

Sum of squares lower bounds for refuting any CSP

Pravesh K. Kothari* Ryuhei Mori† Ryan O’Donnell‡ David Witmer‡

January 16, 2017

Abstract

Let $P : \{0, 1\}^k \rightarrow \{0, 1\}$ be a nontrivial k -ary predicate. Consider a random instance of the constraint satisfaction problem $\text{CSP}(P)$ on n variables with Δn constraints, each being P applied to k randomly chosen literals. Provided the constraint density satisfies $\Delta \gg 1$, such an instance is unsatisfiable with high probability. The *refutation* problem is to efficiently find a proof of unsatisfiability.

We show that whenever the predicate P supports a *t-wise uniform* probability distribution on its satisfying assignments, the sum of squares (SOS) algorithm of degree $d = \Theta(\frac{n}{\Delta^{2/(t-1)} \log \Delta})$ (which runs in time $n^{O(d)}$) *cannot* refute a random instance of $\text{CSP}(P)$. In particular, the polynomial-time SOS algorithm requires $\tilde{\Omega}(n^{(t+1)/2})$ constraints to refute random instances of $\text{CSP}(P)$ when P supports a *t-wise uniform* distribution on its satisfying assignments. Together with recent work of Lee et al. [LRS15], our result also implies that *any* polynomial-size semidefinite programming relaxation for refutation requires at least $\tilde{\Omega}(n^{(t+1)/2})$ constraints.

More generally, we consider the δ -refutation problem, in which the goal is to certify that at most a $(1 - \delta)$ -fraction of constraints can be simultaneously satisfied. We show that if P is δ -close to supporting a *t-wise uniform* distribution on satisfying assignments, then the degree- $\Theta(\frac{n}{\Delta^{2/(t-1)} \log \Delta})$ SOS algorithm cannot $(\delta + o(1))$ -refute a random instance of $\text{CSP}(P)$. This is the first result to show a distinction between the degree SOS needs to solve the refutation problem and the degree it needs to solve the harder δ -refutation problem.

Our results (which also extend with no change to CSPs over larger alphabets) subsume all previously known lower bounds for semialgebraic refutation of random CSPs. For every constraint predicate P , they give a three-way hardness tradeoff between the density of constraints, the SOS degree (hence running time), and the strength of the refutation. By recent algorithmic results of Allen et al. [AOW15] and Raghavendra et al. [RRS16], this full three-way tradeoff is *tight*, up to lower-order factors.

*Princeton University and IAS. kothari@cs.princeton.edu

†Department of Mathematical and Computing Sciences, Tokyo Institute of Technology. mori@is.titech.ac.jp

‡Computer Science Department, Carnegie Mellon University. Supported by NSF grant CCF-1618679. [{odonnell,dwitmer}@cs.cmu.edu](mailto:odonnell,dwitmer}@cs.cmu.edu)

1 Introduction

Where are the hard problems?

In computational complexity, we have a comprehensive theory of worst-case hardness, assuming $P \neq NP$. The theory is particularly rich in the context of constraint satisfaction problems (CSPs) — optimization tasks that are both simple to state and powerfully expressive. (See, e.g., [BJK05, Rag08].) But despite our many successes in the theory of NP-completeness and NP-hardness-of-approximation, we know relatively little about the nature of hard instances. For example, 3-SAT is conjecturally hard to solve — or even approximate to factor $\frac{7}{8} + \epsilon$ — in $2^{o(n)}$ time. But what do hard(-seeming) instances look like? How can we generate one? These sorts of questions are a key part of understanding what makes various algorithmic problems truly hard. They are particularly important for CSPs, as these are nearly always the starting point for hardness reductions; the ability to find hard instances for CSPs yields the ability to find hard instances for many other algorithmic problems.

In some sense, a single instance can never be “hard” because its solution can always be hard-coded into an algorithm. Thus it is natural to turn to *random* instances, and the theory of average-case hardness. Uniformly random instances of CSPs are a particularly simple and natural source of hard(-seeming) instances. Furthermore, they arise as the fundamental object of study in many disparate areas of research, including cryptography [ABW10], proof complexity [BSB02], hardness of approximation [Fei02], learning theory [DLSS14], SAT-solving [SAT], statistical physics [CLP02], and combinatorics.

1.1 Random CSPs

Let Ω be a finite alphabet and let \mathcal{P} be a collection of nontrivial predicates $\Omega^k \rightarrow \{0, 1\}$. An input \mathcal{I} to the problem $\text{CSP}(\mathcal{P})$ consists of n variables x_1, \dots, x_n , along with a list \mathcal{E} of m constraints (P, S) , where P is a predicate from \mathcal{P} , and $S \in [n]^k$ is a scope of k distinct variables. We often think of the associated “factor graph”: that is, the bipartite graph with n “variable-vertices”, m “constraint-vertices” of degree k , and edges defined by the scopes.

Given \mathcal{I} , the algorithmic task is to find an assignment to the variables so as to maximize the fraction of satisfied constraints, $\text{avg}_{(P,S) \in \mathcal{E}} P(x_{S_1}, \dots, x_{S_k})$. We write $\text{Opt}(\mathcal{I})$ for the maximum possible fraction, and say that \mathcal{I} is *satisfiable* if $\text{Opt}(\mathcal{I}) = 1$. For a fixed *constraint density* $\Delta = \Delta(n) > 0$, a *random* instance of $\text{CSP}(\mathcal{P})$ is defined simply by choosing $m = \Delta n$ constraints uniformly at random: random scopes and random $P \in \mathcal{P}$.

The most typical examples involve a binary alphabet $\Omega = \{0, 1\}$, a fixed predicate $P : \{0, 1\}^k \rightarrow \{0, 1\}$, and $\mathcal{P} = P^\pm$, where by P^\pm we mean the collection of all 2^k predicates obtained by letting P act on possibly-negated input bits (“literals”). For example, if P is the k -bit logical OR function, then $\text{CSP}(P^\pm)$ is simply the k -SAT problem. In this introductory section, we’ll focus mainly on these kinds of CSPs.

For random CSPs, the constraint density Δ plays a critical role; naturally, the larger it is, the more likely \mathcal{I} is to be unsatisfiable. For a fixed \mathcal{P} , it is easy to show the existence of constants $\alpha_0 < \alpha_1$ such that when $\Delta < \alpha_0$, a random instance \mathcal{I} of $\text{CSP}(P)$ is satisfiable with high probability (whp), and when $\Delta > \alpha_1$, \mathcal{I} is unsatisfiable whp. For most interesting \mathcal{P} , it is conjectured that there is even a *sharp threshold* $\alpha_0 = \alpha_1 = \alpha_c$. (This has been proven for k -SAT with k large enough [DSS15]. See [CD09] for a characterization of those Boolean CSPs for which a sharp threshold is expected.)

For random instances with subcritical constraint density, $\Delta < \alpha_c$, the natural algorithmic task is to try to efficiently find satisfying assignments. There have been quite a few theoretical and

practical successes for this problem, for Δ quite large and even approaching α_c [Gab16, MPRT16]. On the other hand, for random instances with supercritical constraint density, $\Delta > \alpha_c$, the natural algorithmic task is to try to efficiently *refute* them; i.e., produce a certificate of unsatisfiability. For many CSPs, this task seems much harder, even heuristically. For example, random 3-SAT instances are unsatisfiable (whp) once $\Delta > 4.49$ [DKMPG08]; however, even for Δ as large as $n^{.49}$ there is no known algorithm that efficiently refutes random instances — even heuristically/experimentally. Thus the refutation task for random instances of CSPs with many constraints may be a source of simple-to-generate, yet hard-to-solve problems.

1.2 The importance and utility of hardness assumptions for random CSPs

In this section, we discuss the task of refuting random CSP instances and the importance of understanding the “constraint density vs. running time vs. refutation strength tradeoff” for *all* predicate families \mathcal{P} . To define our terms, a (*weak*) *refutation algorithm* for $\text{CSP}(\mathcal{P})$ is an algorithm that takes as input an instance \mathcal{I} and either correctly outputs “unsatisfiable”, or else outputs “don’t know”. For a given density Δ (larger than the critical density), we say the algorithm “succeeds” if it outputs “unsatisfiable” with high probability (over the choice of \mathcal{I} , and over its internal coins, if any). More generally, we can consider refutation algorithms that always output a correct upper bound on $\text{Opt}(\mathcal{I})$; we call them δ -*refutation algorithms* if they output an upper bound of $1 - \delta$ (or smaller) with high probability. The case of $\delta = 1/m$, where $m = \Delta n$ is the number of constraints, corresponds to the simple weak refutation task described earlier (with an output of “1” corresponding to “don’t know”). In general, we refer to δ as the “strength” of the refutation.

For a wide variety of areas — cryptography, learning theory, and approximation algorithms — it is of significant utility to have concrete hardness assumptions concerning random CSPs. Because uniformly random CSPs are very simply and concretely defined, they form an excellent basis for constructing other potentially hard problems by reduction. An early concrete hypothesis comes from an influential paper of Feige [Fei02]:

Feige’s R3SAT Hypothesis. *For every small $\delta > 0$ and for large enough constant Δ , there is no polynomial-time algorithm that succeeds in δ -refuting random instances of 3-SAT.*

Feige’s main motivation was hardness of approximation; e.g., he showed that the R3SAT Hypothesis implies stronger hardness of approximation results than were previously known for several problems (Balanced Bipartite Clique, Min-Bisection, Dense k -Subgraph, 2-Catalog). By reducing from these problems, several more new hardness of approximation results based on Feige’s Hypothesis have been shown in a variety of domains [BKP04, DFHS06, Bri08, AGT12]. Feige [Fei02] also related hardness of refuting 3-SAT to hardness of refuting 3-XOR. The assumption that refuting 3-XOR is hard has been used to prove new hardness results in subsequent work [OWWZ14]. Alekhovich [Ale03] further showed that certain average-case hardness assumptions for XOR imply additional hardness results, as well as the existence of secure public key cryptosystems.

In even earlier cryptography work, Goldreich [Gol00] proposed using the average-case hardness of random CSPs as the basis for candidate one-way functions. Subsequent work (e.g., [MST03]) suggested using similar functions as candidate pseudorandom generators (PRGs). The advantage of this kind of construction is the extreme simplicity of computing the PRG: indeed, its output bits can be computed in NC^0 , constant parallel time. Further work investigated variations and extensions of Goldreich’s suggestion [ABW10, ABR12, AL16]; see Applebaum’s survey [App13] for many more details. Of course, the security of these candidate cryptographic constructions depends heavily on the hardness of refuting random CSPs. Applebaum, Ishai, and Kushilevitz [AIK06] took

a slightly different approach to showing that PRGs exist in NC^0 , instead basing their result on one of Alekhovich’s average case XOR hardness assumptions [Ale03].

Finally, a recent exciting sequence of works due to Daniely and coauthors [DLSS13, DLSS14, DS14, Dan15] has linked hardness of random CSPs to hardness of learning. By making concrete conjectures about the hardness of refuting random $\text{CSP}(\mathcal{P})$ for various \mathcal{P} and for superpolynomial Δ , they obtained negative results for several longstanding problems in learning theory, such as learning DNFs and learning halfspaces with noise.

1.3 Desiderata for hardness results

While Feige’s R3SAT Hypothesis has proven useful in hardness of approximation, there are several important strengthenings of it that would lead to even further utility. We discuss here four key desiderata for hardness results about random CSPs:

- 1. Predicates other than SAT.** The hardness of random 3-SAT and 3-XOR has been most extensively studied, but for applications it is quite important to consider other predicates. For hardness of approximation, already Feige [Fei02] noted that he could prove stronger inapproximability for the 2-Catalog problem assuming hardness of refuting random k -AND for large k . Subsequent work has used assumptions about the hardness of refuting CSPs with other predicates to prove additional worst-case hardness results [GL04, AAM⁺11, CMVZ12, BCMV12, RSW16]. Relatedly, Barak, Kindler, and Steurer [BKS13] have recently considered a generalization of Feige’s Hypothesis to all Boolean predicates, in which the assumption is that the “basic SDP” provides the best δ -refutation algorithm when $\Delta = O(1)$. They also describe the relevance of predicates over larger alphabet sizes and with superconstant arity for problems such as the Sliding Scale Conjecture and Densest k -Subgraph. Bhaskara et al. [BCG⁺12] prove an SOS lower bound for Densest k -Subgraph via a reduction from Tulsiani’s SOS lower bound for random instances of $\text{CSP}(P)$ with P a q -ary linear code [Tul09]. A computational hardness assumption for refutation of this CSP would therefore give a hardness result for Densest k -Subgraph.

Regarding cryptographic applications, the potential security of Goldreich’s candidate PRGs depends heavily on what predicates they are instantiated with. Goldreich originally suggested a random predicate, with a slightly superconstant arity k . However algorithmic attacks on random $\text{CSP}(\mathcal{P})$ by Bogdanov and Qiao [BQ09] showed that predicates that are not at least “3-wise uniform” do not lead to secure PRGs with significant stretch. Quite a few subsequent works have tried to analyze what properties of a predicate family \mathcal{P} may — or may not — lead to secure PRGs [BQ09, ABR12, OW14, AL16].

Regarding the approach of Daniely et al. to hardness of learning, there are close connections between the predicates for which random $\text{CSP}(\mathcal{P})$ is assumed hard and the concept class for which one achieves hardness of learning. For example, the earlier work [DLSS14] assumed hardness of refuting random $\text{CSP}(P^\pm)$ for P being (i) the “Huang predicate” [Hua13, Hua14], (ii) Majority, (iii) a certain AND of 8 thresholds; it thereby deduced hardness of learning (i) DNFs, (ii) halfspaces with noise, (iii) intersections of halfspaces. Unfortunately, Allen et al. [AOW15] gave efficient algorithms refuting all three hardness assumptions; fortunately, the results were mostly recovered in later works [DS14, Dan15] assuming hardness of refuting random k -SAT and k -XOR. Although these are more “standard” predicates, a careful inspection of [DS14]’s hardness of learning DNF result shows that it essentially works by reduction from $\text{CSP}(P^\pm)$ where P is a “tribes” predicate. (It first shows hardness for this predicate by

reduction from k -SAT.) From these discussions, one can see the utility of understanding the hardness of random $\text{CSP}(\mathcal{P})$ for as wide a variety of predicates \mathcal{P} as possible.

2. **Superlinear number of constraints.** Much of the prior work on hardness of refuting random CSPs (assumptions and evidence for it) has focused on the regime of $\Delta = O(1)$; i.e., random CSPs with $O(n)$ constraints. However, it is quite important in a number of settings to have evidence of hardness even when the number of constraints is superlinear. An obvious case of this arises in the application to security of Goldreich-style PRGs; here the number of constraints directly corresponds to the stretch of the PRG. It's natural, then, to look for arbitrarily large polynomial stretch. In particular, having NC^0 PRGs with $m = n^{1+\Omega(1)}$ stretch yields secure two-party communication with constant overhead [IKOS08]. This motivates getting hardness of refuting random CSPs with $\Delta = n^{\Omega(1)}$. As another example, the hardness of learning results in the work of Daniely et al. [DLSS14, DS14, Dan15] all require hardness of refuting random CSPs with $m = n^C$, for arbitrarily large C . In general, given a predicate family \mathcal{P} , it is interesting to try to determine the least Δ for which refuting random $\text{CSP}(\mathcal{P})$ instances at density Δ becomes easy.
3. **Stronger refutation.** Most previous work on the hardness of refuting random CSPs has focused just on weak refutation (especially in the proof complexity community), or on δ -refutation for arbitrarily small $\delta > 0$. The latter framework is arguably more natural: as discussed in [Fei02], seeking just weak refutation makes the problem less robust to the precise model of random instances, and requiring δ -refutation for some $\delta > 0$ allows some more natural CSPs like k -XOR (where unsatisfiable instances are easy to refute) to be discussed. In fact, it is natural and important to study δ -refutation for *all* values of δ . As an example, given \mathcal{P} it is easy to show that there is a large enough constant Δ_0 such that for any $\Delta \geq \Delta_0$ a random instance \mathcal{I} of $\text{CSP}(\mathcal{P})$ has $\text{Opt}(\mathcal{I}) \leq \mu_{\mathcal{P}} + o(1)$, where $\mu_{\mathcal{P}}$ is the probability a random assignment satisfies a random predicate $P \in \mathcal{P}$. Thus it is quite natural to ask for δ -refutation for $\delta = 1 - \mu_{\mathcal{P}} - o(1)$; i.e., for an algorithm that certifies the *true* value of $\text{Opt}(\mathcal{I})$ up to $o(1)$ (whp). This is sometimes termed *strong refutation*. As an example, Barak and Moitra [BM16] show hardness of tensor completion based on hardness of strongly refuting random 3-SAT with $\Delta \ll n^{1/2}$. In general, there is a very close connection between refutation algorithms for $\text{CSP}(\mathcal{P})$ and approximation algorithms for $\text{CSP}(\mathcal{P})$; e.g., hardness of δ -refutation results for LP- and SDP-based proof systems can be viewed as saying that random instances are $1 - \delta$ vs. $\mu_{\mathcal{P}} + o(1)$ *integrality gap* instances for $\text{CSP}(\mathcal{P})$.
4. **Hardness against superpolynomial time.** Naturally, we would prefer to have evidence against superpolynomial-time refutation, or even subexponential-time refutation, of random $\text{CSP}(\mathcal{P})$; for example, this would be desirable for cryptography applications. This desire also fits in with the recent surge of work on hardness assuming the Exponential Time Hypothesis (ETH). We already know of two works that use a strengthening of the ETH for random CSPs. The first, due to Khot and Moshkovitz [KM16], is a candidate hard Unique Game, based on the assumption that random instances of $\text{CSP}(P^{\pm})$ require time $2^{\Omega(n)}$ to strongly refute, where P is the k -ary ‘‘Hadamard predicate’’. The second, due to Razenshteyn et al. [RSW16] proves hardness for the Weighted Low Rank Approximation problem assuming that refuting random 4-SAT requires time $2^{\Omega(n)}$. An even further interesting direction, in light of the work of Feige, Kim, and Ofek [FKO06], is to find evidence against efficient *nondeterministic* refutations of random CSPs.

These discussions lead us to the following goal:

Goal: *For every predicate family \mathcal{P} , provide strong evidence for the hardness of refuting random instances of $\text{CSP}(\mathcal{P})$, with the best possible tradeoff between number of constraints, refutation strength, and running time.*

The main theorem in this work, stated in Section 1.5, completely accomplishes this goal in the context of the Sum of Squares (SOS) method. Before stating our results, we review this method, as well as prior results in the direction of the above goal.

1.4 Prior results in proof complexity, and the SOS method

Absent the ability to even prove $\text{P} \neq \text{NP}$, the most natural way to get evidence of hardness for refuting random $\text{CSP}(\mathcal{P})$ is to prove unconditional negative results for specific proof systems. It’s particularly natural to consider automatizable proof systems, as these correspond to efficient deterministic refutation algorithms.

Much of the work in this area has focused on random instances of k -SAT. A seminal early work of Chvátal and Szemerédi [CS88] showed that Resolution refutations of random instances of k -SAT require exponential size when Δ is a sufficiently large constant. Ben-Sasson and Wigderson [BSW01, BS01] later strengthened this result to show that Resolution refutations require width $\Omega(\frac{n}{\Delta^{1/(k-2)+\epsilon}})$ for any $\epsilon > 0$. Ben-Sasson and Impagliazzo and Alekhnovich and Razborov further extended these results to the Polynomial Calculus proof system [BSI99, AR01]; for example, the latter work showed that Polynomial Calculus refutations of random k -SAT instances with density Δ require degree $\Omega(\frac{n}{\Delta^{2/(k-2)} \log \Delta})$.

On the other hand, much of the positive work on refuting random k -SAT has used spectral techniques and semialgebraic proof systems. These latter proof systems are often automatizable using linear programming and semidefinite programming, and thereby have the advantage that they can naturally give stronger δ -refutation algorithms. As examples, Goerdt and Krivelevich [GK01] showed that spectral techniques (which can be captured by SDP hierarchies) enable refutation of random k -SAT with $m = n^{\lceil k/2 \rceil}$ constraints; Friedman and Goerdt [FG01] improved this to $m = n^{3/2+o(1)}$ in the case of random 3-SAT. One of the first lower bounds for random CSPs using SDP hierarchies was given by Buresh-Oppenheim et al. [BOGH⁺03]; it showed that the Lovász–Schrijver₊ (LS₊) proof system cannot refute random instances of k -SAT with $k \geq 5$ and constant Δ . Alekhnovich, Arora, and Turlakis [AAT05] extended this result to random instances of 3-SAT.

The strongest results along these lines involve the Sum of Squares (AKA Positivstellensatz or Lasserre) proof system. This system, parameterized by a tuneable “degree” parameter d , is known to be very powerful; e.g., it generalizes the degree- d Sherali–Adams₊ (SA₊) and LS₊ proof systems. In the context of $\text{CSP}(\mathcal{P})$ over domain $\{0, 1\}$, it is also (approximately) automatizable in $n^{O(d)}$ time using semidefinite programming. As such, it has proven to be a very powerful positive tool in algorithm design, both for CSPs and for other tasks; in particular, it has been used to show that several conjectured hard instances for CSPs are actually easy [BBaH⁺12, OZ13, KOTZ14]. Finally, thanks to work of Lee, Raghavendra, and Steurer [LRS15], it is known that constant-degree SOS approximates the optimum value of CSPs at least as well as *any* polynomial-size family of SDP relaxations. See, e.g., [OZ13, BS14, Lau09] for surveys concerning SOS.

Early on, Grigoriev [Gri01] showed that SOS of degree $\Omega(n)$ could not refute k -XOR instances on sufficiently good expanders. Schoenebeck [Sch08] essentially rediscovered this proof and showed that it applied to random instances of k -SAT and k -XOR, specifically showing that SOS degree $\frac{n}{\Delta^{2/(k-2)-\epsilon}}$ is required to refute instances with density Δ . Tulsiani [Tul09] extended this result to the alphabet- q generalization of random 3-XOR.

Much less was previously known about predicates other than k -SAT and k -XOR. Austrin and

Mossel [AM08] established a connection between hardness of $\text{CSP}(\mathcal{P})$ and pairwise-uniform distributions, showing inapproximability beyond the random-threshold subject to the Unique Games Conjecture. A key work of Benabbas et al. [BGMT12] showed an unconditional analog of this result: random instances of $\text{CSP}(P^\pm)$ with sufficiently large constant constraint density require $\Omega(n)$ degree to refute in the SA_+ SDP hierarchy when P is a predicate (over any alphabet) supporting a pairwise-uniform distribution on satisfying assignments. O’Donnell and Witmer [OW14] extended these results by observing a density/degree tradeoff: they showed that if the predicate supports a $(t - 1)$ -wise uniform distribution, then the SA LP hierarchy at degree $n^{\Omega(\epsilon)}$ cannot refute random instances of $\text{CSP}(P^\pm)$ with $m = n^{t/2-\epsilon}$ constraints. They also showed the same thing for the SA_+ SDP hierarchy, provided one can remove a carefully chosen $o(m)$ constraints from the random instance. Extending results of Tulsiani and Worah [TW13], Mori and Witmer [MW16] showed this result for the SA_+ and LS_+ SDP hierarchies, for purely random instances. Finally, Barak, Chan, and Kothari [BCK15] recently extended the [BGMT12] result to the SOS system, though not for purely random instances: they showed that for any Boolean predicate P supporting a pairwise-uniform distribution, if one chooses a random instance of $\text{CSP}(P^\pm)$ with large constant Δ and then carefully removes a certain $o(n)$ constraints, then SOS needs degree $\Omega(n)$ to refute the instance.

Beyond semialgebraic proof systems and hierarchies, even less is known about non-SAT, non-XOR predicates. Feldman, Perkins, and Vempala [FPV15] proved lower bounds for refutation of $\text{CSP}(P^\pm)$ using statistical algorithms when P supports a $(t - 1)$ -wise uniform distribution. Their results are incomparable to the above lower bounds for LP and SDP hierarchies: the class of statistical algorithms is quite general and includes any convex relaxation, but the [FPV15] lower bounds are not strong enough to rule out refutation by polynomial-size SDP and LP relaxations.

Summary. For the strongest semialgebraic proof system, SOS, our evidence of hardness for random CSPs from previous work was somewhat limited. We did not know any hardness results for a superlinear number of constraints, except in the case of k -SAT/ k -XOR and the alphabet- q generalization of 3-XOR. We did not know any results that differentiated weak refutation from δ -refutation. Finally, the results known for refuting $\text{CSP}(P^\pm)$ with pairwise-uniform-supporting P did not hold for purely random instances.

1.5 Our result

We essentially achieve the Goal described in Section 1.3 in the context of the powerful SOS hierarchy. Specifically, for every predicate family \mathcal{P} , we provide a *full three-way tradeoff between constraint density, SOS degree, and strength of refutation*. Our lower bound subsumes all of the hardness results for semialgebraic proof systems mentioned in the previous section. Furthermore, as we will describe, known algorithmic work implies that our full three-way hardness tradeoff is tight, up to lower-order terms.

To state our result, we need a definition. For a predicate $P : \Omega^k \rightarrow \{0, 1\}$ and an integer $1 < t \leq k$, we define $\delta_P(t)$ to be P ’s distance from supporting a t -wise uniform distribution. Formally,

$$\delta_P(t) := \min_{\substack{\mu \text{ is a } t\text{-wise uniform distribution on } \Omega^k, \\ \sigma \text{ is a distribution supported on satisfying assignments for } P}} d_{\text{TV}}(\mu, \sigma),$$

where $d_{\text{TV}}(\cdot, \cdot)$ denotes total variation distance.

We can now (slightly informally) state our main theorem in the context of Boolean predicates:

Theorem 1.1. *Let P be a k -ary Boolean predicate and let $1 < t \leq k$. Let \mathcal{I} be a random instance of $\text{CSP}(P^\pm)$ with $m = \Delta n$ constraints. Then with high probability, degree- $\tilde{\Omega}\left(\frac{n}{\Delta^{2/(t-1)}}\right)$ SOS fails to $(\delta_P(t) + o(1))$ -refute \mathcal{I} .*

Additionally, in the case that $\delta_P(t) = 0$, our result does not need the additive $o(1)$ in refutation strength. That is:

Theorem 1.2. *Let P be a k -ary predicate and let $\mathcal{C}(P)$ be the minimum integer $3 \leq \tau \leq k$ for which P fails to support a τ -wise uniform distribution. Then if \mathcal{I} is a random instance of $\text{CSP}(P^\pm)$ with $m = \Delta n$ constraints, with high probability degree- $\tilde{\Omega}\left(\frac{n}{\Delta^{2/(\mathcal{C}(P)-2)}}\right)$ SOS fails to (weakly) refute \mathcal{I} .*

Remark 1.3. We comment here on the (surprisingly mild) parameter-dependence hidden by the $\tilde{\Omega}(\cdot)$ and $o(1)$ in these bounds. See Section 7 for full details.

- In terms of Δ , the $\tilde{\Omega}(\cdot)$ is only hiding a factor of $\log \Delta$. Thus we get a full linear $\Omega(n)$ -degree lower bound for $m = O(n)$ in both theorems above.
- In terms of k , and t , the $\tilde{\Omega}(\cdot)$ is only hiding a factor of $1/(k2^{O(k/t)})$. There are a number of interesting cases where one may take $t = \Theta(k)$; for example, k -SAT, k -XOR, and $\text{XOR}_{k/2} \oplus \text{MAJ}_{k/2}$, a predicate often used in cryptography (e.g., it was suggested by [AL16] for as the basis for high-stretch PRGs in NC^0). In these cases, the dependence of the degree lower bound depends only *linearly* on k and thus, there's little loss in having k significantly superconstant.
- Indeed in this case of $t = \Theta(k)$, if we also have $\Delta = 2^{\Theta(k)}$ then the degree lower bound for weak refutation in Theorem 1.2 is $\Omega(n)$ for k as large as $\Omega(n)$; here, both $\Omega(\cdot)$'s hide only a *universal* constants. The regime of $\Delta = 2^{\Theta(k)}$ is the algorithmically hardest one for k -SAT, and thus in this very natural case we have a linear-degree lower bound even for $k = \Omega(n)$.
- The refutation strength $\delta_P(t) + o(1)$ in Theorem 1.1 is more precisely $\delta_P(t) + O(1/\sqrt{n})$ whenever $\Delta = n^{\Omega(1)}$.
- Theorem 1.1 also holds for predicates P with alphabet size $q > 2$, with absolutely no additional parameter dependence on q .

The full three-way tradeoff in Theorem 1.1 between constraint density, SOS degree, and strength of refutation is tight up to a polylogarithmic factor in the degree and an additive $o(1)$ term in the strength of the refutation. The tightness follows from the below theorem, which is an immediate consequence of the general δ -refutation framework of Allen et al. [AOW15] and the strong refutation algorithm for XOR due to Raghavendra, Rao, and Schramm [RRS16] (which fits in the SOS framework).

Theorem 1.4. (Follows from [AOW15, RRS16].) *Let P be a k -ary Boolean predicate and let $1 < t \leq k$. Let \mathcal{I} be a random instance of $\text{CSP}(P^\pm)$ with $m = \Delta n$ constraints. Then with high probability, degree- $\tilde{O}\left(\frac{n}{\Delta^{2/(t-2)}}\right)$ SOS does $(\delta_P(t) - o(1))$ -refute \mathcal{I} . Furthermore, with high probability degree- $O(1)$ SOS succeeds in $(\delta_P(2) - o(1))$ -refuting \mathcal{I} , provided Δ is at least some polylog(n).*

An example. As the parameters can be a little difficult to grasp, we illustrate our main theorem and its tightness with a simple example. Let P be the 3-bit predicate that is true if *exactly* one of its three inputs is true. The resulting 3-SAT variant $\text{CSP}(P^\pm)$ is traditionally called 1-in-3-SAT. Let us compute the $\delta(t)$ values. The uniform distribution on the odd-weight inputs is pairwise-uniform, and it only has probability mass $\frac{1}{4}$ off of P 's satisfying assignments. This is minimum possible, and therefore $\delta_{1\text{-in-3-SAT}}(2) = \frac{1}{4}$. The only 3-wise uniform distribution on $\{0,1\}^3$ is the fully uniform one, and it has probability mass $\frac{5}{8}$ off of P 's satisfying assignments; thus $\delta_{1\text{-in-3-SAT}}(3) = \frac{5}{8}$.

Let us also note that as soon as Δ is a large enough constant, $\text{Opt}(\mathcal{I}) \leq \frac{3}{8} + o(1)$ (with high probability, a qualifier we will henceforth omit). Furthermore, it's long been known [BSB02] that for $\Delta = O(\log n)$ there is an efficient algorithm that weakly refutes \mathcal{I} ; i.e., certifies $\text{Opt}(\mathcal{I}) < 1$. But what can be said about stronger refutation? Let us see what our Theorem 1.1 and its counterpart Theorem 1.4 tell us.

Suppose first that there are $m = n \text{ polylog}(n)$ constraints. Theorem 1.4 tells us that constant-degree SOS certifies $\text{Opt}(\mathcal{I}) \leq \frac{3}{4} + o(1)$. However our result, Theorem 1.1, says this $\frac{3}{4}$ cannot be improved: SOS cannot certify $\text{Opt}(\mathcal{I}) \leq \frac{3}{4} - o(1)$ until the degree is as large as $\tilde{\Omega}(n)$. (Of course at degree n , SOS can certify the exact value of $\text{Opt}(\mathcal{I})$.)

What if there are $m = n^{1.1}$ constraints, meaning $\Delta = n^1$? Our result says SOS still cannot certify $\text{Opt}(\mathcal{I}) \leq \frac{3}{4} - o(1)$ until the degree is as large as $n^8/O(\log n)$. On the other hand, as soon as the degree gets bigger than some $\tilde{O}(n^8)$, SOS *does* certify $\text{Opt}(\mathcal{I}) \leq \frac{3}{4} - o(1)$; in fact, it certifies $\text{Opt}(\mathcal{I}) \leq \frac{3}{8} + o(1)$.

Similarly (dropping lower-order terms for brevity), if there are $m = n^{1.2}$ constraints, SOS is stuck at certifying just $\text{Opt}(\mathcal{I}) \leq \frac{3}{4}$ up until degree n^6 , at which point it jumps to being able to certify the truth, $\text{Opt}(\mathcal{I}) \leq \frac{3}{8} + o(1)$. If there are $n^{1.49}$ constraints, SOS remains stuck at certifying just $\text{Opt}(\mathcal{I}) \leq \frac{3}{4}$ up until degree $n^{0.2}$. Finally (as already shown in [AOW15]), once $m = n^{1.5} \text{ polylog}(n)$, constant-degree SOS can certify $\text{Opt}(\mathcal{I}) \leq \frac{3}{8} + o(1)$. **(End of example.)**

More generally, for a given predicate P and a fixed number of random constraints $m = n^{1+c}$, we provably get a “time vs. quality” tradeoff with an intriguing discrete set of breakpoints: With constant degree, SOS can $\delta_P(2)$ -refute, and then as the degree increases to n^{1-2c} , n^{1-c} , $n^{1-2c/3}$, etc., SOS can $\delta_P(3)$ -refute, $\delta_P(4)$ -refute, $\delta_P(5)$ -refute, etc.

An alternative way to look at the tradeoff is by fixing the SOS degree to some n^ϵ and considering how refutation strength varies with the number of constraints. So for m between n and $n^{3/2-\epsilon/2}$ SOS can $\delta_P(2)$ -refute; for m between $n^{3/2-\epsilon/2}$ and $n^{2-\epsilon}$ SOS can $\delta_P(3)$ -refute; for m between $n^{2-\epsilon}$ and $n^{5/2-3\epsilon/2}$ SOS can $\delta_P(4)$ -refute; etc.

It is particularly natural to examine our tradeoff in the case of constant-degree SOS, as this corresponds to polynomial time. In this case, our Theorem 1.1 says that random $\text{CSP}(P^\pm)$ cannot be $(\delta_P(t) + o(1))$ -refuted when $m \ll n^{(t+1)/2}$, and it cannot even be weakly refuted when $m \ll n^{C(P)/2}$. Now by applying the work of Lee, Raghavendra, and Steurer [LRS15], we get the same hardness results for *any* polynomial-size SDP-based refutation algorithm. (See [LRS15] for precise definitions.)

Corollary 1.5. *Let P be a k -ary predicate, and fix a sequence of polynomial-size SDP relaxations for $\text{CSP}(P^\pm)$. If \mathcal{I} is a random instance of $\text{CSP}(P^\pm)$ with $m \leq \tilde{\Omega}(n^{C(P)/2})$ constraints, then whp the SDP relaxation will have value 1 on \mathcal{I} . Furthermore, if $m \leq \tilde{\Omega}(n^{(t+1)/2})$ (for $1 < t \leq k$), then whp the SDP relaxation will have value at least $1 - \delta_P(t) - o(1)$ on \mathcal{I} .*

The results in this corollary are tight up to the polylogs on m , by the SOS algorithms of [AOW15].

2 Technical framework

In Section 1, we described our results as being SOS lower bounds for random CSPs, with constraints chosen randomly from a fixed predicate family \mathcal{P} . However it is conceptually clearest to divorce our results from the “random CSP” model as quickly as possible.

- Our lower bound applies whenever the underlying factor graph (bipartite constraint/variable graph) does not contain certain small forbidden subgraphs, which we call “implausible” subgraphs. Granted, the only examples we know of such graphs are random graphs (whp). Further, the condition of “does not contain any implausible subgraphs” is highly related to the condition of “has very good vertex expansion”. Still, we believe the right way to think about the requirement is in terms of forbidden subgraphs.
- Our lower bound doesn’t really involve CSPs and constraints, per se. For each constraint-vertex f in the underlying factor graph, rather than assuming it comes equipped with a constraint predicate P applied to its vertex-variable neighbors, we assume it comes equipped with a probability distribution μ_f on assignments to its vertex-variable neighbors. We can have a different μ_f for every constraint-vertex f if we want (indeed, the constraints need not even have the same arity).
- Our SOS lower bounds now take the following form: Assume we are given a factor graph G with no implausible subgraphs, and assume each constraint-vertex f has an associated distribution μ_f that is t -wise uniform. Then the low-degree SOS proof system “thinks” that there is a global assignment to the variables such that, at every constraint-vertex f , the local assignment to the neighboring variable-vertices is in the support of μ_f . (Indeed, it “thinks” that there is a *probability distribution* on global assignments such that for almost all f , the marginal distribution on f ’s neighbors is equal to μ_f .)

Let us make some of these notions more precise.

2.1 Constraint satisfaction

Notation 2.1. We fix an *alphabet* Ω of cardinality $q \geq 2$, and a *maximum constraint arity* $K \geq 3$.

The reader is strongly advised to focus on the case $q = 2$, with $\Omega = \{\pm 1\}$, as the only real difficulty posed by larger alphabets is notational. Also, although we describe K as a maximum arity, there will be no loss in thinking of every constraint as having arity K .

Definition 2.2 (*t*-wise uniform distributions). A probability distribution μ on Ω^k is said to be *t*-wise uniform if its marginal on every subset of t coordinates is uniform.

Rather than our full Theorem 1.1 concerning δ -refutation, the reader is advised to mainly keep in mind our Theorem 1.2, which is concerned with (weak) refutation of CSPs for which the predicates support a $(\tau - 1)$ -wise uniform distribution. Given our proof of Theorem 1.2, the more general Theorem 1.1 will fall out fairly easily.

Notation 2.3. We fix an integer τ satisfying $3 \leq \tau \leq K$.

The reader is advised to focus on the simplest case of $\tau = 3$ (corresponding to predicates supporting *pairwise*-uniform distributions), as the value of τ makes no real difference to our proofs.

Notation 2.4 (Instance). The *instance* we work with consists of two parts: a *factor graph* and its *constraint distributions*. The factor graph, denoted G , is a bipartite graph with edges going between n *variable-vertices* and m *constraint-vertices*. For a constraint-vertex f we write $N(f)$ for the *neighborhood* of f , which we take to be an *ordered* list of the variable-vertices adjacent to f . We assume that the degree (“arity”) of every constraint-vertex f satisfies $\tau - 1 \leq |N(f)| \leq K$. Finally, each constraint-vertex f also comes with a constraint distribution μ_f on $\Omega^{N(f)}$. It is assumed that each μ_f is $(\tau - 1)$ -wise uniform.

To orient the reader vis-à-vis our description of CSPs in Section 1.1, consider our Theorem 1.2 in which we have $\text{CSP}(P^\pm)$ instances, where $P : \{\pm 1\}^k \rightarrow \{0, 1\}$ is a k -ary Boolean predicate with complexity $\mathcal{C}(P) = \tau$. This means there exists some $(\tau - 1)$ -wise uniform distribution μ on $\{\pm 1\}^k$ supported on satisfying assignments for P . Note that for any “literal pattern” $\ell \in \{\pm 1\}^k$, the distribution μ_ℓ gotten by negating inputs to μ according to ℓ is also $(\tau - 1)$ -wise uniform. In the $\text{CSP}(P^\pm)$ instance, to every constraint with literal pattern ℓ the associated “constraint distribution” will be μ_ℓ . (In the more general context of Theorem 1.1 where we have a k -ary predicate P with $\delta = \delta_P(t)$, this means there is some distribution μ on $\{\pm 1\}^k$ which is t -wise uniform and which is δ -close to being supported on P . We will take $\tau = t + 1$ and take the constraint distributions to be μ_ℓ again.)

2.2 Plausible factor graphs

As mentioned earlier, our SOS lower bounds will hold whenever the factor graph G has no “implausible” subgraphs. The meaning of this will be discussed in much greater detail in Section 4, but here we will give the briefest possible definition.

Notation 2.5. We introduce two parameters: $1 \leq \text{SMALL} \leq n/2$ and $0 < \zeta < 1$. (For the sake of intuition, the reader might think of, e.g., $\text{SMALL} = n^{\Omega(1)}$ and $\zeta = \frac{1}{\log n}$.) The parameters are assumed to satisfy $K \leq \zeta \cdot \text{SMALL}$.

Plausibility Assumption. *Henceforth the factor graph G is assumed to satisfy the following property: Let H be an edge-induced subgraph in which every constraint-vertex has minimum degree τ . Suppose H has c constraint-vertices, v variable-vertices, and e edges, with $c \leq 2 \cdot \text{SMALL}$. Then $(\tau - \zeta)c \geq 2(e - v)$.*

We call the subgraphs H for which the inequality holds *plausible* because they are indeed the ones that may plausibly show up when the factor graph G is randomly chosen:

Proposition 2.6. (Roughly stated; see Theorem 4.12 for a precise statement.) *A random G with constraint density Δ will satisfy the Plausibility Assumption whp provided $\text{SMALL} \ll \frac{n}{\Delta^{2/(\tau-2-\zeta)}}$.*

The Plausibility Assumption is highly similar to the assumption that G has good vertex-expansion, and indeed our proof of Theorem 4.12 in Appendix A is a completely standard variant of the well-known proof that random bipartite graphs have good vertex-expansion.

2.3 The Sum of Squares algorithm, and pseudoexpectations

We give a brief overview of the Sum of Squares algorithm/proof system here. For more general background see, e.g., [BS]; for more details germane to this paper, see Section 5.3.

The Sum of Squares (SOS) algorithm is a hierarchy of semidefinite programming-based relaxations applicable to *polynomial optimization problems*; i.e., maximizing an n -variate polynomial

subject to polynomial inequality and equality constraints. Each algorithm in the hierarchy is indexed by a parameter d known as the *degree* of the relaxation. Central to the algorithm is the concept of *pseudoexpectations* that describe the feasible points of the SOS algorithm of degree d .

Definition 2.7 (Pseudoexpectations). Given n indeterminates, a degree- d pseudoexpectation is a linear operator $\tilde{\mathbf{E}}$ on the space of real polynomials of degree at most d in those indeterminates, such that $\tilde{\mathbf{E}}[1] = 1$. We also generally want it to satisfy the *Positive Semidefiniteness* condition: $\tilde{\mathbf{E}}[p^2] \geq 0$ for every polynomial p of degree at most $d/2$.

Definition 2.8 (Pseudoexpectations satisfying an identity). A degree- d pseudoexpectation $\tilde{\mathbf{E}}$ is said to *satisfy a polynomial identity* “ $p = 0$ ” if, for every polynomial q with $\deg(p) + \deg(q) \leq d$, we have $\tilde{\mathbf{E}}[pq] = 0$.

Given a polynomial optimization problem — say, maximizing a polynomial p_1 subject to constraints $\{q_i = 0 : i \in [m]\}$ — the degree- d SOS relaxation maximizes $\tilde{\mathbf{E}}[p_1]$ over all degree- d pseudoexpectations $\tilde{\mathbf{E}}$ that satisfy the identities $\{q_i = 0 : i \in [m]\}$. A feasibility problem, in particular, would ask if there is a degree- d pseudoexpectation satisfying certain polynomial equality constraints. These SOS relaxations can be expressed using a semidefinite program (SDP) of size $n^{O(d)}$. The *Sum of Squares algorithm* refers to (approximately) solving the SDP, which can generally be done in $n^{O(d)}$ time.

As suggested by the name, pseudoexpectations generalize the notion of expectations with respect to a *probability distribution* on real indeterminate values satisfying the given polynomial identity constraints. In particular, if there is at least one real solution for the polynomial identity constraints, then *any* probability distribution on solutions yields a valid degree- d pseudoexpectation, for any d . However, even when the polynomial constraints have no real solution, there may well be pseudoexpectations of limited degree that satisfy all the constraints. As one would expect, as the degree d grows, the pseudoexpectations resemble actual expectations more and more. Indeed, if the constraints include that the n indeterminates are Boolean (“ $x_i^2 = x_i$ ” or “ $x_i^2 = 1$ ”) then every degree- $2n$ pseudoexpectation in fact corresponds to an actual distribution on real solutions.

In our context of CSPs, we can think of a constraint satisfaction problem $\mathcal{E} = \{(P_i, S_i)\}$ over n Boolean variables x_1, \dots, x_n as a polynomial feasibility problem, with (the arithmetization of) the constraints $P_i(x_{S_i}) = 1$ as polynomial identities. As we know, randomly chosen CSPs with $\Delta \gg 1$ are unsatisfiable whp; to show a lower bound on the degree- d SOS refutation algorithm amounts to showing that there exists a degree- d pseudoexpectation that satisfies all the constraints. In more casual terminology, we say that degree- d SOS “thinks” that the CSP is satisfiable.

2.4 Main result

We can now describe our main result with the terminology and set-up developed above.

Theorem 2.9 (Roughly stated; cf. Theorem 6.1.). *Suppose we are given an instance, with factor graph G satisfying the Plausibility Assumption, and constraint distributions μ_f for each constraint-vertex. Then for $D = \frac{1}{3}\zeta \cdot \text{SMALL}$, there exists a degree- D pseudoexpectation $\tilde{\mathbf{E}}$ on global variable assignments such that for every constraint-vertex f , the following (suitably encoded) polynomial identity is satisfied: “The marginal distribution on assignments to the variable-neighbors of f is supported within $\text{supp}(\mu_f)$.” (Indeed, for almost all f , a stronger identity is satisfied, that the marginal simply equals μ_f .)*

In particular, if our instance comes from an actual random CSP with predicates, where for each f the distribution μ_f is supported on satisfying assignments for the predicate at f , then the

degree- D SOS algorithm “thinks” that the CSP is completely satisfiable. This is of course despite the fact that, whp, the CSP is not satisfiable.

Given Proposition 2.6 and Theorem 2.9, we can now point out how the constraint density vs. SOS-degree tradeoff arises in our Theorem 1.2. For $\text{CSP}(P^\pm)$ with $\mathcal{C}(P) = \tau$ and Δn random constraints, we get an SOS lower bound for degree roughly $\zeta \cdot \frac{n}{\Delta^{2/(\tau-2-\zeta)}}$. The best choice of ζ is roughly $1/\log \Delta$, and this indeed yields a degree bound of $\tilde{\Omega}\left(\frac{n}{\Delta^{2/(\mathcal{C}(P)-2)}}\right)$. More precise details of parameter-setting are given in Section 7.

3 Sketch of our techniques

Throughout this section, we describe our techniques in the context of CSPs on n Boolean variables and k -ary predicates that are $(\tau - 1)$ -wise uniform. As stated before, almost all of our ideas are present in this special case. Our goal is to build a degree- d pseudoexpectation operator $\tilde{\mathbf{E}}$ as described in Theorem 2.9.

3.1 Constructing the pseudoexpectation

As in all previous works on CSP lower bounds for hierarchies, we use a variant of the natural pseudoexpectation introduced by Benabbas et al. [BGMT12]. This pseudoexpectation is always defined in terms of a certain “closure” operator on instance graphs; previous works have used slightly different notions of “closure”. Our method introduces yet another definition of closure that we believe is the “right” one; at the very least, it seems to be precisely the right definition for facilitating our proofs.

3.1.1 Closures

We can describe a pseudoexpectation by prescribing its values on the basis of monomials of degree at most d . We work with the Fourier basis; i.e., ± 1 notation.

In the context of CSPs, a natural way to come up with a pseudoexpectation is via the idea of *local distributions*. If $\tilde{\mathbf{E}}$ is a degree- d pseudoexpectation, then for every collection S of at most $d/2$ variables, $\tilde{\mathbf{E}}$ agrees with the expectation of an actual probability distribution. In particular, the pseudoexpectation of a monomial $x^S := \prod_{i \in S} x_i$ for $S \subseteq [n]$ (or indeed any function on S) can then be described as the expectation of x^S with respect to the local distribution η_S that $\tilde{\mathbf{E}}$ induces on the set S of variables. For such a definition to make sense, the local distributions must satisfy *consistency*: the pseudoexpectation of x^T should equal the expectation of x^T with respect to the local distribution η_S for any S that includes T and is of size at most d .

We would like to choose local distributions η_S that are supported on satisfying assignments of all constraints completely included in S (we call these the constraints covered by S). At first blush, we could choose the uniform distribution over the set of satisfying assignments for the constraints covered by S . However, this choice doesn’t satisfy the consistency constraints. The t -wise uniform distributions that are supported on satisfying assignments of the predicate P now come to our rescue: if we obtain a local probability distribution that induces μ on the literals of any constraint in our CSP instance, we should intuitively expect be in good shape because t -wise uniformity roughly guarantees that any constraint that intersects S in t or less variables has a satisfying assignment that agrees with the assignment sampled for S . A natural choice is to define the probability of an assignment to S to be the product of the probabilities (with respect to μ) of the partial assignments corresponding to the constraints covered by S . This doesn’t work as-is,

either: there could be constraints that intersect S in many variables and yet are not completely contained inside S . A sample from η_S thus might already force such a constraint to not be satisfied.

To correct for this, we want to collect all such “dependencies” before choosing the local distribution. Benabbas et al. [BGMT12] make this idea precise by defining a notion of *closure* for a set of variables S : intuitively, these are all the variables that one should care about when defining the local distribution on S . Concretely, their closure maps S into a larger set S' such that for any $T \supseteq S'$, the marginal of η_T on S is equal to the marginal of $\eta_{S'}$ on S . We then choose $\eta_{S'}$ to be the local distribution on S' and define η_S to be the marginal of $\eta_{S'}$ on S . For such an effort to be feasible, S' shouldn't be much bigger than S : if in the extreme case the closure happened to be the whole set of variables $[n]$, we cannot define a distribution on satisfying assignments of all constraints covered by S' .

The closure of Benabbas et al. [BGMT12] guarantees local consistency as we wanted. Local consistency is all that is required for showing a Sherali–Adams lower bound and is equivalent to the following local positivity condition, which is weaker than positive semidefiniteness: $\tilde{\mathbf{E}}[p] \geq 0$ for p for every truly nonnegative polynomial p depending on at most d variables. However, when trying to show that the more global $\tilde{\mathbf{E}}[p^2]$ positive-semidefiniteness condition holds, the [BGMT12] construction seems hard to analyze.

To address this problem, Barak, Chan, and Kothari [BCK15] introduced a simpler variant of the [BGMT12] closure in order to show that the $\tilde{\mathbf{E}}$ defined above satisfies the positive-semidefiniteness condition for certain pruned random instances of the CSP(P^\pm), when P supports a pairwise-uniform distribution. However, their definition of closure degenerates into the set of all variables with high probability when the random CSP has $\Delta = \omega(1)$.

Our closure. One of the main innovations in our work is the introduction of a new, simpler definition of closure that plays a key role in our proof of positive semidefiniteness and gives a definition of $\tilde{\mathbf{E}}$ that works even when the number of constraints is superlinear in n . In addition, our definition of closure enables us to extend our results to δ -refutation.

Our closure for a set of variables S is a subgraph of the factor graph of the CSP instance, including both variables and constraints. We think of the closure of S as being the set of variables and constraints that “matter” when defining the distribution η_S . Given that a predicate P supports a $(\tau - 1)$ -wise uniform distribution, any constraint that affects η_S must have at least $\tau - 1$ variables in S . Otherwise, $(\tau - 1)$ -wise uniformity implies that we could ignore such a constraint without changing η_S . Any variable v not in S that occurs in only one constraint isn't necessary for defining η_S , either. We could sum η_S over the two assignments to v to get a new distribution that no longer depends on v . This leads to a natural choice of the closure as the union of all small subgraphs of the factor graph such that each constraint contains at least $\tau - 1$ variables and each variable outside of S occurs in at least two constraints. For a formal definition, see Section 5.

3.2 Proving positivity

Once we have the definition of the pseudoexpectation, we get to the main challenge in showing any SOS lower bound: arguing positive-semidefiniteness of the $\tilde{\mathbf{E}}$ constructed. The high level idea in our analysis builds on the work of Barak, Chan and Kothari [BCK15]. Their idea of proving positive-semidefiniteness is simple. They begin by observing that it suffices to verify positive-semidefiniteness for a basis that satisfies *orthogonality* under $\tilde{\mathbf{E}}[\cdot]$, meaning, the pseudoexpectation of the product of any distinct pair of basis polynomials is 0.

Fact 3.1. *Suppose there exists a basis f_1, f_2, \dots for degree- d polynomials such that the following two properties hold:*

1. $\tilde{\mathbf{E}}[f_i f_j] = 0$ for all $i \neq j$.
2. $\tilde{\mathbf{E}}[f_i^2] \geq 0$ for all i .

Then $\tilde{\mathbf{E}}[g^2] \geq 0$ for all g of degree at most d .

Proof. Write g as $\sum_i a_i f_i$. Then $\tilde{\mathbf{E}}[g^2] = \sum_{i,j} a_i a_j \tilde{\mathbf{E}}[f_i f_j] = \sum_i a_i^2 \tilde{\mathbf{E}}[f_i^2] \geq 0$. \square

Notice that the standard Fourier monomial basis guarantees us positivity (since $\tilde{\mathbf{E}}$ satisfies the local Sherali–Adams positivity condition by construction). However, it is not orthogonal in general. How can we construct such a basis? One way to construct a basis that is orthogonal under $\tilde{\mathbf{E}}[\cdot]$ is to perform the Gram–Schmidt process on, say, the monomial basis $1, x_1, x_2, \dots, x_1 x_2, \dots$ to get a new basis f_1, f_2, \dots . Now, Property 1 above holds for this new basis by construction. However, the Gram–Schmidt process is highly sequential and, in particular, the basis function towards the end could depend on all n variables. Thus, we cannot appeal to local positivity of $\tilde{\mathbf{E}}$ in order to argue positive-semidefiniteness of the newly generated basis. It appears that we have made no progress, ensuring orthogonality but potentially losing positivity.

The idea of Barak et al. to escape this pitfall is to show that *local orthogonalization* is enough. Before the start of the Gram–Schmidt process, we fix an order on basis vectors. In each step of the process, one orthogonalizes a basis function against all previous basis functions in this order by subtracting off its projection onto their span. Barak et al. analyze the variant of this process in which one orthogonalizes a basis function x^S by subtracting off its projection onto the span of all basis functions that precede it in the order *and* are functions of variables that lie in a small “ball” around S in the factor graph G of the instance. This lets them ensure that the new basis satisfies positivity (since it now depends only on a small number of variables, one can appeal to the local positivity of $\tilde{\mathbf{E}}$), and they show that this relaxed variant of the Gram–Schmidt process still ensures orthogonality.

Their proof, however, is highly combinatorial and requires various assumptions on the factor graph of the instance that intuitively shouldn’t matter. In particular, they need that the factor graph have no small cycles (girth should be logarithmic): while this can be ensured by pruning $o(n)$ fraction of the constraints in a random instance with $\Theta(n)$ constraints, this proof strategy breaks down for super-linear number of constraints.

Our approach Our main idea simplifies the analysis without requiring the assumptions of [BCK15] and yields tight results. It also naturally extends to the case of t -wise uniform predicates and further to δ -approximate t -wise uniform predicates. We next describe our key technical ideas that makes this possible.

At a high level, our argument drops the *local orthogonalization* strategy of Barak et al. [BCK15] and instead runs the Gram–Schmidt procedure “as-is”. Thus orthogonality of the resulting basis functions is immediate, and we need only show positive-semidefiniteness. We show that for any sequential ordering of the basis monomials in the Gram–Schmidt procedure, so long as it is of increasing degree, whenever we orthogonalize a monomial x^S , the result basis function depends only on a small number of variables.

To see why such an assertion might be plausible, let us consider the task of orthogonalizing the singletons. The monomial basis may not be orthogonal under $\tilde{\mathbf{E}}[\cdot]$; e.g., consider the following 3-XOR

instance:

$$\begin{array}{ll}
x_1x_2x_3 = 1 & y_1y_2y_3 = -1 \\
x_2x_4x_5 = 1 & y_2y_4y_5 = -1 \\
x_4x_5x_6 = 1 & y_4y_5y_6 = -1 \\
x_6x_7x_8 = 1 & y_6y_7y_8 = -1 \\
x_3x_7x_8 = 1 & y_3y_7y_8 = -1
\end{array}$$

Observe that x_1 and y_1 each appear in exactly one constraint and all other variables each occur in exactly two constraints. Multiplying each block of constraints together, we see that if $\tilde{\mathbf{E}}[\cdot]$ satisfies all constraints then $\tilde{\mathbf{E}}[x_1] = 1$ and $\tilde{\mathbf{E}}[y_1] = -1$. So neither x_1 nor y_1 are orthogonal to 1. Since the two sets of equations are disjoint, we also know that $\tilde{\mathbf{E}}[x_1y_1] = -1$, so x_1 and y_1 are not orthogonal. We note that many such blocks may occur in a random instance with $m \gg n^{1.4}$ constraints. Let's try to understand what happens when we run the Gram-Schmidt procedure on this basis. Consider an instance consisting of n such disjoint blocks of 5 constraints on $8n$ variables. Let x_{i1} be the variables that is fixed in block i . Then every x_{i1} is not orthogonal to 1 and every pair x_{i1}, x_{j1} is not orthogonal. Intuitively, the variables x_{i1}, x_{j1} behave independently, but are biased. To fix this bias, consider the functions \bar{x}_{i1} (where we use the notation $\bar{z} := z - \tilde{\mathbf{E}}[z]$). Now we have that \bar{x}_{i1} is orthogonal to 1 and, by independence of the blocks, $\tilde{\mathbf{E}}[\bar{x}_{i1} \cdot \bar{x}_{j1}] = 0$ for all i, j .

Ideally, we might hope this this new basis satisfies orthogonality when we move to degree 2, as well. Unfortunately, in general the basis $\{1, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_n, \bar{x}_1\bar{x}_2, \dots\}$ again need not be orthogonal. Consider a 3-XOR instance with n constraints $x_0x_iy_i = b_i$ for $i \in [n]$; call this an n -star. Random instances contain stars of superconstant size with high probability. For all $\binom{n}{2}$ pairs i, j , it holds that \bar{x}_iy_i and \bar{x}_jy_j are not orthogonal under $\tilde{\mathbf{E}}[\cdot]$:

$$\tilde{\mathbf{E}}[\bar{x}_iy_i \cdot \bar{x}_jy_j] = \tilde{\mathbf{E}}[x_iy_i \cdot x_jy_j] - \tilde{\mathbf{E}}[x_iy_i] \tilde{\mathbf{E}}[x_jy_j] = b_ib_j - 0 = b_ib_j.$$

Instead, consider the basis

$$\widehat{1} = 1, \widehat{x}_0 = x_0, \widehat{x}_1 = x_1, \dots, \widehat{y}_1 = y_1, \widehat{y}_2 = y_2, \dots, \widehat{x}_1y_1 = x_1y_1 - b_1x_0, \widehat{x}_2y_2 = x_2y_2 - b_2x_0, \dots$$

A simple calculation shows that these basis functions are orthogonal. Each basis function depends on at most 3 variables, so the degree-3 Sherali-Adams positivity condition and Fact 3.1 imply that degree-2 positive semidefiniteness holds. We give a proof of orthogonality of \widehat{x}_iy_i and \widehat{x}_jy_j that illustrates the underlying intuition. Observe that \widehat{x}_iy_i and \widehat{x}_jy_j are independent *conditioned on* x_0 for all $i \neq j$, and we can write

$$\begin{aligned}
\tilde{\mathbf{E}}[\widehat{x}_iy_i \cdot \widehat{x}_jy_j] &= \mathbf{E}[\widehat{x}_iy_i \cdot \widehat{x}_jy_j] && (\tilde{\mathbf{E}}[\cdot] \text{ is a valid expectation on small sets}) \\
&= \mathbf{E}[\mathbf{E}[\widehat{x}_iy_i \cdot \widehat{x}_jy_j | x_0]] && (\text{law of total expectation}) \\
&= \mathbf{E}[\mathbf{E}[\widehat{x}_iy_i | x_0] \cdot \mathbf{E}[\widehat{x}_jy_j | x_0]] && (\text{conditional independence of } \widehat{x}_iy_i \text{ and } \widehat{x}_jy_j \text{ given } x_0).
\end{aligned}$$

Next, note that

$$\mathbf{E}[\widehat{x}_iy_i | x_0 = b] = \frac{1}{\Pr[x_0 = b]} \mathbf{E}[\widehat{x}_iy_i \cdot 1_{\{x_0=b\}}(x_0)],$$

where $1_{\{x_0=b\}}$ is the indicator function for $x_0 = b$. Since we have orthogonalized \widehat{x}_iy_i against all degree-1 basis functions and $1_{\{x_0=b\}}$ is a degree-1 polynomial, this expression is equal to 0. Therefore, $\mathbf{E}[\widehat{x}_iy_i | x_0] = 0$ and \widehat{x}_iy_i and \widehat{x}_jy_j are orthogonal. In this case, x_iy_i and x_jy_j are correlated

because they are connected by x_0 . After subtracting off their correlation with x_0 , the resulting functions are orthogonal and no longer correlated.

Let us now formalize this intuition and generalize it to higher degree. At a high level, our idea is to show that the Gram–Schmidt process produces a basis such that each new basis element depends only on a small number of variables. Let y_S be the result of applying the Gram–Schmidt process to x^S . If y_T appears in y_S with a nonzero coefficient, then it must be the case that $\tilde{\mathbf{E}}[x^S \cdot y_T] \neq 0$. That is, x^S and y_T are correlated under $\tilde{\mathbf{E}}[\cdot]$. We show that this correlation is “witnessed” by some small, “dense” subgraph containing many constraints covered by few variables. If y_S has many variables in its support, then there must be many such subgraphs. We show that the union of these subgraphs is dense enough to be “implausible”. This means that y_S cannot have too many variables in its support.

Our witness can be seen as a generalization of the connected sets in the degree-2 case discussed above. Call two sets of vertices c -connected if removing any set of $c - 1$ vertices cannot disconnect them. In the degree-1 case, nonzero correlation between x^S and y_T with $|S| = |T| = 1$ is witnessed by a small, dense, connected (1-connected) subgraph. In the degree-2 case after orthogonalizing against degree-1 terms, we expect based on the star example that if S and T are only 1-connected, then x^S and y_T will no longer be correlated. We show that nonzero correlation between x^S and y_T with $|S| = |T| = 2$ is then witnessed by a small, dense, 2-connected subgraph. In general, we show that nonzero correlation between x^S and y_T with $|S| = |T| = d$ is witnessed by a small, dense, d -connected subgraph. This stronger connectivity requirement enables us to show that these witness subgraphs and their unions are dense enough to be implausible if the support of a basis function grows too large. For details of this argument, see Section 6.

4 Forbidden subgraphs for the factor graph

Let us make a few definitions concerning factor graphs, after which we will elaborate on the “Plausibility Assumption”.

Definition 4.1 (Subgraphs). We call H a *subgraph* of G if it is an *edge-induced* subgraph; i.e., $H = G[A]$ for some subset A of the edges of G . We explicitly allow $A = \emptyset$ and hence $H = \emptyset$. The subgraph H need not be connected.

Notation 4.2. For H a subgraph, we write $\text{vbls}(H)$ for the set of variables appearing in H , $\text{cons}(H)$ for the set of constraints appearing in H , and $\text{edges}(H)$ for the set of edges appearing in H .

Notation 4.3. Given $f \in \text{cons}(H)$, we write $N_H(f) = \{i \in \text{vbls}(H) : (f, i) \in \text{edges}(H)\}$. Note that this is *not* necessarily the same thing as $N(f) \cap \text{vbls}(H)$.

We will typically measure the “size” of a subgraph by the number of constraints in it:

Definition 4.4 (Small subgraphs). We say that subgraph H is *small* if $|\text{cons}(H)| \leq \text{SMALL}$.

Now regarding the Plausibility Assumption, for intuition’s sake let us suppose we are concerned with weak refutation and degree- $O(1)$ SOS, as in Corollary 1.5. Thus we have some k -ary predicate P with $\mathcal{C}(P) = \tau$, and we are selecting a random CSP with slightly fewer than $n^{\tau/2}$ constraints; say $m = n^{(\tau-\zeta)/2}$. What does a random factor graph look like in this case? Which small subgraphs may appear? A quick-and-dirty method to analyze this is as follows. Consider the fixed small subgraph in Figure 1; call it H .

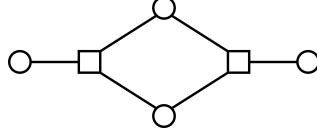


Figure 1: An example small subgraph. Constraint-vertices are squares, variable-vertices are circles.

What is the expected number of copies of H in a random factor graph G with n variable-vertices and $m = n^{(\tau-\zeta)/2}$ constraint-vertices? There are $\binom{m}{2} \approx m^2$ choices for H 's 2 constraint-vertices and $\binom{n}{4} \approx n^4$ choices for H 's 4 variable-vertices. Thinking of each constraint-vertex as choosing $k = O(1)$ random neighbors, the chance that the 6 edges of H show up is roughly n^{-6} . Thus, very roughly, we expect about $m^2 n^4 n^{-6} = n^{2 \cdot (\tau-\zeta)/2 + (4-6)}$ copies of H in a random G . Thus copies of H “plausibly” show up if and only if $2 \cdot (\tau - \zeta)/2 + (4 - 6) \geq 0$; i.e., if and only if $\tau \geq 2 + \zeta$. Since $\tau \geq 3$ always, this means we should certainly expect copies of H in G .

For a general subgraph H with $c = |\text{cons}(H)|$, $v = |\text{vbls}(H)|$, $e = |\text{edges}(H)|$,

$$\mathbf{E}[\# \text{ copies of } H] \approx m^c n^v n^{-e} = n^{c \cdot (\tau-\zeta)/2 + v - e} \implies H \text{ “plausibly occurs” iff } c \cdot (\tau-\zeta)/2 + (v-e) \geq 0. \quad (1)$$

This inequality is precisely the one occurring in the Plausibility Assumption from Section 2.2.

Despite the simple form of the inequality, we will find it helpful to view it in a different way. For reasons that will become clear in Section 5, we will be concerned almost exclusively with subgraphs of G in which all constraint-vertices have degree at least τ :

Definition 4.5 (τ -subgraphs). Let H be a subgraph. We will call H a τ -subgraph if every constraint-vertex in H has degree at least τ within H ; i.e., $|N_H(f)| \geq \tau$ for all $f \in \text{cons}(H)$.

Remark 4.6. The empty subgraph \emptyset is always trivially a τ -subgraph. Also, if H and H' are τ -subgraphs then so is $H \cup H'$.

Definition 4.7 (Leaf vertices and interior vertices). Given a subgraph H , we classify the variable-vertices in H as either *leaf* or *interior* depending on whether they have degree 1 or at least 2. (Since H is an edge-induced subgraph, it does not have any isolated vertices.)

For τ -subgraphs, there is a different way to view the “plausibility inequality” that will be more useful for us. We define it with some “accounting” terminology.

Definition 4.8 (Credit, debit, excess, revenue, cost, income). Let H be a τ -subgraph. For the purposes of this definition, consider each of its edges to be two directed edges.

- For each variable-vertex, we assign it a *credit* of 1 if it is a leaf vertex. We’ll write ℓ for the total credits.
- For each variable-vertex, any out-edges in excess of 2 are called *excess*, and we assign a *debit* for each. We’ll write e_v for the total number of these.
- For each constraint-vertex, any out-edges in excess of τ are called *excess*, and we assign a debit for each. We’ll write e_c for the total number of these, and $e = e_c + e_v$ for the total debit (number of excess edges).
- The sum of credits minus the sum of debits, $\ell - e$, is called the *revenue*. We denote it by $R(H)$.
- Each constraint-vertex has a *cost* of ζ . We write $C(H) = \zeta \cdot |\text{cons}(H)|$ for the total *cost*.

- The *income* is $I(H) = R(H) - C(H)$.

Definition 4.9 (Plausible τ -subgraphs). Let H be a τ -subgraph. We say that H is *plausible* if $I(H) \geq 0$.

Remark 4.10. H being plausible implies (indeed, is equivalent to) $|\text{cons}(H)| \leq \frac{1}{\zeta} \cdot R(H)$. Thus controlling a subgraph’s revenue is equivalent to controlling its size.

The next lemma implies that the inequality $I(H) \geq 0$ is the same as the inequality appearing in the Plausibility Assumption and in (1).

Lemma 4.11. *Let H be a τ -subgraph with $c = |\text{cons}(H)|$, $v = |\text{vbls}(H)|$, $e = |\text{edges}(H)|$, and $I = I(H)$. Then $e = \frac{\tau - \zeta}{2} \cdot c + v - \frac{I}{2}$.*

Proof. We count the number of “directed edges” in H . Counting those coming out of variable-vertices, the ℓ leaf vertices contribute 1 each, and the $v - \ell$ interior vertices contribute $2(v - \ell) + e_v$. Counting the directed edges coming out of constraint-vertices yields $\tau c + e_c$. Thus

$$\# \text{ directed edges} = 2e = \ell + 2(v - \ell) + e_v + \tau c + e_c = \tau c + 2v - (\ell - e) = \tau c + 2v - (\zeta c + I),$$

since $\ell - e = R(H) = C(H) + I(H)$. The claim follows. \square

In light of this, we may restate the Plausibility Assumption:

Plausibility Assumption, Restated. *Henceforth we assume the factor graph G has the following property: All τ -subgraphs H of G with $|\text{cons}(H)| \leq 2 \cdot \text{SMALL}$ are plausible.*

As mentioned earlier, for an appropriate choice of SMALL, the Plausibility Assumption holds for a random instance. More precisely, in Appendix A we prove the below theorem. The reader is advised that in this theorem, the first claim is the main one; it is used to show our Theorem 1.2 concerning weak refutation. The second claim (“Moreover...”) is a technical variant needed to extend our results to give Theorem 1.1 concerning δ -refutation.

Theorem 4.12. *Let $\lambda = \tau - 2 \geq 1$. Fix $0 < \zeta \leq .99\lambda$, $0 < \beta < \frac{1}{2}$. Then except with probability at most β , when \mathbf{G} is a random instance with $m = \Delta n$ constraints, the Plausibility Assumption holds provided*

$$\text{SMALL} \leq \gamma \cdot \frac{n}{\Delta^{2/(\lambda - \zeta)}},$$

where $\gamma = \frac{1}{K} \left(\frac{\beta^{1/\lambda}}{2^{K/\lambda}} \right)^{O(1)}$. Moreover, assuming $\zeta < 1$, except with probability at most β we have

$$\#\{\text{nonempty } \tau\text{-subgraphs } H \text{ with } \text{cons}(H) \leq 2 \cdot \text{SMALL} : I(H) \leq \tau - 1\} \leq \Delta n^{\frac{1+\zeta}{2}}.$$

5 Defining the pseudoexpectation

5.1 Closures

In this section we define the “closure” of a set of variables. Roughly speaking, this can be thought of as the smallest τ -subgraph of G that fully determines the distribution on S under a natural “planted distribution”.

Definition 5.1 (*S*-closed subgraph). Let S be a set of variables. We say that a subgraph H is *S*-closed if it is a τ -subgraph and all its leaf vertices are in S .

Remark 5.2. For every constraint in G , if H is taken to be the full neighborhood of that constraint, and S is the set of variables in that constraint, then H is *S*-closed.

Note that a union of *S*-closed τ -subgraphs is *S*-closed. This leads us to the following definition:

Definition 5.3 (Closure, $\text{cl}(S)$). Let S be a set of variables. We define the *closure* of S , written $\text{cl}(S)$, to be the union of all *small* *S*-closed τ -subgraphs H . Note that $\text{cl}(S)$ is itself an *S*-closed τ -subgraph.

Remark 5.4. A key warning to remember: we do not necessarily have $S \subseteq \text{vbls}(\text{cl}(S))$.

Remark 5.5. Let $T \subseteq S$. Then if H is *T*-closed, it is also *S*-closed. It follows that $\text{cl}(T) \subseteq \text{cl}(S)$.

Fact 5.6. *The only plausible \emptyset -closed τ -subgraph H is $H = \emptyset$. It follows that $\text{cl}(\emptyset) = \emptyset$.*

Proof. If H is \emptyset -closed then its revenue is at most 0. Hence if it is plausible, its cost is 0. \square

We will now give an important generalization of this fact for *S*-closures, $|S| > 0$

Theorem 5.7. *Let S be a set of variables with $|S| \leq \zeta \cdot \text{SMALL}$. Then $\text{cl}(S)$ is small and satisfies $R(\text{cl}(S)) \leq |S|$.*

Proof. Since $\text{cl}(S)$ is *S*-closed, all its leaf vertices are in S ; thus $\text{cl}(S)$ has at most $|S|$ credits and so $R(\text{cl}(S)) \leq |S|$, as claimed. Observe that if H_1, \dots, H_t is the complete list of *S*-closed τ -subgraphs, we may make the same deduction about $H_1 \cup \dots \cup H_j$ for any $1 \leq j \leq t$, in particular deducing that $R(H_1 \cup \dots \cup H_j) \leq \zeta \cdot \text{SMALL}$ for each j . The smallness of $\text{cl}(S)$ is now a consequence of the lemma that immediately follows. \square

Lemma 5.8. *Suppose that H is a τ -subgraph formed as a union, $H = H_1 \cup \dots \cup H_t$, where each H_j is small and where we have $R(H_1 \cup \dots \cup H_j) \leq \zeta \cdot \text{SMALL}$ for all $1 \leq j \leq t$. Then H is small.*

Proof. The proof is by induction on t , with the base case of $t = 1$ being immediate. In general, suppose $H' = H_1 \cup \dots \cup H_{t-1}$ is small. Since H_t is also small we have $|\text{cons}(H')|, |\text{cons}(H_t)| \leq \text{SMALL}$ and hence $|\text{cons}(H' \cup H_t)| \leq 2 \cdot \text{SMALL}$. Thus $H' \cup H_t$ is plausible and so

$$\text{cons}(H' \cup H_t) = (1/\zeta) \cdot C(H' \cup H_t) \leq (1/\zeta) \cdot R(H' \cup H_t) \leq (1/\zeta) \cdot \zeta \cdot \text{SMALL} = \text{SMALL},$$

showing that $H' \cup H_t$ is small, completing the induction. \square

In proving Theorem 5.7, we iteratively formed the union of all small *S*-closed subgraphs, at each step verifying that we have a small τ -subgraph of revenue at most $|S|$. Once we finish producing $\text{cl}(S)$ in this way, let $V = \text{vbls}(\text{cl}(S))$, and suppose we *continue* iteratively adding in small τ -subgraphs that are $(S \cup V)$ -closed. This process cannot add any leaf vertices except possibly in S ; thus we will still have that revenue is bounded by $|S| \leq \zeta \cdot \text{SMALL}$, and Lemma 5.8 will still imply the resulting τ -subgraph is small. Thus we end up with a small, *S*-closed τ -subgraph—which by definition is already contained in $\text{cl}(S)$. Thus we have shown:

Theorem 5.9. *Let S be a set of variables with $|S| \leq \zeta \cdot \text{SMALL}$. Then $\text{cl}(S \cup \text{vbls}(\text{cl}(S))) = \text{cl}(S)$.*

5.2 The planted distribution

Definition 5.10 (Planted distribution on a small subgraph). Let H be a small subgraph of G . The *planted distribution on H* is a probability distribution on assignments $\mathbf{x} \in \Omega^n$ to the variables of G , defined as follows: For each constraint $f \in \text{cons}(H)$ we independently draw an assignment $\mathbf{w}_f \in \Omega^{N(f)}$ according to μ_f . We write its component associated to variable $i \in N_H(f)$ as $\mathbf{w}_{f,i}$, and think of it as an assignment “suggested” for this variable. (Note that we will ignore the components of \mathbf{w}_f corresponding to variables not in $N_H(f)$.) Now each variable $i \in \text{vbls}(H)$ has one or more assignments in Ω suggested by its adjacent constraints. We get a unique assignment \mathbf{x}_i for it by *conditioning* on all the suggestions being consistent. (We will show later in (8) that this occurs with nonzero probability.) Finally, assignments for variables not in H are chosen independently and uniformly from Ω .

We’ll write η_H for the probability distribution on Ω^n associated to this planted distribution on H , and we’ll write $\mathbf{E}_H[\cdot]$ for the associated expectation.

Definition 5.11. For each $i \in \text{vbls}(G)$ and each $c \in \Omega$, we introduce an “indeterminate” $1_c(x_i)$ that is supposed to stand for 1 if variable i is assigned c and 0 otherwise.

The key theorem about the planted distributions is that as soon as a subgraph H contains $\text{cl}(S)$, the marginal of η_H on S is determined. In some sense, this property is exactly the reason we defined the closure the way we did.

Theorem 5.12. *Let S be a set of variables and let $H \supseteq \text{cl}(S)$ be a small subgraph. Then the marginal of η_H on S is the same as the marginal of $\eta_{\text{cl}(S)}$ on S .*

Remark 5.13. Although the notation in the below proof looks cumbersome, the calculations are actually fairly straightforward. We strongly encourage the reader to work through the proof in the case of $q = 2$, $\Omega = \{\pm 1\}$, with “ $\overline{1}_c(x_i)$ ” replaced by $cx_i \in \{\pm 1\}$.

Proof. For brevity we write $v_H = |\text{vbls}(H)|$ and $e_H = |\text{edges}(H)|$. We also introduce the notation $\overline{1}_c(x_i) = q1_c(x_i) - 1$. Recalling that η_H puts the uniform distribution on the $n - v_H$ variables outside H , we have

$$\begin{aligned} p_H(x) &:= \mathbf{Pr}[\mathbf{w} \text{ suggestions consistent, and the assignment to vbls}(H) \text{ they agree on is } x] \\ &= q^{-(n-v_H)} \cdot \mathbf{E}_{\mathbf{w}} \prod_{(f,i) \in \text{edges}(H)} (q^{-1}(1 + \overline{1}_{\mathbf{w}_{f,i}}(x_i))) \\ &= q^{v_H - e_H - n} \cdot \sum_{H' \subseteq H} \mathbf{E}_{\mathbf{w}} \prod_{(f,i) \in \text{edges}(H')} \overline{1}_{\mathbf{w}_{f,i}}(x_i) \tag{2} \\ &= q^{v_H - e_H - n} \cdot \sum_{H' \subseteq H} \prod_{f \in \text{cons}(H')} \mathbf{E}_{\mathbf{w}_f \sim \mu_f} \prod_{i \in N_{H'}(f)} \overline{1}_{\mathbf{w}_{f,i}}(x_i), \tag{3} \end{aligned}$$

where we used that the draws $\mathbf{w}_f \sim \mu_f$ are independent across f ’s. Now whenever $f \in \text{cons}(H')$ has $|N_H(f)| < \tau$, the $(\tau - 1)$ -wise uniformity of μ_f implies that

$$\mathbf{E}_{\mathbf{w}_f \sim \mu_f} \prod_{i \in N_{H'}(f)} \overline{1}_{\mathbf{w}_{f,i}}(x_i) = \mathbf{E}_{\substack{\mathbf{w}_{f,i} \sim \Omega \\ \text{uniform, indep.}}} \prod_{i \in N_{H'}(f)} \overline{1}_{\mathbf{w}_{f,i}}(x_i) = \prod_{i \in N_{H'}(f)} \mathbf{E}_{\substack{\mathbf{w}_{f,i} \sim \Omega \\ \text{uniform}}} \overline{1}_{\mathbf{w}_{f,i}}(x_i) = 0,$$

since $\mathbf{E}_{c \sim \Omega}[\overline{1}_c(x_i)] = 0$ for any fixed value $x \in \Omega$. Thus in (3) it is equivalent to sum over τ -subgraphs H' , and so returning to (2) we get

$$p_H(x) = q^{v_H - e_H - n} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ a } \tau\text{-subgraph}}} \mathbf{E}_{\mathbf{w}} \prod_{(f,i) \in \text{edges}(H')} \overline{1}_{\mathbf{w}_{f,i}}(x_i). \tag{4}$$

Suppose now that $T \subseteq [n]$ is a set of variables. We'll decompose an $x \in \Omega^n$ into its projection x_T onto the coordinates in T and $x_{\bar{T}}$ onto the coordinates not in T . Then

$$\begin{aligned}
p_H(x_T) &:= \Pr_{\mathbf{w}}[\mathbf{w} \text{ suggestions consistent, and the assignment to } T \text{ they agree on is } x_T] \\
&= \sum_{x_{\bar{T}} \in \Omega^{\bar{T}}} p_H(x_T, x_{\bar{T}}) = q^{n-|T|} \cdot \mathbf{E}_{\substack{x_{\bar{T}} \sim \Omega^{\bar{T}} \\ \text{uniform}}} [p_H(x_T, x_{\bar{T}})] \\
&= q^{v_H - e_H - |T|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ a } \tau\text{-subgraph}}} \mathbf{E}_{\mathbf{w}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in T}} \overline{1_{\mathbf{w}_{f,i}}}(x_i) \cdot \mathbf{E}_{\substack{x_{\bar{T}} \sim \Omega^{\bar{T}} \\ \text{uniform}}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in \bar{T}}} \overline{1_{\mathbf{w}_{f,i}}}(x_i), \quad (5)
\end{aligned}$$

where we used (4). Now suppose the τ -subgraph H' has a leaf vertex j that is in \bar{T} ; i.e., it's *not* in T . Then \mathbf{x}_j appears exactly once in the above, within the expression

$$\mathbf{E}_{\substack{x_{\bar{T}} \sim \Omega^{\bar{T}} \\ \text{uniform}}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in \bar{T}}} \overline{1_{\mathbf{w}_{f,i}}}(x_i). \quad (6)$$

As \mathbf{x}_j is chosen uniformly and independently of all random variables, the above contains a factor of the form $\mathbf{E}_{\mathbf{x}_j \sim \Omega} [\overline{1_{\mathbf{w}_{f,j}}}(x_j)]$. But for *any* fixed outcome of $\mathbf{w}_{f,j}$, this expectation is 0, meaning (6) will vanish. Thus any summand H' in (5) will vanish if H' has a leaf variable outside T . Thus we may equivalently sum only over T -closed H' . That is,

$$\begin{aligned}
p_H(x_T) &= \Pr_{\mathbf{w}}[\mathbf{w} \text{ suggestions consistent, and the assignment to } T \text{ they agree on is } x_T] \\
&= q^{v_H - e_H - |T|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ is } T\text{-closed}}} \mathbf{E}_{\mathbf{w}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in T}} \overline{1_{\mathbf{w}_{f,i}}}(x_i) \cdot \mathbf{E}_{\substack{x_{\bar{T}} \sim \Omega^{\bar{T}} \\ \text{uniform}}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in \bar{T}}} \overline{1_{\mathbf{w}_{f,i}}}(x_i) \\
&= q^{v_H - e_H - |T|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ is } T\text{-closed}}} \mathbf{E}_{\mathbf{w}} \mathbf{E}_{\substack{\mathbf{x} \sim \Omega^n \text{ unif.,} \\ \text{condit. on } \mathbf{x}_T = x_T}} \prod_{(f,i) \in \text{edges}(H')} \overline{1_{\mathbf{w}_{f,i}}}(x_i). \quad (7)
\end{aligned}$$

Suppose we took $T = \emptyset$ above. Since H is small, every subgraph H' is plausible and hence Fact 5.6 implies that the above has only one summand, corresponding to $H' = \emptyset$. The summand is trivially 1, and hence

$$\Pr_{\mathbf{w}}[\mathbf{w} \text{ suggestions consistent}] = q^{v_H - e_H}. \quad (8)$$

Observe that this does not depend at all on the μ_f 's; in particular, it is easily seen to be the probability of consistent suggestions under completely uniform μ_f 's. In any case, since (8) is positive, as promised, we may condition on the associated event; thus from (7) we obtain

$$\begin{aligned}
&\Pr_{\mathbf{w}}[\text{the suggested assignment to } S \text{ is } x_S \mid \text{the suggestions } \mathbf{w} \text{ are consistent}] \\
&= q^{-|S|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ is } S\text{-closed}}} \mathbf{E}_{\mathbf{w}} \mathbf{E}_{\substack{\mathbf{x} \sim \Omega^n \text{ unif.,} \\ \text{condit. on } \mathbf{x}_S = x_S}} \prod_{(f,i) \in \text{edges}(H')} \overline{1_{\mathbf{w}_{f,i}}}(x_i).
\end{aligned}$$

This formula visibly has the property that once $H \supseteq \text{cl}(S)$, it does not depend on H . \square

5.3 Pseudoexpectations

In this section, we formally define the pseudoexpectation with which we will work.

Definition 5.14. Given a polynomial expression $p(x)$ in the indeterminates $1_c(x_i)$, we write

$$\begin{aligned} \text{vbls}(p) &= \{i : \text{at least one } 1_c(x_i) \text{ appears in } p(x)\}, \\ \text{deg}_{\text{mlin}}(p) &= \max\{|\text{vbls}(M)| : M(x) \text{ is a monomial in } p(x)\}. \end{aligned}$$

We call the latter the *multilinear-degree*; note that $\text{deg}_{\text{mlin}}(p) \leq \text{deg}(p)$ always.

Recall that a *pseudoexpectation* on polynomials of degree at most D is a linear map $\tilde{\mathbf{E}}[\cdot]$ satisfying $\tilde{\mathbf{E}}[1] = 1$. We can uniquely define it by specifying its values on all monomials of degree at most D . Further, recall that if $p(x)$ is a polynomial, we say that $\tilde{\mathbf{E}}[\cdot]$ *satisfies the identity* $p(x) = 0$ if $\tilde{\mathbf{E}}[p(x) \cdot q(x)] = 0$ for all polynomials $q(x)$ with $\text{deg}(p \cdot q) \leq D$.

Definition 5.15 (Our pseudoexpectation). We'll define our pseudoexpectation $\tilde{\mathbf{E}}[\cdot]$ on all polynomials of *multilinear-degree* at most $\zeta \cdot \text{SMALL}$; in particular, this defines it for all polynomials of (usual) degree at most $\zeta \cdot \text{SMALL}$. We define it by imposing that $\tilde{\mathbf{E}}[M(x)] = \mathbf{E}_{\text{cl}(\text{vbls}(M))}[M(\mathbf{x})]$ for all monomials $M(x)$ having $\text{deg}_{\text{mlin}}(M) \leq \zeta \cdot \text{SMALL}$. (Here we are using the abbreviation $\mathbf{E}_C[M(\mathbf{x})]$ for $\mathbf{E}_{\mathbf{x} \sim \eta_C}[q(\mathbf{x})]$.) By Theorem 5.7, this makes sense in that $\text{cl}(\text{vbls}(M))$ will always be small. Note that we have $\tilde{\mathbf{E}}[1] = \mathbf{E}_{\text{cl}(\emptyset)}[1] = 1$, as required.

Theorem 5.16. *Let $p(x)$ be a polynomial expression of multilinear-degree at most $\zeta \cdot \text{SMALL}$. Let H be any small subgraph containing*

$$H' = \bigcup \{\text{cl}(\text{vbls}(M)) : M(x) \text{ is a monomial of } p(x)\}.$$

For example, if $\text{cl}(\text{vbls}(p))$ is small then it would qualify for H . Then

$$\tilde{\mathbf{E}}[p(x)] = \mathbf{E}_H[p(\mathbf{x})] = \mathbf{E}_{H'}[p(\mathbf{x})].$$

Proof. This is immediate from Theorem 5.12 and Remark 5.5. □

Theorem 5.17. *Let $p(x)$ be a polynomial with $S = \text{vbls}(p)$ satisfying $|S| \leq \text{deg}(p)$, $|S| \leq \zeta \cdot \text{SMALL}$. Assume that $p(\mathbf{x})$ is identically zero for $\mathbf{x} \sim \eta_{\text{cl}(S)}$. (Note that $\text{cl}(S)$ is small by Theorem 5.7.) Then our $\tilde{\mathbf{E}}[\cdot]$ satisfies the identity $p(x) = 0$.*

Proof. Let $q(x)$ be a nonzero polynomial with $\text{deg}(p \cdot q) \leq \zeta \cdot \text{SMALL}$. Writing $q(x) = \sum_j M_j(x)$ where each $M_j(x)$ is a monomial, we have

$$\tilde{\mathbf{E}}[p(x) \cdot q(x)] = \sum_j \tilde{\mathbf{E}}[p(x) \cdot M_j(x)] = \sum_j \mathbf{E}_{\text{cl}(S \cup \text{vbls}(M_j))}[p(\mathbf{x}) \cdot M_j(\mathbf{x})]. \quad (9)$$

Here the last equality used Theorem 5.16 and the smallness of $\text{cl}(S \cup \text{vbls}(M_j))$, which follows from Theorem 5.7 and the fact that $|S \cup \text{vbls}(M_j)| \leq \text{deg}(p) + \text{deg}(q) = \text{deg}(p \cdot q) \leq \zeta \cdot \text{SMALL}$. But since $\text{cl}(S \cup \text{vbls}(M_j)) \supseteq \text{cl}(S)$ (Remark 5.5), Theorem 5.12 tells us that $p(\mathbf{x})$ has the same distribution under $\eta_{\text{cl}(S \cup \text{vbls}(M_j))}$ and $\eta_{\text{cl}(S)}$; i.e., it is identically 0. Thus (9) vanishes, as needed. □

We have the following immediate corollaries:

Corollary 5.18. *Our pseudoexpectation $\tilde{\mathbf{E}}[\cdot]$ satisfies the following identities:*

- $\sum_{c \in \Omega} 1_c(x_i) = 1$ for all $i \in [n]$ (i.e., the identity $\sum_{c \in \Omega} 1_c(x_i) - 1 = 0$).
- $1_c(x_i)^2 = 1_c(x_i)$ for all $c \in \Omega, i \in [n]$.

As an immediate consequence of the latter, we always have $\tilde{\mathbf{E}}[p(x)] = \tilde{\mathbf{E}}[\text{multilin}(p(x))]$, where $\text{multilin}(p(x))$ is defined by replacing any positive power of $1_c(x_i)$ in $p(x)$ with just $1_c(x_i)$.

Another corollary is the following (cf. the rough statement of our main technical result, Theorem 2.9):

Corollary 5.19. *Our pseudoexpectation $\tilde{\mathbf{E}}[\cdot]$ satisfies the identity*

$$s_f(x) := \sum_{\vec{c} \in \text{supp}(\mu_f)} \prod_{i \in N(f)} 1_{c_i}(x_i) = 1$$

for all $f \in \text{cons}(G)$; i.e., “ $\tilde{\mathbf{E}}[\cdot]$ ’s distribution on $N(f)$ is always in $\text{supp}(\mu_f)$ ”.

Proof. We apply Theorem 5.17, with $S = N(f)$, which satisfies $|S| = \deg(s_f)$ and $|S| \leq K \leq \zeta \cdot \text{SMALL}$. Note that if H_f denotes the τ -subgraph induced by all edges of G incident on constraint-vertex f , then H_f is S -closed and so $H_f \subseteq \text{cl}(S)$. It then follows from the definition of $\mathbf{x} \sim \eta_{\text{cl}(S)}$ that $s_f(\mathbf{x}) \equiv 1$, since the restriction of \mathbf{x} to $N(f)$ will always be supported on $\text{supp}(\mu_f)$. \square

6 The proof of positive semidefiniteness

6.1 Setup

Throughout this section, fix a degree D satisfying $1 \leq D \leq \frac{1}{3}\zeta \cdot \text{SMALL}$. Our goal will be to establish:

Theorem 6.1. *If $p(x)$ is a polynomial expression of degree at most D , then $\tilde{\mathbf{E}}[p(x)^2] \geq 0$.*

In light of Corollary 5.18, we may assume that $p(x)$ is “multilinear” (i.e., does not contain $1_c(x_i)^k$ for any $k > 1$). Another way to state this assumption is $p(x) \in \text{span}(x^S : S \in \mathcal{M}^{\leq D})$, where we introduce the following notation:

Definition 6.2. A *monomial index* will be a set S of pairs $(i, c) \in [n] \times \Omega$, with no variable $i \in [n]$ occurring more than once. We write x^S for the monomial $\prod_{(i,c) \in S} 1_c(x_i)$, with the usual convention that $x^\emptyset = 1$. Finally, we write $\mathcal{M}^{\leq D}$ for the collection of monomial indices S with $|S| \leq D$.

Notation 6.3. We abuse notation as follows: If a monomial index S occurs in a place where a subset of variables is expected, we intend the subset of variables $\{i : (i, c) \in S \text{ for some } c\}$.

Remark 6.4. All of the ideas in our proof of Theorem 6.1 are present in the $q = 2$ case; only notational complexities arise for $q > 2$. Thus the reader is encouraged to keep the Boolean case $\Omega = \{\text{false}, \text{true}\}$ in mind. In this case, since $\tilde{\mathbf{E}}[\cdot]$ satisfies the identity $1_{\text{false}}(x_i) = 1 - 1_{\text{true}}(x_i)$, one can also ignore the indeterminate $1_{\text{false}}(x_i)$ (since $1 \in \text{span}(\mathcal{M}^{\leq 0})$ already). Then one can more naturally write the indeterminate $1_{\text{true}}(x_i)$ as x_i and the monomial x^S becomes $\prod_{i \in S} x_i$.

6.2 Gram–Schmidt overview

Notation 6.5. Let \preceq denote any total ordering on $\mathcal{M}^{\leq D}$ that respects cardinality, so that if T and S are monomial indices with $|T| < |S|$, then $T \prec S$. For $S \neq \emptyset$, let $\text{pr}(S)$ denote the immediate predecessor of S under \preceq .

Our goal in this section is to show that the *modified Gram–Schmidt process* from linear algebra can be successfully applied to the monomials $(x^S : S \in \mathcal{M}^{\leq D})$, in the ordering \preceq , using $\tilde{\mathbf{E}}[\cdot]$ as the “inner product”: $\langle p(x), q(x) \rangle := \tilde{\mathbf{E}}[p(x) \cdot q(x)]$. Of course, we don’t know that this is a genuine inner product (indeed, that’s essentially what we’re trying to prove). We will discuss this issue shortly, but we first remind the reader that the modified Gram–Schmidt process would typically produce a collection of polynomials $y_S = y_S(x)$, for $S \in \mathcal{M}^{\leq D}$, that are *orthogonal* under $\tilde{\mathbf{E}}[\cdot]$ (meaning $\tilde{\mathbf{E}}[y_S \cdot y_{S'}] = 0$ if $S \neq S'$) and that have the same span as $(x^S : S \in \mathcal{M}^{\leq D})$. As well, it would produce “normalized” versions of these polynomials $z_S = y_S / \sqrt{\tilde{\mathbf{E}}[y_S^2]}$, satisfying $\tilde{\mathbf{E}}[z_S^2] = 1$.

We now address the obviously difficulty that $\tilde{\mathbf{E}}[\cdot]$ is not (known to be) an inner product, because we don’t know it’s positive definite on the monomials of $\mathcal{M}^{\leq D}$. Our goal will be to show that as we follow the Gram–Schmidt process, it never encounters any “positive definiteness problems”, and therefore “succeeds”. The main “positive definiteness problem” Gram–Schmidt might encounter would be if it creates a polynomial y_S with $\tilde{\mathbf{E}}[y_S^2] < 0$. In this case, when it tries to produce the normalized polynomial z_S , it would certainly fail.

There is one additional potential problem, occurring if Gram–Schmidt produces a y_S with $\tilde{\mathbf{E}}[y_S^2] = 0$. In the usual process from linear algebra this may indeed occur, and the Gram–Schmidt algorithm copes by treating z_S as 0 (effectively, throwing it out of the span). This is a valid strategy because genuine inner products are strictly positive definite. However we only expect our “inner product” $\tilde{\mathbf{E}}[\cdot]$ to be positive *semidefinite*. We therefore need a different coping mechanism. For us, when $\tilde{\mathbf{E}}[y_S^2] = 0$ occurs, we will simply define its “normalized” version z_S to be y_S . The challenge of this is that Gram–Schmidt’s guarantee of producing an orthogonal collection $(y_S : S \in \mathcal{M}^{\leq D})$ relies syntactically on all the z_S polynomials satisfying $\tilde{\mathbf{E}}[z_S^2] = 1$. Thus we will have an additional burden: we will have to “manually” show that $\tilde{\mathbf{E}}[y_S^2] = 0$ implies that y_S is orthogonal under $\tilde{\mathbf{E}}[\cdot]$ to all other polynomials. It will count as a “positive definiteness problem” if we are unable to show this; we will call this the “pseudovariance zero problem”. We remark that the main positive definiteness problem is fundamentally more important than this “pseudovariance zero problem”, and the reader may wish to ignore the pseudovariance zero issue on first reading.

We now describe the modified Gram–Schmidt process in detail. The process works in *stages*, named after the elements of $\mathcal{M}^{\leq D}$ and in order of \preceq . At the end of stage S it creates a certain polynomial z_S . Stage \emptyset always “succeeds” and simply consists of defining $z_\emptyset = 1$. In some cases it may happen that $\tilde{\mathbf{E}}[z_S^2] = 0$. In this case we say that z_S has *pseudovariance zero*, and the Gram–Schmidt algorithm will add S to a growing collection called PvZ.

Each stage S is further divided into *substages*, associated to monomial indices $T \prec S$ in order of \preceq . Let us introduce some notation:

Notation 6.6. Let $\mathcal{M}_2^{\leq D}$ denote the collection of all pairs $(S, T) \in \mathcal{M}^{\leq D} \times \mathcal{M}^{\leq D}$ with $T \prec S$. We define a total ordering \preceq_2 on $\mathcal{M}_2^{\leq D}$ via

$$(S', T') \preceq_2 (S, T) \iff S' \prec S, \text{ or } S' = S \text{ and } T' \preceq T.$$

Thus the overall progression of substages in Gram–Schmidt is through the elements of $\mathcal{M}_2^{\leq D}$ in order of \preceq_2 . Substage (S, T) creates a polynomial $y_{S,T}$ as follows:

$$y_{S,T} = \begin{cases} x^S - \tilde{\mathbf{E}}[x^S] & \text{if } T = \emptyset; \\ y_{S,\text{pr}(T)} - \tilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T] z_T & \text{else.} \end{cases}$$

Stage S ends just after substage $(S, \text{pr}(S))$. At this point, the Gram–Schmidt process defines

$$y_S = y_{S,\text{pr}(S)}, \quad z_S = \begin{cases} y_S / \sqrt{\tilde{\mathbf{E}}[y_S^2]} & \text{if } \tilde{\mathbf{E}}[y_S^2] > 0; \\ y_S & \text{if } \tilde{\mathbf{E}}[y_S^2] = 0, \text{ in which case } S \text{ is placed into PvZ.} \end{cases}$$

Of course, if $\tilde{\mathbf{E}}[y_S^2] < 0$ then we have encountered a positive definiteness problem. Indeed, to be conservative we will treat it as a problem if $\tilde{\mathbf{E}}[y_{S,T}^2] < 0$ for *any* $(S, T) \in \mathcal{M}_2^{\leq D}$.

It is a syntactic property of the usual modified Gram–Schmidt process that when $y_{S,T}$ is produced, it is orthogonal to z_T under $\tilde{\mathbf{E}}[\cdot]$. However this relies on $\tilde{\mathbf{E}}[z_T^2] = 1$, which fails for us if $T \in \text{PvZ}$. Thus we will need to explicitly prove that $T \in \text{PvZ}$ implies $\tilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T] = 0$. If this doesn't hold, we've encountered the pseudovariance zero problem. But assuming it does hold, $y_{S,T}$ will simply become $y_{S,\text{pr}(T)}$ and we will have the desired orthogonality of $y_{S,T}$ and z_T . We remark that the usual Gram–Schmidt property of $y_{S,T}$ being orthogonal to *all* $z_{T'}$ with $T' \preceq T$ follows by induction in the usual way; this only needs the inductive property that the z_T 's are orthogonal (not that they're orthonormal).

We may now summarize the discussion so far:

Definition 6.7. A *positive definiteness problem* occurs at substage (S, T) of modified Gram–Schmidt if either $\tilde{\mathbf{E}}[y_{S,T}^2] < 0$, or if $T \in \text{PvZ}$ but $\tilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T] \neq 0$. (The latter is called a *pseudovariance zero* problem.) We say that the modified Gram–Schmidt process *succeeds through substage* (S, T) if it encounters no positive definiteness problem at any substage $(S', T') \preceq_2 (S, T)$.

Proposition 6.8. *Suppose the modified Gram–Schmidt process succeeds through substage (S, T) . Then we have:*

- $y_{S,T} = x^S - p(x)$ for some polynomial $p(x)$ supported on monomials $x^{T'}$ with $T' \preceq T$;
- $\tilde{\mathbf{E}}[y_{S,T} \cdot z_{T'}] = 0$ for all $T' \preceq T$, and hence $\tilde{\mathbf{E}}[y_{S,T} \cdot q(x)] = 0$ for all polynomials $q(x)$ supported on monomials $x^{T'}$ with $T' \preceq T$;
- $\tilde{\mathbf{E}}[y_{S,T}^2] \geq 0$.

In particular, if the process succeeds through stage S , we have:

- $z_S = c \cdot x^S - p(x)$ for some positive constant $c > 0$ and some polynomial $p(x)$ supported on monomials x^T with $T \prec S$;
- $\text{span}(x^{S'} : S' \preceq S) = \text{span}(z_{S'} : S' \preceq S)$;
- $\tilde{\mathbf{E}}[z_S \cdot z_T] = 0$ for all $T \prec S$, and hence $\tilde{\mathbf{E}}[z_S \cdot q(x)] = 0$ for all polynomials $q(x)$ supported on monomials $x^{T'}$ with $T' \prec S$;
- $\tilde{\mathbf{E}}[z_S^2] = 0$ if S is put in PvZ , else $\tilde{\mathbf{E}}[z_S^2] = 1$.

Our main Theorem 6.1 follows provided the modified Gram–Schmidt process succeeds through all substages in $\mathcal{M}_2^{\leq D}$. The reason is that then any multilinear $p(x)$ of degree at most D can be expressed as $p(x) = \sum_{|T| \leq D} c_T z_T$. This implies

$$\tilde{\mathbf{E}}[p(x)^2] = \sum_{|T|, |T'| \leq D} c_T c_{T'} \tilde{\mathbf{E}}[z_T \cdot z_{T'}] = \sum_{\substack{|T| \leq D \\ T \notin \text{PvZ}}} c_T^2 \geq 0,$$

using Proposition 6.8.

6.3 Advanced accounting

Definition 6.9. A τ -subgraph⁺ is defined to be a τ -subgraph, together with zero or more isolated variable-vertices.

We still have that the union of τ -subgraphs⁺ is a τ -subgraph⁺. We extend the $\text{cons}(H)$ and $\text{vbls}(H)$ notation to τ -subgraphs⁺, and also the planted distribution notation η_H (being the same as $\eta_{H'}$ where H' is formed from H by deleting its isolated vertices).

Definition 6.10. For a τ -subgraph⁺ H , we extend the definition of revenue by assigning *two* credits for all isolated variable-vertices in H .

Remark 6.11. If H is a τ -subgraph⁺ and H' is the τ -subgraph formed by deleting isolated vertices, then $\text{cons}(H') = \text{cons}(H)$, $C(H') = C(H)$, and $R(H') \leq R(H)$. Thus the Plausibility Assumption immediately implies that all τ -subgraphs⁺ with at most $2 \cdot \text{SMALL}$ constraints are also plausible.

Lemma 6.12. *Let H be a small τ -subgraph⁺ with $R(H) \leq r$. Let H' be a small τ -subgraph with at most s leaf variables that are not in H . Assume $r + s \leq \zeta \cdot \text{SMALL}$. Then $H \cup H'$ is small and satisfies $R(H \cup H') \leq r + s$.*

Proof. Adding H' into H cannot remove any of the debits of H , and the only additional credits that can be created come from the s leaf variables in H' that are not in H . (Since H' is only a τ -subgraph it has no isolated variables.) This establishes $R(H \cup H') \leq r + s$. The smallness conclusion follows immediately from Lemma 5.8 (here it does not matter that H is a τ -subgraph⁺). \square

A key aspect to our main theorem will be that in some cases this revenue bound can be improved:

Lemma 6.13. *In the setup of Lemma 6.12, suppose also that H' has b edges that are “boundary” for H , in the sense that each has exactly one endpoint in H . Then in fact $R(H \cup H') \leq r + s - b$.*

Proof. Let a be an edge in H' with exactly one endpoint, call it w , in H . We show that the addition of this edge to H causes a drop of 1 in revenue. If w is a constraint-vertex, then this follows because w already had degree at least τ in H , so a becomes a new excess edge in H , creating a new debit. So suppose w is a variable-vertex. If w had degree at least 2 in H then a is again excess and creates a new debit. If w had degree 1 in H then the addition of a changes w from a leaf variable to an interior variable, removing 1 credit from H . Finally, if w was isolated in H then the addition of a turns it into a leaf variable, again removing 1 credit from H . Repeating this argument for all b boundary edges completes the proof. \square

6.4 The key lemma

Lemma 6.14. *Let $y = y(x)$ be a polynomial expression of degree d . Assume $2d \leq \zeta \cdot \text{SMALL}$ and that $\tilde{\mathbf{E}}[y \cdot p(x)] = 0$ for all polynomials p of degree strictly less than d . Let H be a small τ -subgraph⁺ with $\text{vbls}(H) \supseteq \text{vbls}(y)$ and $R(H) \leq r$, where we assume $r + d \leq \zeta \cdot \text{SMALL}$. Finally, suppose T is a monomial index with $|T| = d$ such that*

$$\tilde{\mathbf{E}}[y \cdot x^T] \neq 0.$$

Then there exists a small τ -subgraph⁺ $H_{\text{new}} \supseteq H$ with $\text{vbls}(H_{\text{new}}) \supseteq \text{vbls}(y) \cup T$ and $R(H_{\text{new}}) \leq r$. (In writing $\text{vbls}(y) \cup T$, we are using the abuse described in Notation 6.3.)

Proof. Let us define

$$T_{\text{new}} = T \setminus \text{vbls}(H), \quad T_{\text{old}} = T \cap \text{vbls}(H), \quad B = \text{cl}(\text{vbls}(H) \cup T), \quad H_{\text{new}} = H \cup B.$$

First, we show that the τ -subgraph⁺ H_{new} is small; it follows that the τ -subgraph B is also small.

Claim 6.15. *H_{new} is small.*

Proof. Write $\text{cl}(\text{vbls}(H) \cup T) = H'_1 \cup \dots \cup H'_t$ for small $(\text{vbls}(H) \cup T)$ -closed τ -subgraphs H'_i . Let $H'_{<j} := H'_1 \cup \dots \cup H'_{j-1}$, and let s_j denote the number of leaves of H_j that are not in $H \cup H'_{<j}$. Then it is easy to see that $\sum_{j=1}^t s_j \leq d$. Now, iteratively apply Lemma 6.12 to $H \cup H'_1$, $(H \cup H'_1) \cup H'_2$, $((H \cup H'_1) \cup H'_2) \cup H'_3$, \dots to prove the claim. \square

Next, observe that we have $\text{vbls}(H_{\text{new}}) \supseteq \text{vbls}(H) \supseteq \text{vbls}(y)$; therefore to prove the lemma, it suffices to show that $\text{vbls}(H_{\text{new}}) \supseteq T_{\text{new}}$ and that $R(H_{\text{new}}) \leq R(H)$.

For the first of these, given an $(i, c) \in T$ we write $\bar{x}_i = 1_c(x_i) - \tilde{\mathbf{E}}[1_c(x_i)]$ and $\bar{x}^T = \prod_{i \in T} \bar{x}_i$. Observe that $\bar{x}^T - x^T$ is a polynomial of degree strictly less than d ; thus $\tilde{\mathbf{E}}[y \cdot (\bar{x}^T - x^T)] = 0$ and so $\tilde{\mathbf{E}}[y \cdot \bar{x}^T] \neq 0$. Now using Theorem 5.16 and $B \supseteq \text{cl}(\text{vbls}(H) \cup T) \supseteq \text{cl}(\text{vbls}(y \cdot \bar{x}^T))$, we conclude

$$\mathbf{E}_B[y \cdot \bar{x}^T] \neq 0. \tag{10}$$

In light of this, we claim that every variable $j \in T_{\text{new}}$ must appear as a vertex in B (and hence in $\text{vbls}(H_{\text{new}})$, as needed). For if $j \notin \text{vbls}(B)$, then \bar{x}_j is independent of all other random variables \mathbf{x}_i under η_B , and so

$$\begin{aligned} \mathbf{E}_B[y \cdot \bar{x}^T] &= \mathbf{E}_B[\bar{x}_j] \cdot \mathbf{E}_B[y \cdot \bar{x}^{T_{\text{old}}} \cdot \bar{x}^{T_{\text{new}} \setminus \{j\}}] && \text{(using } j \notin \text{vbls}(H) \supseteq \text{vbls}(y)) \\ &= \tilde{\mathbf{E}}[\bar{x}_j] \cdot \mathbf{E}_B[y \cdot \bar{x}^{T_{\text{old}}} \cdot \bar{x}^{T_{\text{new}} \setminus \{j\}}] = 0 && \text{(using } B \supseteq \text{cl}(T) \supseteq \text{cl}(\{j\}) \text{ and } \tilde{\mathbf{E}}[\bar{x}_j] = 0) \end{aligned}$$

in contradiction to (10).

It remains to show that $R(H_{\text{new}}) \leq R(H)$, which we will do using Lemma 6.13 (with $H' = B$, and $s = |T_{\text{new}}|$, recalling that all of B 's leaves are in $\text{vbls}(H) \cup T$). We must show that the number of ‘‘boundary edges’’ — i.e., edges in B that have exactly one endpoint in H — is at least $|T_{\text{new}}|$. Supposing otherwise, the set

$$V = \{\text{variable-vertices } v \in B : v \text{ is incident on a boundary edge}\} \cup T_{\text{old}}$$

would satisfy $|V| < |T_{\text{new}}| + |T_{\text{old}}| = |T| \leq d$. We will show that this contradicts (10).

Claim 6.16. *The deletion of variable-vertices V from B disconnects all variables in T from all variables in $\text{vbls}(H)$ within B . (Note that when a variable does not even appear in a subgraph, it is trivially disconnected from all other variables.)*

Proof. It suffices to show that deleting V disconnects T_{new} from $\text{vbls}(H)$ within B , as the vertices of T_{old} are already in V . Suppose $j \in T_{\text{new}}$ is connected to some variable $i \in \text{vbls}(H)$ by a path within B . Since $j \notin \text{vbls}(H)$, there must be some edge in this path that has exactly one endpoint in H . This edge is a boundary edge, and hence the variable-vertex incident on it is in V . Thus we have indeed established that *every* path within B from a variable in T_{new} to a variable in $\text{vbls}(H)$ must pass through a variable in V . \square

Recall that the proof is complete once we show that $|V| < d$ contradicts (10). Now

$$\begin{aligned}\mathbf{E}_B[\mathbf{y} \cdot \bar{\mathbf{x}}^T] &= \mathbf{E}_B\left[\mathbf{y} \cdot \bar{\mathbf{x}}^T \cdot \sum_{\vec{c} \in \Omega^V} \mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]\right] \\ &= \sum_{\vec{c} \in \Omega^V} \mathbf{E}_B\left[\mathbf{y} \cdot \bar{\mathbf{x}}^T \cdot \mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]\right].\end{aligned}\tag{11}$$

We claim that every summand above equals 0. The reason is that for each summand \vec{c} , either $\mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]$ is always 0 under η_B (establishing the claim), or else we may condition on the event, yielding

$$\mathbf{E}_B[\mathbf{y} \cdot \bar{\mathbf{x}}^T \cdot \mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]] = \Pr[\mathbf{x}_i = c_i \ \forall i \in V] \cdot \mathbf{E}_B[\mathbf{y} \cdot \bar{\mathbf{x}}^T \mid \mathbf{x}_i = c_i \ \forall i \in V].$$

By Claim 6.16 and the definition of the planted distribution η_B (and $\text{vbls}(y) \subseteq \text{vbls}(H)$), we have that \mathbf{y} and $\bar{\mathbf{x}}^T$ are conditionally independent under η_B , conditioned on all $(\mathbf{x}_i : i \in V)$. Therefore

$$\mathbf{E}_B[\mathbf{y} \cdot \bar{\mathbf{x}}^T \mid \mathbf{x}_i = c_i \ \forall i \in V] = \mathbf{E}_B[\mathbf{y} \mid \mathbf{x}_i = c_i \ \forall i \in V] \cdot \mathbf{E}_B[\bar{\mathbf{x}}^T \mid \mathbf{x}_i = c_i \ \forall i \in V].$$

Combining the previous two equations yields

$$\mathbf{E}_B[\mathbf{y} \cdot \bar{\mathbf{x}}^T \cdot \mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]] = \mathbf{E}_B[\mathbf{y} \cdot \mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]] \cdot \mathbf{E}_B[\bar{\mathbf{x}}^T \mid \mathbf{x}_i = c_i \ \forall i \in V].$$

Finally, using $|V| < d$ we will show that the first factor above is 0 (thereby establishing the claim that every term in (11) is 0, in contradiction to (10)). To see this, we have

$$\mathbf{E}_B[\mathbf{y} \cdot \mathbf{1}[\mathbf{x}_i = c_i \ \forall i \in V]] = \tilde{\mathbf{E}}\left[\mathbf{y} \cdot \prod_{i \in V} 1_{c_i}(x_i)\right]$$

because $\text{cl}(\text{vbls}(y) \cup V) \subseteq \text{cl}(\text{vbls}(H) \cup \text{vbls}(B)) \subseteq B$, where we used Theorem 5.9. But this pseudoexpectation is indeed 0 by the lemma's assumption, because $\prod_{i \in V} 1_{c_i}(x_i)$ is a polynomial expression of degree at most $|V| < d$. \square

6.5 Gram–Schmidt details

We wish to show that Gram–Schmidt succeeds through substage (S, T) for all $(S, T) \in \mathcal{M}_2^{\leq D}$. We will do this by induction along the order \preceq_2 . The key to showing that no positive definiteness problem is encountered at stage (S, T) will be the existence of a *witness*:

Definition 6.17. A *witness* for substage $(S, T) \in \mathcal{M}_2^{\leq D}$ is defined to be a small τ -subgraph⁺ $H_{S,T}$ with $\text{vbls}(H_{S,T}) \supseteq \text{vbls}(y_{S,T})$ and $R(H_{S,T}) \leq 2D$.

Remark 6.18. For any substage of the form (S, \emptyset) , we may always take as a witness the τ -subgraph⁺ consisting of all variables in S as isolated vertices.

As the below proposition shows, witnesses are useful for showing that one kind of positive definiteness problem does not occur. (They will also assist in showing the other kind does not occur.)

Proposition 6.19. *The existence of a witness $H_{S,T}$ for substage (S, T) implies $\tilde{\mathbf{E}}[y_{S,T}^2] \geq 0$.*

Proof. By Lemma 6.12, we have that $\bar{H} := H_{S,T} \cup \text{cl}(\text{vbls}(y_{S,T}))$ is small. Thus $\tilde{\mathbf{E}}[y_{S,T}^2] = \mathbf{E}_{\bar{H}}[y_{S,T}^2] \geq 0$, using Theorem 5.16. \square

We now come to our main technical theorem:

Theorem 6.20. *Let $(S, T) \in \mathcal{M}_2^{\leq D}$. Then:*

(i) *Given any witness $H_{S, \emptyset}$ for substage (S, \emptyset) , there is a witness $H_{S, T}$ for substage (S, T) satisfying $H_{S, T} \supseteq H_{S, \emptyset}$.*

(ii) *The Gram–Schmidt process succeeds through substage (S, T) .*

Proof. The proof will be by (strong) induction on (S, T) along \preceq_2 . Observe that in proving part (ii) of the theorem, by induction we only need to show that no positive definiteness problem occurs at substage (S, T) . Further, if we can inductively establish part (i) of the theorem, then Remark 6.18 and Proposition 6.19 imply that $\tilde{\mathbf{E}}[y_{S, T}^2] \geq 0$. Thus to also establish part (ii), it would only remain to prove that no “pseudovariance zero problem” problem occurs. Also, observe that the pseudovariance zero problem can never occur when $T = \emptyset$. Thus for substages (S, \emptyset) , we only need to establish part (i) of the theorem statement. But part (i) is trivial for (S, \emptyset) substages. Thus all substages of the form (S, \emptyset) are taken care of, including the base case of the induction (namely substage $(\{(i_0, c_0)\}, \emptyset)$, where $\{(i_0, c_0)\}$ is the first singleton in the order \preceq).

Thus it remains to establish, for a particular substage (S, T) with $T \neq \emptyset$, that part (i) of the theorem statement holds, and also that no pseudovariance zero problem occurs. Given any witness $H_{S, \emptyset}$ for substage (S, \emptyset) , by induction we may obtain a witness $H_{S, \text{pr}(T)} \supseteq H_{S, \emptyset}$ for substage $(S, \text{pr}(T))$. We now distinguish two cases.

Case 1: $\tilde{\mathbf{E}}[y_{S, \text{pr}(T)} \cdot z_T] = 0$. In this case, $y_{S, T} = y_{S, \text{pr}(T)}$ and therefore $\tilde{\mathbf{E}}[y_{S, T} \cdot z_T] = 0$. Thus certainly no pseudovariance zero problem occurs, and also we can establish part (i) of the theorem statement simply by taking $H_{S, T} = H_{S, \text{pr}(T)}$. Thus the inductive step is completed in this case.

Case 2: $\tilde{\mathbf{E}}[y_{S, \text{pr}(T)} \cdot z_T] \neq 0$. This is where the main work in the proof occurs. First, we will show in this case that $T \in \text{PvZ}$ is impossible, and hence the pseudovariance zero problem cannot have occurred. We can then complete the induction by finding a witness $H_{S, T} \supseteq H_{S, \text{pr}(T)}$ for substage (S, T) .

First, suppose for contradiction that $T \in \text{PvZ}$. We have that $y_{S, \text{pr}(T)} = x^S - q(x)$ for some $q(x)$ supported on monomials $x^{T'}$ with $T' \preceq \text{pr}(T) \prec T$. By Proposition 6.8 and induction, z_T is orthogonal to all such polynomials. Thus we deduce

$$0 \neq \tilde{\mathbf{E}}[y_{S, \text{pr}(T)} \cdot z_T] = \tilde{\mathbf{E}}[x^S \cdot z_T] = \tilde{\mathbf{E}}[x^S \cdot y_{T, \text{pr}(T)}], \quad (12)$$

the last equality because $T \in \text{PvZ}$ and hence $z_T = y_T = y_{T, \text{pr}(T)}$. By induction (and using Remark 6.18), we have a witness $H_{T, \text{pr}(T)}$ for $y_{T, \text{pr}(T)}$. By Lemma 6.12 (using $2D + |S| \leq 3D \leq \zeta \cdot \text{SMALL}$) we have that $\overline{H} := H_{T, \text{pr}(T)} \cup \text{cl}(S)$ is small. (In writing $\text{cl}(S)$ we used the abuse from Notation 6.3.). Now $\text{vbls}(\overline{H}) \supseteq \text{vbls}(H_{T, \text{pr}(T)}) \cup S \supseteq \text{vbls}(x^S \cdot y_{T, \text{pr}(T)})$, so by Theorem 5.16 we have

$$\tilde{\mathbf{E}}[x^S \cdot y_{T, \text{pr}(T)}] = \frac{\mathbf{E}}{\overline{H}}[x^S \cdot \mathbf{y}_{T, \text{pr}(T)}], \quad \text{and also} \quad \frac{\mathbf{E}}{\overline{H}}[\mathbf{y}_{T, \text{pr}(T)}^2] = \tilde{\mathbf{E}}[y_{T, \text{pr}(T)}^2] = \tilde{\mathbf{E}}[z_T^2] = 0,$$

the last equality because we’re assuming $T \in \text{PvZ}$. But the second identity above shows that $\mathbf{y}_{T, \text{pr}(T)}^2$ is identically 0 under $\eta_{\overline{H}}$, meaning the first expression above must be 0. This contradicts (12).

Having ruled out the pseudovariance zero problem, we can complete the induction by finding a witness $H_{S, T} \supseteq H_{S, \text{pr}(T)}$ for substage (S, T) . By Proposition 6.8 we have that $z_T = c \cdot x^T - p(x)$

for some constant $c > 0$ and some polynomial $p(x)$ supported on monomials $x^{T'}$ with $T' \preceq \text{pr}(T)$. Furthermore, $y_{S,\text{pr}(T)}$ is orthogonal to $p(x)$ under $\tilde{\mathbf{E}}[\cdot]$. Thus, since we are in Case 2, we may deduce that

$$\tilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot x^T] \neq 0. \quad (13)$$

We may now apply Lemma 6.14 (with $y = y_{S,\text{pr}(T)}$, $H = H_{S,\text{pr}(T)}$, and $r = 2D$) to obtain a small τ -subgraph⁺ $H_{\text{new}} \supseteq H_{S,\text{pr}(T)}$ with $\text{vbls}(H_{\text{new}}) \supseteq \text{vbls}(y_{S,\text{pr}(T)}) \cup T$ and $R(H_{\text{new}}) \leq 2D$. This H_{new} is *almost* able to serve as the witness for substage (S, T) . The only deficiency is that, although it contains all the variables in $y_{S,\text{pr}(T)}$ and x^T , it doesn't necessarily contain all the variables appearing in z_T — as it would need to in order to contain all variables in the new $y_{S,T} = y_{S,\text{pr}(T)} - \tilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T]z_T$. However, we can fix this by induction; we apply the induction hypothesis to substage $(T, \text{pr}(T))$, taking H_{new} as the “given witness $H_{T,\emptyset}$ ”. This produces a witness — call it H'_{new} — for substage $(T, \text{pr}(T))$ that satisfies $H'_{\text{new}} \supseteq H_{\text{new}}$. This witness H'_{new} now additionally contains all variables in $y_{T,\text{pr}(T)} = z_T$, and therefore it can now serve as the needed witness for substage (S, T) . \square

7 Wrapping things up by setting parameters

To prove our main result on weak refutation, Theorem 1.2, we simply need to combine Theorems 4.12 and Theorem 6.1. Together these give us a pseudoexpectation defined up to degree

$$D = \Omega(\gamma) \cdot \zeta \cdot \frac{n}{\Delta^{2/(\lambda-\zeta)}}, \quad \text{where } \gamma = \frac{\beta^{O(1/\lambda)}}{K \cdot 2^{O(K/\lambda)}}.$$

We need to decide how to best set parameters, which we do under the assumption that $\Delta \geq 10$.

We start with the special but interesting case when λ is thought of very large; specifically, $\lambda \geq \Omega(\log \Delta)$. This case arises, e.g., for high-arity K -SAT (where $\lambda = K - 2$) with clause density $2^{\Theta(K)}$. In this case, by choosing $\zeta = \frac{1}{2}\lambda$ and $\beta = e^{-O(K)}$ for our probability bound, we get $D = n/2^{O(K/\lambda)}$. Note that if $\lambda = \Theta(K)$, as it is in the case of K -SAT, then our SOS degree lower bound is linear in n with absolutely *no* dependence on $K = K(n)$ (all the way up to $K = \Omega(n)$)!

In the more general regime (e.g., when one thinks of K as “constant” and Δ as asymptotically large), a good choice for ζ is $\frac{1}{\log \Delta}$, which entails

$$D = \Omega(\gamma) \cdot \frac{n}{\Delta^{2/\lambda} \log \Delta}.$$

With this setting, Theorem 4.12 tells us that with high probability we get a pseudoexpectation satisfying Corollaries 5.18, 5.19. Thus we have established the following more precise version of Theorem 1.2:

Theorem 7.1. *Let P be a k -ary Boolean predicate and let $\mathcal{C}(P)$ be the minimum integer $3 \leq \tau \leq k$ for which P fails to support a τ -wise uniform distribution. Then if \mathcal{I} is a random instance of $\text{CSP}(P^\pm)$ with $m = \Delta n$ constraints ($\Delta \geq 10$), then except with probability at most β , degree- D SOS fails to (weakly) refute \mathcal{I} , where*

$$D = \frac{\beta^{O(1/\mathcal{C}(P))}}{k \cdot 2^{O(k/\mathcal{C}(P))}} \cdot \frac{n}{\Delta^{2/(\mathcal{C}(P)-2)} \log \Delta}.$$

The result also holds if P is a predicate over an alphabet of size $q > 2$ (with an appropriate notion of “literals”), with no change in parameters.

Proving our main result on δ -refutation, Theorem 1.1, requires just a little work. We now imagine that our instance comes from a random CSP(P^\pm) as in Theorem 1.1. As discussed at the end of Section 2.1, given t and taking $\tau = t + 1$, we have some t -wise uniform distribution μ on $\{\pm 1\}^k$ which is δ -close to being supported on P , where $\delta = \delta_P(t)$. We assume that all of the constraint distributions μ_f are now simply equal to μ , up to the appropriate negation pattern. Thus a draw from μ_f satisfies the constraint at f except with probability at most δ .

With the parameter settings chosen earlier, Theorem 4.12 tells us moreover that

$$\#\{\text{nonempty } \tau\text{-subgraphs } H \text{ with } |\text{cons}(H)| \leq 2 \cdot \text{SMALL} : I(H) \leq \tau - 1\} \leq \Delta n^{\frac{1+1/\log \Delta}{2}} = 2^{\frac{\log n}{2 \log \Delta}} \cdot \frac{m}{\sqrt{n}}. \quad (14)$$

Observe that this bound is *always* $o(m)$, and in the very typical case that $\Delta \geq n^{\Omega(1)}$, the bound is $O(\frac{m}{\sqrt{n}})$. Let us see what this bound means for the pseudodistribution.

Supposing (14) holds, let f be any constraint-vertex in G , let $S = N(f)$, and let H_f be the (small) τ -subgraph induced by the edges between f and S . Certainly $\text{cl}(S) \supseteq H_f$, but we may ask whether $\text{cl}(S)$ is strictly bigger than H_f . Suppose this is the case; i.e., there is some small S -closed $H \not\subseteq H_f$. Then $H' = H_f \cup H$ is a τ -subgraph satisfying $|\text{cons}(H')| \leq 2 \cdot \text{SMALL}$. Furthermore, the number of leaf variables in H' must be at least 1 (else H' is \emptyset -closed and hence empty by Fact 5.6) and strictly less than K (else $H \setminus H_f$ will be \emptyset -closed and hence empty). Finally, we claim $R(H') \leq \tau - 1$. This is because $R(H_f) = \tau$, the addition of H cannot add any new credits (since all its leaf variables are already in H_f), and in fact the addition of H must cause a drop of at least one in revenue since H must have at least one edge not in H_f . (This argument is similar to Lemma 6.13.) We conclude that whenever $\text{cl}(N(f)) \neq H_f$, there must exist a nonempty τ -subgraph H' with the following properties: (i) $|\text{cons}(H')| \leq 2 \cdot \text{SMALL}$; (ii) $I(H') \leq R(H') \leq \tau - 1$; (iii) H' has at least one leaf variable; (iv) all leaves of H' are adjacent to f .

But (14) bounds the number of τ -subgraphs with the first two properties above, and every τ -subgraph with the latter two properties uniquely determines f . Thus we conclude:

$$\#\{\text{constraints } f : \text{cl}(N(f)) \neq H_f\} \leq 2^{\frac{\log n}{2 \log \Delta}} \cdot \frac{m}{\sqrt{n}}.$$

Finally, when $\text{cl}(N(f)) = H_f$, observe that the planted distribution $\eta_{\text{cl}(N(f))}$ is just μ_f , and hence

$$\tilde{\mathbf{E}}[1[x \text{ satisfies } f]] = \Pr_{\mathbf{x} \sim \mu_f}[1[\mathbf{x} \text{ satisfies } f]] \geq 1 - \delta.$$

Combining the last two deductions yields

$$\tilde{\mathbf{E}}\left[\text{fraction of constraints satisfied}\right] \geq 1 - \delta - 2^{\frac{\log n}{2 \log \Delta}} \cdot \frac{1}{\sqrt{n}}.$$

In summary, we have proven the following more precise version of Theorem 1.1:

Theorem 7.2. *Let P be a k -ary Boolean predicate and let $1 < t \leq k$. Let \mathcal{I} be a random instance of CSP(P^\pm) with $m = \Delta n$ constraints. Then except with probability at most β , degree- D SOS fails to $(\delta_P(t) + \epsilon)$ -refute \mathcal{I} , where*

$$\epsilon = 2^{\frac{\log n}{2 \log \Delta}} \cdot \frac{1}{\sqrt{n}}, \quad D = \frac{\beta^{O(1/t)}}{k \cdot 2^{O(k/t)}} \cdot \frac{n}{\Delta^{2/(t-1)} \log \Delta}.$$

We remark that $\epsilon = o(1)$ always, and $\epsilon = O(\frac{1}{\sqrt{n}})$ whenever $\Delta = n^{\Omega(1)}$. Finally, the result also holds if P is a predicate over an alphabet of size $q > 2$ (with an appropriate notion of “literals”), with no change in parameters.

Remark 7.3. We should mention that in our δ -refutation result Theorem 7.2, our pseudoexpectation does *not* satisfy “solution value = $1 - \delta_0$ ” as a constraint for any $\delta_0 \leq \delta$; it merely has $\tilde{\mathbf{E}}[\text{solution value}] \geq 1 - \delta$. Achieving the (stronger) former condition is a direction for future work. By contrast, for our weak refutation result Theorem 1.2, the pseudoexpectation *does* satisfy all the constraints and hence also satisfies $\tilde{\mathbf{E}}[\text{solution value}] = 1$ as a constraint.

Acknowledgment

We would like to thank the Institute for Mathematical Sciences, National University of Singapore in 2016; a visit there was where some of the initial research for this work began.

References

- [AAM⁺11] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Inapproximability of densest κ -subgraph from average case hardness. 2011. [1](#)
- [AAT05] Mikhail Alekhnovich, Sanjeev Arora, and Iannis Tourlakis. Towards strong nonapproximability results in the Lovász-Schrijver hierarchy. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 294–303, 2005. [1.4](#)
- [ABR12] Benny Applebaum, Andrej Bogdanov, and Alon Rosen. A dichotomy for local small-bias generators. In Ronald Cramer, editor, *Theory of Cryptography*, volume 7194 of *Lecture Notes in Computer Science*, pages 600–617. Springer Berlin Heidelberg, 2012. [1.2](#), [1](#)
- [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](#), [1.2](#)
- [AGT12] Noga Alon, Iftah Gamzu, and Moshe Tennenholtz. Optimizing budget allocation among channels and influencers. In *Proceedings of the 21st International Conference on World Wide Web*, pages 381–388, 2012. [1.2](#)
- [AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC⁰. *SIAM Journal on Computing*, 36(4):845–888, 2006. [1.2](#)
- [AL16] Benny Applebaum and Shachar Lovett. Algebraic Attacks against Random Local Functions and Their Countermeasures. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 1087–1100, 2016. [1.2](#), [1](#), [1.3](#)
- [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. [1.2](#)
- [AM08] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. In *Proceedings of the 23rd IEEE Conference on Computational Complexity*, pages 249–258, 2008. [1.4](#)
- [AOW15] Sarah R. Allen, Ryan O’Donnell, and David Witmer. How to refute a random CSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 689–708, 2015. [\(document\)](#), [1](#), [1.5](#), [1.4](#), [1.5](#), [1.5](#)

- [App13] Benny Applebaum. Cryptographic hardness of random local functions—survey. In *10th Theory of Cryptography Conference*, 2013. [1.2](#)
- [AR01] Michael Alekhnovich and Alexander A. Razborov. Lower bounds for polynomial calculus: non-binomial case. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 190–199, 2001. [1.4](#)
- [BBaH⁺12] Boaz Barak, Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, Sum-of-Squares Proofs, and their Applications. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, pages 307–326, 2012. [1.4](#)
- [BCG⁺12] Aditya Bhaskara, Moses Charikar, Venkatesan Guruswami, Aravindan Vijayaraghavan, and Yuan Zhou. Polynomial integrality gaps for strong sdp relaxations of densest k -subgraph. In *Proceedings of the 23rd ACM-SIAM Symposium on Discrete Algorithms*, pages 388–405, 2012. [1](#)
- [BCK15] Boaz Barak, Siu On Chan, and Pravesh K. Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, 2015. [1.4](#), [3.1.1](#), [3.2](#), [3.2](#)
- [BCM⁺12] Aditya Bhaskara, Moses Charikar, Rajsekar Manokaran, and Aravindan Vijayaraghavan. On quadratic programming with a ratio objective. In *Proceedings of the 39th International Colloquium on Automata, Languages and Programming*, pages 109–120, 2012. [1](#)
- [BGMT12] Siavosh Benaïm, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(12):269–289, 2012. [1.4](#), [3.1](#), [3.1.1](#)
- [BJK05] Andrei Bulatov, Peter Jeavons, and Andrei Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34(3):720–742, 2005. [1](#)
- [BKP04] Punit Bhargava, Sriram C. Krishnan, and Rina Panigrahy. Efficient multicast on a terabit router. In *Proceedings of the 12th Annual IEEE Symposium on High Performance Interconnects*, pages 61–67, 2004. [1.2](#)
- [BKS13] Boaz Barak, Guy Kindler, and David Steurer. On the optimality of semidefinite relaxations for average-case and generalized constraint satisfaction. In *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 197–214, 2013. [1](#)
- [BM16] Boaz Barak and Ankur Moitra. Noisy Tensor Completion via the Sum-of-Squares Hierarchy. In *Proceedings of the 29th Annual Conference on Learning Theory*, pages 417–445, 2016. [3](#)
- [BOGH⁺03] Joshua Buresh-Oppenheim, Nicola Galesi, Shlomo Hoory, Avner Magen, and Toniann Pitassi. Rank bounds and integrality gaps for cutting planes procedures. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 318–327, 2003. [1.4](#)

- [BQ09] Andrej Bogdanov and Youming Qiao. On the security of Goldreich’s one-way function. In Irit Dinur, Klaus Jansen, Joseph Naor, and José Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques*, volume 5687 of *Lecture Notes in Computer Science*, pages 392–405. Springer Berlin Heidelberg, 2009. 1
- [Bri08] Patrick Briest. Uniform Budgets and the Envy-Free Pricing Problem. In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming*, pages 808–819. 2008. 1.2
- [BS] Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares. <http://sumofsquares.org/public/index.html>. 2.3
- [BS01] Eli Ben-Sasson. *Expansion in Proof Complexity*. PhD thesis, Hebrew University, 2001. 1.4
- [BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014. 1.4
- [BSB02] Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 9(3), 2002. 1, 1.5
- [BSI99] Eli Ben-Sasson and Russell Impagliazzo. Random CNF’s are hard for the polynomial calculus. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 415–421, 1999. 1.4
- [BSW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *J. ACM*, 48(2):149–169, 2001. 1.4
- [CD09] Nadia Creignou and Hervé Daudé. The SAT-UNSAT transition for random constraint satisfaction problems. *Discrete Math.*, 309(8):2085–2099, 2009. 1.1
- [CLP02] A Crisanti, L Leuzzi, and G Parisi. The 3-sat problem with large number of clauses in the ∞ -replica symmetry breaking scheme. *Journal of Physics A: Mathematical and General*, 35(3):481, 2002. 1
- [CMVZ12] Julia Chuzhoy, Yury Makarychev, Aravindan Vijayaraghavan, and Yuan Zhou. Approximation algorithms and hardness of the k -route cut problem. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 780–799, 2012. 1
- [CS88] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *J. Assoc. Comput. Mach.*, 35(4):759–768, 1988. 1.4
- [Dan15] Amit Daniely. Complexity Theoretic Limitations on Learning Halfspaces. *CoRR*, abs/1505.05800, 2015. 1.2, 1, 2
- [DFHS06] Erik D. Demaine, Uriel Feige, Mohammad Taghi Hajiaghayi, and Mohammad R. Salavatipour. Combination can be hard: Approximability of the unique coverage problem. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 162–171, 2006. 1.2

- [DKMPG08] Josep Diaz, Lefteris Kirousis, Dieter Mitsche, and Xavier Perez-Gimenez. A new upper bound for 3-SAT. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 2, pages 163–174, 2008. 1.1
- [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*, pages 145–153, 2013. 1.2
- [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. ACM, 2014. 1, 1.2, 1, 2
- [DS14] Amit Daniely and Shai Shalev-Shwartz. Complexity theoretic limitations on learning DNF’s. Technical Report 1404.3378, arXiv, 2014. 1.2, 1, 2
- [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*, pages 59–68, 2015. 1.1
- [Fei02] Uriel Feige. Relations Between Average Case Complexity and Approximation Complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 534–543, 2002. 1, 1.2, 1.2, 1, 3
- [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. 1.4
- [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, 2006. 4
- [FPV15] Vitaly Feldman, Will Perkins, and Santosh Vempala. On the Complexity of Random Satisfiability Problems with Planted Solutions. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 77–86, 2015. 1.4
- [Gab16] Oliver Gableske. dimetheus. In *Proceedings of SAT Competition 2016: Solver and Benchmark Descriptions*, pages 37–38, 2016. 1.1
- [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. 1.4
- [GL04] Andreas Goerdt and André Lanka. An approximation hardness result for bipartite Clique. *Electronic Colloquium on Computational Complexity (ECCC)*, (048), 2004. 1
- [Gol00] Oded Goldreich. Candidate One-Way Functions Based on Expander Graphs. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 7, 2000. 1.2
- [Gri01] Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *Computational Complexity*, 10(2):139–154, 2001. 1.4
- [Hua13] Sangxia Huang. Approximation resistance on satisfiable instances for predicates with few accepting inputs (extended abstract). In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 457–466, 2013. 1

- [Hua14] Sangxia Huang. Approximation Resistance on Satisfiable Instances for Predicates with Few Accepting Inputs. *Theory of Computing*, 10(14):359–388, 2014. [1](#)
- [IKOS08] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Cryptography with constant computational overhead. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 433–442, 2008. [2](#)
- [KM16] Subhash Khot and Dana Moshkovitz. Candidate hard unique game. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 63–76, 2016. [4](#)
- [KOTZ14] Manuel Kauers, Ryan O’Donnell, Li-Yang Tan, and Yuan Zhou. Hypercontractive inequalities via SOS, and the Frankl-Rödl graph. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1644–1658, 2014. [1.4](#)
- [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. [1.4](#)
- [LRS15] James R Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. ACM, 2015. [\(document\)](#), [1.4](#), [1.5](#)
- [MPRT16] Raffaele Marino, Giorgio Parisi, and Federico Ricci-Tersenghi. The backtracking survey propagation algorithm for solving random K-SAT problems. *Nature Communications*, 7(12996), 2016. [1.1](#)
- [MST03] Elchanan Mossel, Amir Shpilka, and Luca Trevisan. On ϵ -biased generators in NC^0 . In *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, pages 136–145, 2003. [1.2](#)
- [MW16] Ryuhei Mori and David Witmer. Lower bounds for CSP refutation by SDP hierarchies. In *RANDOM ’16*, 2016. [1.4](#)
- [OW14] Ryan O’Donnell and David Witmer. Goldreich’s PRG: Evidence for near-optimal polynomial stretch. In *Proceedings of the 29th Annual Conference on Computational Complexity*, pages 1–12, 2014. [1](#), [1.4](#)
- [OWWZ14] Ryan O’Donnell, John Wright, Chenggang Wu, and Yuan Zhou. Hardness of robust graph isomorphism, Lasserre gaps, and asymmetry of random graphs. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1659–1677, 2014. [1.2](#)
- [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, 2013. [1.4](#)
- [Rag08] Prasad Raghavendra. Optimal Algorithms and Inapproximability Results for Every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 245–254, 2008. [1](#)
- [RRS16] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random cps below the spectral threshold. *CoRR*, abs/1605.00058, 2016. [\(document\)](#), [1.5](#), [1.4](#)

- [RSW16] Ilya Razenshteyn, Zhao Song, and David P. Woodruff. Weighted low rank approximations with provable guarantees. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 250–263, 2016. [1](#), [4](#)
- [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 Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602, 2008. [1.4](#)
- [Tul09] Madhur Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 303–312, 2009. [1](#), [1.4](#)
- [TW13] Madhur Tulsiani and Pratik Worah. LS_+ lower bounds from pairwise independence. In *Proceedings of the 28th Annual Conference on Computational Complexity*, pages 121–132, 2013. [1.4](#)

A Proof that random graphs satisfy the Plausibility Assumption

Here we prove Theorem [4.12](#), which we restate for convenience:

Theorem 4.12 restated. *Let $\lambda = \tau - 2 \geq 1$. Fix $0 < \zeta \leq .99\lambda$, $0 < \beta < \frac{1}{2}$. Then except with probability at most β , when \mathbf{G} is a random instance with $m = \Delta n$ constraints, the Plausibility Assumption holds provided*

$$\text{SMALL} \leq \gamma \cdot \frac{n}{\Delta^{2/(\lambda-\zeta)}}, \quad (15)$$

where $\gamma = \frac{1}{K} \left(\frac{\beta^{1/\lambda}}{2^{K/\lambda}} \right)^{O(1)}$. Moreover, assuming $\zeta < 1$, except with probability at most β we have

$$\#\{\text{nonempty } \tau\text{-subgraphs } H \text{ with } \text{cons}(H) \leq 2 \cdot \text{SMALL} : I(H) \leq \tau - 1\} \leq \Delta n^{\frac{1+\zeta}{2}}. \quad (16)$$

Proof. A remark before we begin: the expression in [\(15\)](#) was chosen precisely so that

$$c \leq 2 \cdot \text{SMALL} \implies 20^K \cdot \Delta \cdot \left(\frac{Kc}{n} \right)^{\frac{\lambda-\zeta}{2}} \leq \beta/50^K, \quad (17)$$

provided the $O(1)$ in the definition of γ is a sufficiently large universal constant.

The proof is a standard argument of the kind used to show that a random bipartite graph has good expansion. Fixing $I_0 \in \{0, \tau - 1\}$, $1 \leq c \leq 2 \cdot \text{SMALL}$, and $1 \leq v \leq Kc$, let us upper-bound

$$\mathbf{E}[\#\{\tau\text{-subgraphs with } c \text{ constraints, } v \text{ vertices, and income at most } I_0\}]. \quad (18)$$

There are $\binom{m}{c}$ choices for the constraints and $\binom{n}{v}$ choices for the variables. Then by using Lemma [4.11](#),

$$(18) \leq \binom{m}{c} \binom{n}{v} \mathbf{Pr}[\text{fixed set of } c \text{ constraints and } v \text{ variables gets at least } A \text{ edges}], \quad (19)$$

where $A := \frac{\tau-\zeta}{2} \cdot c + v - \frac{I_0}{2}$. In [\(19\)](#), we may imagine that a constraint's variables are chosen uniformly and independently (i.e., *without* conditioning on them being distinct), as this only increases the

probability in question. Now any fixed set of c constraints has at most Kc edges coming out it, so the probability that some integer $a > A$ of them will go into a fixed set of v variables is at most

$$\binom{Kc}{a} \cdot \left(\frac{v}{n}\right)^a \leq 2^{Kc} \cdot \left(\frac{v}{n}\right)^a \leq 2^{Kc} \cdot \left(\frac{v}{n}\right)^A.$$

Thus

$$\begin{aligned} (19) &\leq 2^{Kc} \binom{m}{c} \binom{n}{v} \left(\frac{v}{n}\right)^A \leq 2^{Kc} \left(\frac{em}{c}\right)^c \left(\frac{en}{v}\right)^v \left(\frac{v}{n}\right)^A = \left(e2^K e^{v/c} (v/c)\right)^c \cdot \Delta^c \cdot \left(\frac{v}{n}\right)^{\frac{\lambda-\zeta}{2} \cdot c - I_0/2} \\ &\leq (20^K)^c \cdot \Delta^c \cdot \left(\frac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2} \cdot c - I_0/2}, \end{aligned} \quad (20)$$

where the equality used the definition of A and the subsequent inequality used $v \leq Kc$.

We now split into two cases, depending on whether I_0 is 0 or $\tau - 1$. When $I_0 = 0$ we use

$$(18) \leq (20) = \left(20^K \cdot \Delta \cdot \left(\frac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2}}\right)^c \leq \left(\frac{\beta}{50^K}\right)^c,$$

using (17). Summing over the at most Kc possibilities for v gives

$$\mathbf{E}[\#\{\tau\text{-subgraphs with } c \text{ constraints and income at most } 0\}] \leq Kc \left(\frac{\beta}{50^K}\right)^c.$$

Now summing this expression over all $1 \leq c \leq 2 \cdot \text{SMALL}$ we get

$$\mathbf{E}[\#\{\text{implausible } \tau\text{-subgraphs } H : |\text{cons}(H)| \leq 2 \cdot \text{SMALL}\}] \leq \sum_{c=1}^{\infty} Kc \left(\frac{\beta}{50^K}\right)^c \leq \beta.$$

Thus Markov's inequality implies that the Plausibility Assumption holds except with probability at most β .

The analysis for $I_0 = \tau - 1$ is similar. In this case, we use

$$(18) \leq (20) = \left(20^K \cdot \Delta \cdot \left(\frac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2}}\right)^{c-1} \cdot 20^K \cdot \Delta \cdot \left(\frac{n}{Kc}\right)^{\frac{1+\zeta}{2}} \leq \left(\frac{\beta}{50^K}\right)^{c-1} \cdot 20^K \cdot \Delta n^{\frac{1+\zeta}{2}},$$

again using (17). We again sum this over the at most Kc possibilities for v . We also only need to sum this over all $c \geq 2$, since if $\text{cons}(H) = 1$ then $I(H) = \tau - \zeta > \tau - 1$. We then obtain

$$\begin{aligned} \mathbf{E}[\#\{\text{nonempty small } \tau\text{-subgraphs } H \text{ with } |\text{cons}(H)| \leq 2 \cdot \text{SMALL} : I(H) \leq \tau - 1\}] \\ \leq \sum_{c=2}^{\infty} Kc \left(\frac{\beta}{50^K}\right)^{c-1} 20^K \cdot \Delta n^{\frac{1+\zeta}{2}} \leq \beta \cdot n^{\frac{1+\zeta}{2}}, \end{aligned}$$

and again Markov's inequality establishes that (16) holds except with probability at most β . \square