A_i is equivalent to sampling from ν as well, and we are done. ◀

From the claim above, if we condition on the value in $d < t-2$ columns, the remaining $t-2-d$ columns will remain identically distributed; this is because the rows are $(t-1)$-wise uniform, so after conditioning the distribution in each row will remain $(t-1-d)$-wise uniform. Thus, proving that each column is $(t-1)$-wise uniform suffices to prove $(t-1)$-wise uniformity on the whole.

The following lemma states that we may in fact choose κ so that this condition holds exactly.

▶ **Lemma 14.** *Let $y \in \{\pm\}^{r-a\ell}$ be a perfectly balanced string. If $a\ell > h_1 \cdot \sqrt{tr}$ for a fixed constant h_1 and $\sqrt{r} \geqslant (t-1)\ell 2^{h_2 t}$ for a fixed constant h_2, then there is a distribution κ over $\{\pm 1\}^{a\ell}$, supported on strings which have a number of 1s which is a multiple of $\ell/2$, such that if x is sampled by choosing $z \sim \kappa$, concatenating z with y and applying a random permutation, then for any $S \subset [r]$ with $|S| \leqslant t-1$, $\mathbb{E}[x^S] = 0$.*

Since each column is distributed as the string x described in the lemma statement, the lemma suffices to give us $(t-1)$-wise uniformity of the columns. We'll prove the lemma below, but first we conclude the proof of the theorem statement.

We now choose the parameters to satisfy our constraints. We have the requirements:

$$\begin{aligned}
\Lambda &= n2^k L + b_n\beta\ell + c_n\delta\ell \quad \forall n \in [ak\ell/(2^k L)] && \text{(from (4.5))}\\
r - ak\ell &= b_n\ell + c_n\ell \quad \forall n \in [ak\ell/(2^k L)] && \text{(from (4.4))}\\
a\ell &\geqslant h_1\sqrt{tr} && \text{(from Lemma 14)}\\
\sqrt{r} &\geqslant (t-1)\ell 2^{h_2 t} && \text{(from Lemma 14)}
\end{aligned}$$

where h_1 and h_2 are universal constants. The below choice of integer parameters satisfies these requirements, as well as the requirement of always being non-negative:

$$u = \left(\beta - \mathop{\mathbb{E}}_{x\sim\{\pm1\}^k}[P(x)]\right)2^k|S|; \qquad L = u \cdot \max(1, \lceil h_1 + h_2\rceil) \cdot \left\lceil\frac{1}{\varepsilon}\right\rceil \cdot k; \quad \ell = 2^k L|S|;$$

$$a = \left\lceil h_1 2^{h_2 t/2} tk\right\rceil; \qquad r = \left\lfloor\frac{a^2}{h_1^2 t}\right\rfloor \cdot \ell^2;$$

$$b_0 = \frac{1}{2}\left(\frac{1}{\ell}r - ak\right); \qquad b_{n+1} = b_n - \frac{2^k L}{u};$$

$$c_0 = \frac{1}{2}\left(\frac{1}{\ell}r - ak\right); \qquad c_{n+1} = c_n + \frac{2^k L}{u}.$$

Finally, we have that the fraction of satisfying rows in M is always exactly

$$\frac{\Lambda}{r} = \frac{\frac{1}{2}(r - ak\ell)\beta + \frac{1}{2}(r - ak\ell)\delta}{r} = \beta - \frac{\varepsilon'}{2} - O\left(\frac{ak\ell}{r}\right).$$

The latter term is $O(\frac{1}{\ell})$, and we have chosen L large enough so that it is smaller than $\varepsilon'/2$. ◀

Proof of Lemma 14. For convenience, call $m \stackrel{\text{def}}{=} r - a\ell$. Recall that we take x to be sampled by taking a balanced string $y \in \{\pm 1\}^m$, sampling $z \sim \kappa$, appending z to y and then applying a uniform permutation to the coordinates.

We will solve for κ with a linear program (LP) over the probability p_z of each string $z \in \{\pm 1\}^{a\ell}$. We have the program

$$\forall S \in [m], |S| \in \{1, \dots, t-1\}: \qquad \sum_{\substack{z \in \{\pm 1\}^{a\ell} \\ (\ell/2)|\sum_j z_j}} \mathbb{E}\left[x^S \;\middle|\; \sum_i x_i = \sum_{j \in [a\ell]} z_j\right] \cdot p_z = 0$$

$$\forall z \in \{\pm 1\}^{a\ell} \text{ s.t. } \frac{\ell}{2}\left|\sum_j z_j\right|: \qquad p_z \geqslant 0$$

Since we can take any solution to this LP and scale the p_z so that they sum to 1, the feasibility of this program implies our conclusion. So suppose by way of contradiction that this LP is infeasible. Then Farkas' lemma implies that there exists a $q \in \mathbb{R}^{t-1}$ such that

$$\forall z \in \{\pm 1\}^{a\ell} \text{ s.t. } \frac{\ell}{2}\left|\sum_j z_j, \qquad \sum_{\substack{S \subseteq [m] \\ |S| \in \{1,\dots,t-1\}}} \mathbb{E}\left[x^S \;\middle|\; \sum_i x_i = \sum_{j \in [a]} z_j\right] \cdot y_s > 0.\right.$$

Without loss of generality, we scale q so that $\sum_S q_S^2 = 1$. Moreover by the symmetry of the expectation over subsets S, we can assume that $q_S = q_T$ whenever $|S| = |T|$. This implies that the degree-t mean-zero polynomial

$$q(x) = \sum_{\substack{S \subseteq [m] \\ 1 \leqslant |S| \leqslant t-1}} q_S \cdot \chi_S(x)$$

has positive expectation over every layer of the hypercube with $|\sum_i x_i| = d$ such that $d \leqslant a\ell$ and $\frac{\ell}{2}|d$. Furthermore, q is a symmetric polynomial, which implies that it takes the same value on all inputs of a fixed Hamming weight; this implies that it takes positive values on every inputs x with $|\sum x_i| \in [2a] \cdot \frac{\ell}{2}$.

The following fact will give us the contradiction we desire:

▶ **Fact 15** (Tails of low-degree polynomials [6], see Theorem 4.1 in [1])**.** *Let $f : \{0,1\}^m \to \mathbb{R}$ be a degree-t polynomial with mean zero and variance 1. Then, there exist universal constants $c_1, c_2 > 0$ such that $\mathbb{P}[p \leqslant -2^{-c_1 t}] \geqslant 2^{-c_2 t}$.*

We will show that since q takes positive value on every hypercube slice of discrepancy $\frac{\ell}{2} \cdot [2a]$, this implies that it takes positive values on most hypercube slices with discrepancy at most $a\ell$. Because we have chosen $a\ell$ so that this comprises the bulk of the hypercube, this in turn will contradict Fact 15.

In fact, because q is symmetric and of degree $t-1$ over the hypercube, we can equivalently write q as a degree-$(t-1)$ polynomial in the single variable $x' := \sum_i x_i$, $q(x) = g(\sum_i x_i) = \sum_{s\in\{0,\ldots,t-1\}} g_s \cdot (\sum_i x_i)^s$. Viewing g as a univariate polynomial over the reals, g has at most $t-1$ roots. Therefore we conclude that q can only be non-positive on at most $t-1$ intervals of layers of discrepancy $(i\ell/2, (i+1)\ell/2)$. So for at least $2a\ell - (t-1)\frac{\ell}{2}$ slices of the hypercube around 0, q takes positve value.

Each slice of the hypercube has probability mass at most $\frac{1}{\sqrt{\frac{\pi}{2}(m+a\ell)}}$. By our choice of a, ℓ, m and by a Chernoff bound,

$$\mathbb{P}(q(x) > 0) \geqslant \mathbb{P}\left(\left|\sum_i x_i\right| \leqslant a\ell\right) - \frac{(t-1)\ell}{2\sqrt{m+a\ell}} > 1 - 2^{-c_2 t},$$

which contradicts Fact 15. ◀

5 SOS lower bounds for CSPs with exact objective constraints

In this section we put things together and show how to extend Theorem 12 to prove Theorem 2. As discussed briefly in Section 1.1, our random instance of the $CSP(P^{\pm})$ will be sampled in a somewhat non-standard way, which we will refer to as "batch-sampling." This is because, in order to apply Theorem 12 to a random instance Φ of a Boolean CSP, we need to partition Φ's constraints into groups of r non-intersecting constraints for some integer r, while also maintaining the expansion properties required by Theorem 5.

We first prove Theorem 2 as stated, for random CSPs sampled from a slightly different distribution. Then in Section 5.2 we show that for a "standard" random CSP with $m = o(n^{3/2})$ constraints, we can still get a theorem along the lines of Theorem 2.

5.1 Exact objective constraints for batch-sampled random CSPs

Suppose that P is a k-ary predicate, and let r be some positive integer which divides m. We'll "batch-sample" an n-variate random $CSP(P^{\pm})$ with m clauses as follows:

1. Choose independently m/r subsets each of $r \cdot k$ distinct variables uniformly at random from $[n]$, $S_1, \ldots, S_m$
2. For each $j \in [m/r]$, $S_j = \{xi_1, \ldots, x_{i_{rk}}\}$:
   - Choose a random signing of P, $z_j \in \{\pm 1\}^k$
   - To each block of k variables in S_j, $(x_{i_{(\ell-1)\cdot k+1}}, \ldots, x_{i_{\ell k}})$ for $\ell \in [r]$, add the predicate P with signing z_j.

▶ **Theorem 16** (Restatement of Theorem 2). *Let P be a k-ary predicate, and let ν be a $(t-1)$-wise uniform distribution over $\{\pm 1\}^k$ under which $\mathbb{E}_\nu[P] = \beta$. Then for each constant $\varepsilon > 0$ there is a choice of positive integer r such that for a random instance of $CSP(P^{\pm})$ on n variables with $m = \Delta n$ constraints for sufficiently large Δ and $r|m$, sampled as detailed above, there is a degree-$\Omega(\frac{n}{\Delta^{2/(t-2)}\log\Delta})$ SOS pseudodistribution whch satisfies with pseudovariance zero the constraint $\mathrm{OBJ}(x) = \beta - \varepsilon_r$, where $\varepsilon_r < \varepsilon$. This is also true when cardinality constraints are imposed as in Theorem 1.*

Proof. This distribution over instances is equivalent to the standard notion of sampling a random CSP with m/r constraints in the composite predicates from Theorem 12: a scope is chosen independently and uniformly at random for each predicate. Therefore, if we replace each collection of constraints corresponding to S_j with the composite predicate from Theorem 12, and modify ν in accordance with the signing z_j, we have a $(t-1)$-wise uniform

distribution over solutions to the composite predicates supported entirely on assignments which satisfy exactly $\beta - \varepsilon_r$ of the clauses. Combining this with the expansion theorem (Theorem 6), we have our conclusion. ◀

5.2 Exact objective constraints for sparse random CSPs.

Though the batch-sampled distribution over CSPs for which Theorem 2 holds is slightly non-standard, here we show that with minimal effort, we can prove a similar theorem for sparse random instances sampled in the usual manner, when $m = o(n^{3/2})$.

▶ **Theorem 17.** *Let P be a k-ary predicate, let ν be a $(t-1)$-wise uniform distribution over $\{\pm 1\}^k$ such that $\mathbb{E}_\nu[P] = \beta$, and suppose we sample a random instance Φ of a $CSP(P^{\pm})$ in the usual way, by selecting m random signed P-constraints on n variables. Then if $m = \Delta n = o(n^{3/2})$ for sufficiently large Δ, with high probability over the choice of Φ, for each $\varepsilon > 0$ there exists some constant $\varepsilon_\Phi \leqslant \varepsilon$ such that there is a degree-$\Omega_\varepsilon(\frac{n}{\Delta^{2/(t-2)}\log\Delta})$ pseudodistribution which satisfies with pseudovariance zero the constraint $\mathrm{OBJ}(x) = \beta - \varepsilon_\Phi$ and a Hamming weight constraint $\sum_{i\in[n]} x_i = B$ for $|B| = O(\sqrt{n})$.*

Proof. Fix ε, and let r be the corresponding constant required to achieve objective $\mathrm{OBJ}(x) = \beta - \varepsilon^*$ under Theorem 12 for $\varepsilon^* < \varepsilon/2$.

We first couple the standard sampling procedure for a random P-CSP to a different sampling procedure. For simplicity we at first ignore the possible signings of P, and assume we work only with un-negated variables; later we explain how to modify the proof to accomodate negative literals.

We sample a random CSP by independently and uniformly choosing m random scopes $S_1, \ldots, S_m$. For each $\ell \in \{0, 1, \ldots \lfloor m/r \rfloor - 1\}$, the probability that $S_{\ell r+1}, \ldots, S_{(\ell+1)r}$ have non-intersecting scopes is at least

$$\mathbb{P}[\cap_{j\in[r]} S_j = \emptyset] = \prod_{i=2}^{r} \mathbb{P}[S_i \cap (\cap_{j<i} S_j) = \emptyset \mid \cap_{j<i} S_j = \emptyset] = \prod_{i=2}^{r} \left(1 - i\frac{k}{n}\right) \geqslant 1 - O\left(\frac{r\cdot k}{n}\right).$$

So with high probability for all but $O(\frac{m}{n})$ of the intervals of constraints $j \in [\ell \cdot r + 1, (\ell+1)r]$, the constraints will be non-intersecting. Call this the "non-intersecting configuration".

Define a "collision configuration" to be a choice of scopes for which the above condition does not hold; that is, a specific way in which S_j intersects with one or more $S_{j'}$ when $j, j' \in [\ell \cdot r + 1, (\ell+1)r]$. Each of the $\approx \binom{2kr}{r}$ collision configurations has a fixed probability of ocurring (which may be easily calculated), and the total sum of these probabilities is at most $O(r \cdot k/n)$.

Let $\mathcal{D}_r^{(m)}$ be the multinomial distribution which describes the number of occurrences of each configuration for a random CSP with m constraints ($\lfloor m/r \rfloor$ configurations). We couple the standard sampling procedure with the following alternative sampling procedure: we first sample $c \sim \mathcal{D}_r^{(m)}$ to determine how many configurations of each type there are. Then, for each collision configuration specified by c, sample the scope (of size $< k \cdot r$) for each of the collision configurations independently and uniformly at random. Also, sample and additional ($m \mod r$) scopes of k variables for the "leftover copies" of P. Finally, sample the scopes of the non-intersecting configurations specified by c independently uniformly at random. The coupling of the two processes is immediate.

Let C be the number of collision configurations plus ($m \mod r$), the number of leftover copies. As shown above, with high probability over $c \sim \mathcal{S}_r^{(m)}$, the number of collision configurations is at most $O(m/n) = o(n^{1/2})$, so $C = o(n^{1/2}) = o(m)$.

From our alternate sampling procedure, we conclude that with high probability we can meet the conditions of Theorem 6 by fixing an arbitrary variable assignment to any collision configuration. That is, we could alternately first sample the collision configurations and leftover copies, and then set all of the variables present inside be set to (say) False. We take note of how many constraints in the P-CSP are and are not satisfied by this unary assignment, and we correspondingly amend ε^* to ε_Φ. Since with high probability at most $o(m)$ constraints are fixed, we retain the property that $\varepsilon_\Phi \leqslant \varepsilon^* + o(1) \leqslant \varepsilon$.

Now, if we wish to satisfy a Hamming weight constraint, we add arbitrary matching and unary constraints to get the desired Hamming weight; at most $O(C)$ unary constraints are needed to compensate for the $\leqslant Ckr$ variables we set to False.

Finally, we sample the remaining non-intersecting configurations independently; by Theorem 6 when $C = o(n^{1/2})$, the expansion properties we require are met for the composite predicates on the non-intersecting configurations. Since this occurs with high probability, we are done.

To extend the argument to allow predicates on negative literals, we couple with a slightly more elaborate sampling procedure: for each signing pattern $z \in \{\pm 1\}^k$, we draw a separate set of m_z predicates (where m_z may either be deterministic or sampled from a multinomial distribution). For each signing separately we repeat the argument above, and then in the final sampling procedure we sample counts c_z for each signing z, add the leftover copies and collision configurations separately for each signing, add the unary constraints, and then sample the remaining non-intersecting copies. ◀

6 Conclusions

In this work we have shown that, in the context of random Boolean CSPs, the following strategies do *not* give SOS any additional refutation power: (i) trying out all possible Hamming weights for the solution; (ii) trying out all possible (exact) values for the objective function. We also gave the first known SOS lower bounds for the Min- and Max-Bisection problems.

We end by mentioning some open directions. There are two technical challenges arising in our work that look approachable. The first is to extend our results from Section 4 on "exactifying" distributions to the case of larger alphabets. The second is to prove (or disprove) that the "random*" and "purely random" distributions discussed in Remark 1.1 are $o(1)$-close (depending on $m(n)$).

Finally, we suggest investigating further strategies for handling hard constraints in the context of SOS lower bounds. Sometimes this is not too difficult, especially when reducing from linear predicates such as 3XOR, where there are perfectly satisfying SOS solutions. Other times, it's of moderate difficulty, perhaps as in this paper's main Theorem 1 and Theorem 2. In still other cases it appears to be very challenging.

One difficult case seems to be in the context of SOS lower bounds for refuting the existence of large cliques in random graphs. In [4] it is shown that in a $G(n, 1/2)$ random graph, with high probability degree-$\Omega(\log n)$ SOS thinks there is a clique of size $\omega := n^{1/2-\varepsilon}$. (Here $\varepsilon > 0$ can be any constant.) However, it's merely the case that $\widetilde{\mathbb{E}}[\text{clique size}] \geqslant \omega$, and it is far from clear how to upgrade the SOS solution so as to actually satisfy the constraint "clique size $= \omega$" with pseudovariance zero. Besides being an improvement for its own sake, it would be very desirable to have such an SOS solution for the purposes of further reductions; for example, it would greatly simplify the recent proofs of SOS lower bounds for approximate Nash welfare in [15]. It also seems it might be useful for tackling SOS lower bounds for coloring and stochastic block models.

Finally, we leave as open one more "hard constraint" challenge that arises even in the simple context of random 3XOR or 3SAT. Suppose one tried to refute random m-constraint 3XOR instances by trying to refute the following statement for all quadruples $(k_{001}, k_{010}, k_{100}, k_{111})$ that sum to m:

"exactly k_a constraints are satisfied with assignment a", for each $a \in \{001, 010, 100, 111\}$.

As far as we know, constant-degree SOS may succeed with this strategy when $m = O(n)$. It is natural to believe that there is (whp) an $\Omega(n)$-degree SOS pseudodistribution that satisfies all of the above constraints with pseudovariance zero when $k_{001} = k_{010} = k_{100} = k_{111} = m/4$, but we do not know how to construct one.

References

1. Per Austrin and Johan Håstad. Randomly supported independence and resistance. *SIAM Journal on Computing*, 40(1):1–27, 2011.
2. Boaz Barak, Fernando Brandão, Aram Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, Sum-of-Squares Proofs, and their Applications. In *Proc. of the 44th Annual ACM Symposium on Theory of Computing*, pages 307–326, 2012.
3. Boaz Barak, Siu On Chan, and Pravesh Kothari. Sum of Squares Lower Bounds from Pairwise Independence. In *Proc. of the 47th Annual ACM Symposium on Theory of Computing*, pages 97–106, 2015.
4. Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. In *Proc. of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.
5. Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares, 2016. URL: `http://sumofsquares.org/public/index.html`.
6. Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel Journal of Mathematics*, 160(1):389–412, 2007.
7. Uriel Feige. Relations between average case complexity and approximation complexity. In *Proc. of the 34th Annual ACM Symposium on Theory of Computing*, pages 543–543, 2002.
8. Dima Grigoriev. Complexity of Positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001.
9. Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613–622, 2001.
10. Dima Grigoriev, Edward Hirsch, and Dmitrii Pasechnik. Complexity of semialgebraic proofs. *Moscow Mathematical Journal*, 2(4):647–679, 2002.
11. Venkatesan Guruswami, Ali Kemal Sinop, and Yuan Zhou. Constant factor Lasserre integrality gaps for graph partitioning problems. *SIAM Journal on Optimization*, 24(4):1698–1717, 2014.
12. Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.
13. Subhash Khot. Ruling out PTAS for Graph Min-Bisection, Dense k-Subgraph, and Bipartite Clique. *SIAM Journal on Computing*, 36(4):1025–1071, 2006.
14. Pravesh Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proc. of the 49th Annual ACM Symposium on Theory of Computing*, pages 132–145, 2017.
15. Pravesh K. Kothari and Ruta Mehta. Sum-of-squares meets Nash: lower bounds for finding any equilibrium. In *Proc. of the 50th Annual ACM Symposium on Theory of Computing*, pages 1241–1248, 2018.

16 Ryan O'Donnell. SOS is not obviously automatizable, even approximately. In *Proc. of the 8th Annual Innovations in Theoretical Computer Science conference*, 2017.

17 Ryan O'Donnell and John Wright. A new point of NP-hardness for Unique-Games. In *Proc. of the 44th Annual ACM Symposium on Theory of Computing*, pages 289–306, 2012.

18 Prasad Raghavendra and Benjamin Weitz. On the Bit Complexity of Sum-of-Squares Proofs. Technical report, arXiv, 2017. `arXiv:1702.05139`.

19 Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k-CSPs. In *Proc. of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602, 2008.

20 Luca Trevisan, Gregory Sorkin, Madhu Sudan, and David Williamson. Gadgets, Approximation, and Linear Programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000.

21 Madhur Tulsiani. CSP gaps and reductions in the Lasserre hierarchy. In *Proc. of the 41st Annual ACM Symposium on Theory of Computing*, pages 303–312, 2009.