Refutation of random constraint satisfaction problems using the sum of squares proof system

David Witmer

CMU-CS-17-114
April 5, 2017

Computer Science Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213
dwitmer@cs.cmu.edu

Thesis Committee:
Anupam Gupta (co-chair)
Ryan O'Donnell (co-chair)
Alan Frieze
Boaz Barak (Harvard University)
Eli Ben-Sasson (Technion – Israel Institute of Technology)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This research was sponsored by the National Science Foundation under grant numbers CCF-1016799, CCF-1116594, CCF-1319743, CCF-1319811, CCF-1618679, CCF-1617790, and DGE-1252522; the Microsoft MSR-CMU Center for Computational Thinking; and the Presidential Fellowship in the School of Computer Science.

The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

Keywords: constraint satisfaction problems, average-case complexity, semidefinite programming, proof complexity

For the glory of God

Here I raise my Ebenezer;
Here by Thy great help I've come

–Robert Robinson

Abstract

Given a k-ary predicate P, a random instance of a constraint satisfaction problem with predicate P ($\mathrm{CSP}(P)$) is a set of m constraints on n variables. Each constraint is P applied to k random literals. Such an instance is unsatisfiable with high probability when $m \gg n$. Although this fact can be proven via a simple probabilistic argument, certifying that a given randomly chosen instance is unsatisfiable, a task called refutation, remains a challenge. Refutation, besides being a natural problem in its own right, has applications in cryptography, hardness of approximation, and learning theory. This thesis studies refutation using sum-of-squares (SOS) proof systems [GV01]. SOS is a sequence of increasingly powerful methods for proving polynomial inequalities parameterized by a value called the degree: the higher the degree, the more powerful the proof system. On the other hand, the amount of computation needed to find an SOS proof increases with degree: a degree-d proof can be found in time $n^{O(d)}$ [Sho87, Par00, Las00, Las01].

First, we consider refutation via constant-degree SOS proofs, which can be found in polynomial time. We show that the number of constraints required for constant-degree SOS refutation of a random instance of $\mathrm{CSP}(P)$ is determined by a complexity parameter $\mathcal{C}(P)$, the minimum t for which there is no t-wise uniform distribution over $\{0,1\}^k$ supported on satisfying assignments to P. With Allen and O'Donnell [AOW15], we proved that constant-degree SOS can refute a random instance of $\mathrm{CSP}(P)$ when $m = \widetilde{O}(n^{\mathcal{C}(P)/2})$. With Kothari, Mori, and O'Donnell [KMOW17], we showed a nearly matching lower bound: SOS requires superconstant degree to refute random instances of $\mathrm{CSP}(P)$ when $m = \widetilde{\Omega}(n^{\mathcal{C}(P)/2})$.

More generally, we consider the stronger notion of δ-refutation, certifying that at most a $1-\delta$ fraction of constraints can be simultaneously satisfied. We also consider SOS proofs with arbitrary, possibly superconstant, degree. In [AOW15], we proved that if every t-wise uniform distribution on $\{0,1\}^k$ is δ-far from every distribution supported on satisfying assignments to P, then constant-degree SOS can $(\delta - o(1))$-refute a random instance of $\mathrm{CSP}(P)$ with $m = \widetilde{O}(n^{t/2})$. For such P, this can be extended using a result of Raghavendra, Rao, and Schramm [RRS17] to obtain $(\delta - o(1))$-refutation of random instances of $\mathrm{CSP}(P)$ with $m = \Delta n$ constraints via degree-$\widetilde{O}(n/\Delta^{2/(t-2)})$ SOS. In [KMOW17], we proved that $(\delta + o(1))$-refutation of a random instance of $\mathrm{CSP}(P)$ with $m = \Delta n$ constraints requires SOS degree $\widetilde{\Omega}(n/\Delta^{2/(t-1)})$ when there exists a t-wise uniform distribution that is δ-close to being supported on satisfying assignments to P. These results establish a three-way trade-off among number of constraints, SOS degree, and refutation strength δ that is tight up to logarithmic factors.

Acknowledgments

I'm thankful to my advisors Anupam Gupta and Ryan O'Donnell. Anupam taught me about the nuts and bolts of being a researcher: how to read papers, use my time effectively to make progress, talk to people at conferences, and everything else like that. Ryan taught me the importance of having the right way of thinking about a problem, to not be satisfied until I fully understand an idea, and through your own ideas gave me a taste for elegance and simplicity. Ryan and Anupam taught me so much about how to write, speak, and teach. I appreciate the confidence they have had in me, even from the start of grad school. I knew nothing about theoretical computer science research when I arrived at CMU, yet both were willing to advise me. During the Computer Science Department (CSD) Open House, I remember Anupam asking me if I was considering any other schools. Somewhat embarrassed, I answered that I wasn't really considering any other schools because I had only been accepted to two schools and CMU was the best of the two. Anupam didn't miss a beat and replied that CMU had made the right choice in accepting me. Even though I've asked countless dumb questions and making countless obviously wrong statements in meetings, Ryan, too, has shown this same confidence in me. Their confidence was a constant encouragement when I frequently doubted my ability to do research. More than that, I appreciate their consistent patience and kindness. They both care deeply about their students.

I also thank the other members of my thesis committee Alan Frieze, Boaz Barak, and Eli Ben-Sasson. I wish I had had more time to investigate all of the fascinating directions you all suggested.

Apart from my advisors, I worked most closely with Sarah Allen, Pravesh Kothari, Ryuhei Mori, Aravindan Vijayaraghavan, and John Wright. I thank them for everything they taught me, for their patience, and for their contributions to this thesis. I also thank the rest of my coauthors – Guy Kindler, Kunal Talwar, Ankur Moitra, Prasad Raghavendra, Oded Regev, David Steurer, Luca Trevisan, Sarah Loos, Peter Steenkiste, and André Platzer – for their contributions to work not included in this thesis.

I thank Yury Makarychev for inviting me to spend the summer of 2014 at TTIC. It was a pleasure working with Yury, Julia Chuzhoy, and Euiwoong Lee for that summer.

I'm grateful to Uri Feige for inviting me to spend the summer of 2015 at the Weizmann Institute. I thoroughly enjoyed working with Uri, Roee David, and Dan Vilenchik. I'm also thankful for friendships with Ilan Komargodski and Alon Ivtsan. The kehila in Rehovot welcomed me and treated me like family. They showed me what Jesus meant when He said"By this all people will know that you are my disciples, if you have love for one another" (John 13:35), and they exemplified a love for God and His Word that profoundly shaped me.

The CSD administrative staff are incredible. Nancy Conway and Chase Klingensmith helped me with everything from passing back exams to getting reimbursed to making room reservations. It was great working with Nicole Stenger, Jenn Landefeld, and Martha Clarke on Open House. From the first few weeks of grad school when I was worried about finding an advisor to taking care of everything needed to graduate over these last few weeks, Deb Cavlovich has helped me over the last six years with every administrative task that I haven't known how to handle and been consistently supportive and encouraging. Visiting their offices or spending time with them at department events has always been a pleasure.

I also want to thank my friends in CSD. Alex Beutel, Nico Feltman, David Naylor,

Richard Wang, and I started the program together almost six years ago. I'll miss hanging out with them. Guru Guruganesh and Laxman Dhulipala have become close friends as well; I've appreciated all of our conversations about life outside of computer science. Arbob Ahmad, Mehdi Samadi, Nika Haghtalab, and Michael Kuchnik have been great officemates. Outside of CSD, Aleksandr Kazachkov has become a great friend.

I'm grateful for my friends in Pittsburgh outside of CMU. Tom O'Toole and Regina Triplett were my family in Pittsburgh. I'm thankful to have also lived with Gabriel Crawford, Ken Nakajima, Jeffery Song, and Peter Story. Nathan Van Patter has also been a great friend.

Though we no longer live in the same place, my friends from college have continued to be there for me throughout grad school. Ian Sugel, Mark Van de Loo, and Rafael Oliveira have become family. I'm also grateful for my brothers Daniel Kim, Ken Warnock, and Gabriel Ha. Our conversations have been a great source of refreshment and encouragement, and I'm so thankful for their prayers. I'm also grateful to Mitch Westwood, Ken Van Tilburg, Colin Hom, Drew Musacchio, Toomas Sepp, Eric Lubell, and Shreya Dave.

I can't begin to say how thankful I am for my family. My parents Karl and Debbie Witmer have loved and supported me through everything. The older I get, the more I appreciate them and their example of quiet faithfulness, selflessness, humility, and love. I'm also thankful for my sister, Karen Coley, and her husband Will. They have also been endlessly supportive. Having them nearby in both college and grad school has been a huge blessing. My niece and nephew Lydia and Nate have also been a joy to spend time with. I'm also thankful for my brother Joel Witmer and his wife Theresa. Visiting them over the last few years has been a needed refreshing and restful break from work. My grandmother, Thelma Deutsch, has kept me grounded and has constantly reminded me to trust God. I'm also thankful that she lets me visit her. Every conversation and visit with her has been a blessing.

Finally, I have to thank Jesus, my Creator, my Savior, and my God. Every good idea, every relationship, every moment of these last six years has been an instance of His goodness, love, and provision. In the words of Psalm 73:25-26,

> Whom have I in heaven but you?
> And there is nothing on earth that I desire besides you.
> My flesh and my heart may fail,
> but God is the strength of my heart and my portion forever.

Additional acknowledgements

The work described in Chapter 3 was done in collaboration with Sarah R. Allen (Computer Science Department, Carnegie Mellon University; srallen@cs.cmu.edu) and Ryan O'Donnell (Computer Science Department, Carnegie Mellon University; odonnell@cs.cmu.edu). It was originally published as [AOW15] (copyright © 2015 IEEE), and most of Chapter 3 is reprinted, with permission, from [AOW15]. Sarah R. Allen and Ryan O'Donnell were also supported by National Science Foundation grants CCF-0747250 and CCF-1116594. Sarah R. Allen was also supported by the National Science Foundation Graduate Research Fellowship Program under grant DGE-1252522. Some of this work was performed while Ryan O'Donnell was at the Boğaziçi University Computer Engineering Department, supported by Marie Curie International Incoming Fellowship project number 626373. We would like to thank Amin Coja–Oghlan for help

with the literature, and Boaz Barak and Ankur Moitra for permission to reprint the proof of the strong k-XOR refutation result.

The work described in Chapter 4 was done in collaboration with Pravesh K. Kothari (Princeton University and the Institute for Advanced Study; kothari@cs.princeton.edu), Ryuhei Mori (Department of Mathematical and Computing Sciences, Tokyo Institute of Technology; mori@is.titech.ac.jp), and Ryan O'Donnell. It was originally published as [KMOW17] (copyright © 2017 ACM), and most of Chapter 4 is reprinted, with permission, from [KMOW17]. Ryan O'Donnell was also supported by National Science Foundation grant CCF-1618679. Pravesh K. Kothari and Ryan O'Donnell thank the Institute for Mathematical Sciences, National University of Singapore for supporting a 2016 visit to the Institute; this work began during their visit.

Contents

Chapter 1

Introduction

1.1 Random CSPs and refutation

Constraint satisfaction problems (CSPs) are fundamental objects capturing a myriad of important computational problems, from scheduling to resource allocation to planning. In addition to their practical importance, the study of CSPs has played a significant role in the development of theoretical computer science. This is in large part due to their tantalizing simplicity: A CSP instance is a set of m local constraints on a set of n variables and the objective is to find an assignment to the variables satisfying all constraints. Locally, any single constraint is easy to satisfy, but, globally, deciding if a CSP has a solution is the canonical NP-hard problem. Understanding the computational hardness of CSPs requires understanding how this global difficulty arises from local constraints.

Despite making great strides over the last several decades in our knowledge of the worst-case relative hardness of solving (e.g., [Sch78, BJK05]) and approximating (e.g., [Hås01, Rag08]) CSPs, developing a more concrete understanding of what makes a CSP hard remains a challenge. For example, being able to generate hard instances is often useful in proving hardness results and in cryptographic applications.

Randomly chosen CSPs are natural candidate hard instances studied in cryptography [ABW10], proof complexity [BB02], hardness of approximation [Fei02], learning theory [DLSS14], and SAT-solving [SAT]. Let $P : \{0,1\}^k \to \{0,1\}$ be a k-ary predicate. An instance $\mathcal{I}$ of CSP(P) on n variables $x_1, x_2, \ldots, x_n$ consists of a set of m constraints of the form $P(x_{S_1} \oplus c_1, x_{S_2} \oplus c_2, \ldots x_{S_k} \oplus c_k) = 1$ with $S \subseteq [n]^k$ and $c \in \{0,1\}^k$. We let $\mathrm{Opt}(\mathcal{I})$ denote the maximum fraction of constraints that can be simultaneously. To construct a random instance of CSP(P), m constraints are chosen uniformly at random.

The satisfiability of a random CSP is governed by its density $\Delta = m/n$. When $\Delta \gg 1$, straightforward application of the Chernoff and Union Bounds shows that instances are unsatisfiable with very high probability. For small enough $\Delta = O(1)$, it is also easy to show that an instance is satisfiable with high probability. In the case of k-SAT with $k \geq 3$, it is widely conjectured that there exists a critical density Δ_c below which instances of k-SAT are satisfiable with high probability and above which they are unsatisfiable with high probability. Friedgut showed that there exists a sequence $\Delta_c(n)$ of such thresholds [Fri99]. For sufficiently large k, Ding, Sly, and Sun [DSS15] recently proved that the sequence $\Delta_c(n)$ converges to some Δ_c, but proving this for all k remains an open problem. Creignou and Daudé showed that CSP(P) has a sharp satisfiability threshold sequence if and only if P's satisfying assignments are not a subset of those of a dictator function or its negation and are not a subset of those of a 2-XOR function or its negation [CD09]. Again, it

has not yet been proven that these threshold sequences converge to a constant as n increases.

In each of these two phases, there is a natural algorithmic task. In the satisfiable regime, this task is to find a satisfying assignment. For SAT instances, there have been some practical successes in solving instances at densities approaching Δ_c [Gab16, MPRT16]. In the unsatisfiable regime, the natural algorithmic task is to *refute* the instance, meaning to certify that it is unsatisfiable. For refutation we know that 3-SAT instances are unsatisfiable for $\Delta > 4.49$ with high probability [DKMPG08], but we do not know of any efficient algorithms that can refute for $\Delta < n^{0.49}$ in theory or in practice. In both regimes, random instances are used by practitioners as benchmarks for SAT solvers [BBHJ13, BDHJ14]. Despite the fact that constraints are chosen independently, the uniform distribution over optimal assignments to a random CSP is a complex object with nontrivial global correlations that has been heavily studied, especially in the statistical physics literature (e.g., [CLP02, KMRT$^+$07]).

In this thesis, we focus on the problem of refuting random CSPs. Besides being a natural problem in its own right, refutation has arisen in such seemingly unrelated areas as hardness of approximation [Fei02], cryptography [ABW10], and learning theory [DLSS14]. We also consider stronger variants of the basic refutation problem. In the *δ-refutation* problem, the goal is to certify that $\mathrm{Opt}(\mathcal{I}) \leq 1-\delta$. Simple probabilistic arguments show that when $m \gg n$, $\mathrm{Opt}(\mathcal{I}) = \mathbf{E}[P] + o(1)$ with high probability, where $\mathbf{E}[P]$ is the expected value of P over the uniform distribution on $\{0,1\}^k$. We will refer to the standard refutation problem, in which $\delta = \frac{1}{m}$, as *weak refutation*. *Nondeterministic refutation* is showing that there exists a short certificate of unsatisfiability, though it may not be possible to actual find such a proof efficiently.

In the next section, we will describe applications of refutation to hardness of approximation, cryptography, and learning theory. Using these applications, we will motivate our study of δ-refuting CSPs with arbitrary predicates at superconstant constraint densities. Then we will survey previous work on refutation, and describe our results on refutation in the SOS proof system. We will conclude the chapter with an overview of the rest of this thesis. For clearer exposition, we restrict our discussion in this chapter to Boolean CSPs with alphabet size 2, but all of our results also apply to CSPs with any constant alphabet size.

1.2 Applications of refutation

Assumptions about the computational hardness of refuting random CSPs have proven useful in hardness of approximation, cryptography, and learning theory.

1.2.1 Hardness of approximation

Hardness of refutation of random CSP(P) implies *worst-case* hardness of approximation results for other natural computational problems. 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 hardness of a robust variant of graph isomorphism [OWWZ14]. In addition, Alon et al. showed that if random k-AND instances are hard to refute, then densest k-subgraph cannot be approximated to within any constant factor [AAM+11]. Barak, Kindler, and Steurer pointed out that the stronger assumption that the basic SDP relaxation is an optimal polynomial time refutation algorithm implies a stronger hardness of approximation for this problem [BKS13]. They also observe that a generalization of this assumption to larger alphabets implies hardness of random instances of label cover. Goerdt and Lanka weakened Feige's assumption to show that hardness of refuting random 4-SAT implies that bipartite clique is hard to approximate to within a factor of n^ϵ for fixed $\epsilon > 0$ [GL04].

1.2.2 Cryptography: Goldreich's pseudorandom generator

Given a predicate $P : \{0,1\}^k \to \{0,1\}$, Goldreich [Gol00] suggested the following function $f_P : \{0,1\}^n \to \{0,1\}^m$ as a candidate one-way function. Choose a collection $S_1, S_2, \ldots, S_m$ of k-tuples of elements of $[n]$ so that the corresponding k-uniform hypergraph has high expansion. Generate the ith output bit by applying P to the input variables indexed by S_i. A random k-uniform hypergraph has high expansion and here we will consider functions f_P based on a random hypergraph. 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. In one line of subsequent work, this function has also been studied as a candidate pseudorandom generator (PRG) [CM01, MST03, ABW10, App12, ABR12, OW14, AL16].

To show that f_P is a PRG, we need to show that it is hard to distinguish a uniform random $y \in \{0,1\}^m$ from y such that $f_P(x) = y$ for some x. We can think of the pair (f_P, y) as forming a CSP: Each constraint has the form $P(x_{S_i}) = y_i$. Note that the scope of each constraint is chosen randomly. When y is chosen uniformly at random, it is easy to show that the corresponding CSP is unsatisfiable. To show that f_P is a PRG, it suffices to show that it is hard distinguish between satisfiable instances of this random CSP in which there exists x such that $f_P(x) = y$ and unsatisfiable instances in which the constraint scopes and y are chosen uniformly at random. In particular, if we could refute instances with random y, f_P would not be a secure as a PRG. We point out that the hardness of refutation of this CSP is not equivalent to pseudorandomness of f_P, but is necessary for pseudorandomness to hold.

The pseudorandomness of such local functions implies existence of other cryptographic constructions. Applebaum, Barak, and Wigderson also showed that a public-key cryptosystem can be constructed based on the assumption that f_P is pseudorandom for a particular choice of P and another assumption related to hardness of the densest k-subgraph problem [ABW10]. The existence of a secure PRG with constant locality and $m = n^{1+\epsilon}$ for some constant $\epsilon > 0$ implies that secure two-party computation with constant overhead is possible [IKOS08].

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 Alekhnovich's average case XOR hardness assumptions [Ale03]. Alekhnovich [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.

1.2.3 Learning theory

In a sequence of recent works, Daniely and coauthors have made connections between hardness of refutation and hardness of natural learning problems [DLSS13, DLSS14, DS14, Dan15]. By making concrete conjectures about the hardness of refuting random CSP(P) for various P and for super-polynomial Δ, they obtained negative results for several longstanding problems in learning theory, such as learning DNFs and learning halfspaces with noise. In unrelated work, Barak and Moitra showed a connection between a learning problem called tensor prediction and strong refutation of k-XOR [BM16].

1.3 The problem

Most previous work on both upper and lower bounds for refutation has been on weak refutation of k-SAT. Most lower bounds have been proven for instances with only $O(n)$ constraints. However, the three applications above highlight the importance of studying arbitrary predicates, superlinear number of constraints, the stronger notion of δ refutation, and refutation in superpolynomial time.

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 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 CSP(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 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 CSP(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 CSP(P) 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, in this thesis we give 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 inspec-

tion of [DS14]'s hardness of learning DNF result shows that it essentially works by reduction from CSP(P) 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 CSP(P) for as wide a variety of predicates 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 CSP(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 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 CSP(P) has $\mathrm{Opt}(\mathcal{I}) \leq \mathbf{E}[P] + o(1)$, where $\mathbf{E}[P]$ is the probability a random assignment satisfies P. Thus it is quite natural to ask for δ-refutation for $\delta = 1 - \mathbf{E}[P] - o(1)$; i.e., for an algorithm that certifies the *true* value of $\mathrm{Opt}(\mathcal{I})$ 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 CSP(P) and approximation algorithms for CSP(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. $\mathbf{E}[P] + o(1)$ *integrality gap* instances for CSP(P).

4. **Refutation in superpolynomial time.** Naturally, we would prefer to have evidence against superpolynomial-time refutation, or even subexponential-time refutation, of random CSP(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 CSP(P) 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 P, determine the best possible tradeoff between number of constraints, refutation strength, and running time for random instances of* $\mathrm{CSP}(P)$.

This thesis, together with a result of Raghavendra, Rao, and Schramm [RRS17], completely accomplishes this goal for the sum of squares (SOS) proof system. Before stating our results, we review prior results in the direction of the above goal.

1.4 A brief history of refutation

Lower bounds. Nondeterministic refutation was first studied in the context of proof complexity. Cook and Reckhow showed that $\mathsf{NP} = \mathsf{coNP}$ if and only if the set of logical formulas encoding tautologies admits proofs of polynomials size, or equivalently, if and only if the set of contradictions admits refutations of polynomial size [CR79]. One avenue toward showing that $\mathsf{NP} \neq \mathsf{coNP}$ is then to prove lower bounds on the size of refutations in increasingly strong proof systems. Chvátal and Szemerédi [CS88] were the first to prove lower bounds for refutation of random instances, showing that any Resolution refutation of a random k-SAT instance with cn clauses must have size $2^{\Omega(n)}$ with high probability. [1]

Lower bounds for nondeterministic refutation have been proven in several other proof systems, usually for the special case of k-SAT or k-XOR. All of these results rely on high expansion of the underlying hypergraph. Ben-Sasson and Wigderson strengthened and simplified Resolution lower bounds for 3-SAT [BSW99, BS01], showing that Resolution refutations require width $\Omega(\frac{n}{\Delta^{1/(k-2)+\epsilon}})$ for any $\epsilon > 0$. This implies that Resolution refutations must have superpolynomial size when $m \leq n^{3/2-\epsilon}$. Ben-Sasson and Impagliazzo [BSI99] and Alekhnovich and Razborov [AR01] 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 $\Omega(\frac{n}{\Delta^{2/(k-2)}\log\Delta})$. This implies superpolynomial size lower bounds when $m = O(n)$ via the work of Impagliazzo, Pudlák, and Sgall [IPS99]. All of these results also extend to random k-SAT with $m \leq n^{k/2-\epsilon}$. Alekhnovich showed superpolynomial lower bounds for k-DNF Resolution refutations of 3-SAT when $m = O(n)$ [Ale05]. For k-SAT and k-XOR, Buresh-Oppenheim et al. proved linear rank lower bounds for Cutting Planes refutations when $m = O(n)$ [BOGH+03].

Lower bounds have been proven for polynomial-time SDP-based refutation. The strongest results are known for k-SAT and k-XOR. For the sum of squares (SOS) SDP relaxation, Grigoriev [Gri01] proved superconstant degree lower bounds on sum of squares refutations of $\mathrm{CSP}(P)$ for predicates P whose satisfying assignments include those of k-XOR (e.g., k-SAT) when $m \leq n^{k/2}$. 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. When $m = O(n)$, these results also imply superpolynomial size lower bounds [KI06]. Tulsiani [Tul09] extended this result to the alphabet-q generalization of random 3-XOR. Buresh-Oppenheim et al. [BOGH+03] 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 Tourlakis [AAT05] extended this result to random instances of 3-SAT.

[1] Although Chvátal and Szemerédi were the first to explicitly consider lower bounds for refutation of random instances, Galil [Gal77b, Gal77a] had earlier proven lower bounds for weaker versions of Resolution using SAT instances constructed from expander graphs. Galil's instances were not constructed in quite the same way as the random instances we study here. Instead, he used Tseitin's approach of constructing a 2-XOR instance such that all variables occur twice and the parity of the right-hand side is 1 and then translated this this to a SAT instance [Tse66]. Nevertheless, the underlying hypergraph of a random instance is a good expander and expansion has been the key property used to prove lower bounds for refutation of random CSPs (e.g., [CS88, BSW99, Gri01] etc.).

For other predicates, weaker lower bounds are known. Austrin and Mossel [AM08] established a connection between hardness of CSP(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 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. This result implies superconstant rank lower bounds for the proof system corresponding to SA_+ SDP relaxation and superpolynomial size lower bounds when $m = O(n)$ (again, via [KI06]). Tulsiani and Worah extended this result to the Lovász-Schrijver$_+$ SDP relaxation and its corresponding proof system [TW13]. In work not included in this thesis [OW14], we 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 CSP($P^{\pm}$) with $m = n^{t/2-\epsilon}$ constraints. We also showed the same result for the SA_+ SDP hierarchy, provided one can remove a carefully chosen $o(m)$ constraints from the random instance. Extending a result of Tulsiani and Worah [TW13], we showed in work not included in this thesis [MW16] that the degree-$n^{\Omega(\epsilon)}$ SA_+ and LS_+ SDP hierarchies cannot refute purely random instances of CSP($P^{\pm}$) with $m = n^{t/2-\epsilon}$ constraints. 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 CSP(P) 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. Feldman, Perkins, and Vempala [FPV15] proved lower bounds for refutation of CSP($P^{\pm}$) using statistical algorithms when P supports a $(t-1)$-wise uniform distribution, showing that any statistical algorithm based on an oracle taking L values requires $m \geq \widetilde{O}(n^t/L)$ to refute random instances of CSP(P). They further show that the dimension of any convex program refuting such a CSP must be at least $\widetilde{\Omega}(n^{t/2})$. 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.

Upper bounds. Beame and Pitassi [BP96] give the first nontrivial bounds for refutation runtime, observing that random 3-SAT can be refuted in quasipolynomial time when $m = O(n^2)$ by combining Fu's linear-size tree-like Resolution refutation [Fu96] with Clegg et al.'s algorithm for finding tree-like Resolution proofs in quasipolynomial time [CEI96]. Beame et al. showed that Resolution refutations of random k-SAT can be found efficiently when $m \geq O(n^{k-1}/\log^{k-2} n)$ [BKPS98], the first nontrivial polynomial time refutation algorithm.

Since then, 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. 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 LS_+ proof system cannot refute random instances of k-SAT with $k \geq 5$ and constant Δ. Alekhnovich, Arora, and Tourlakis [AAT05] extended this result to random instances of 3-SAT.

The next major advance came with the work of Goerdt and Krivelevich, who used spectral techniques to show that random k-SAT instances can be refuted when $m \geq \widetilde{O}(n^{\lceil k/2 \rceil})$ [GK01]. Friedman and Goerdt later used the same techniques to show that random 3-SAT instances can be refuted once $m \geq O(n^{3/2+\epsilon})$ [FG01]. Spectral methods underlie most subsequent work. Coja-

Oghlan, Goerdt, and Lanka gave strong refutation algorithms for 3-SAT with $m \geq \widetilde{O}(n^{3/2})$ and 4-SAT with $m \geq n^2$ [COGL04]. Coja-Oghlan, Cooper, and Frieze proved that and k-CSP can be strongly refuted when $m \geq \widetilde{O}(n^{\lceil k/2 \rceil})$ [COCF09]. Independently of the work in this thesis, Barak and Moirta showed that k-XOR instances could be refuted strongly when $m \geq \widetilde{O}(n^{k/2})$ [BM16].

Fewer upper bounds for nondeterministic refutation are known. Fu was the first to prove such a result, showing that polynomial-size Resolution refutations of random k-SAT instances exist with high probability when $m \geq O(n^{k-1})$ [Fu96]. Using both combinatorial and spectral methods, Feige, Kim, and Ofek showed that nondeterministic refutation of random 3-SAT is possible when $m \geq O(n^{1.4})$ [FKO06]. Up to this point, this is the only case in which nondeterministic refutation is possible at lower densities than polynomial-time refutation.

Beyond the uniform random model, Feige [Fei07] considered a *semi-random* model in which an instance is generated by starting with an arbitrary instance and then flipping the polarity of every literal independently with constant probability ϵ. Such instances are unsatisfiable with high probability. Feige showed that semi-random 3-SAT instances in this model can be refuted in polynomial time when $m = \widetilde{O}(n^{3/2})$ [Fei07]. Since random instances can be generated in this model, all of the above lower bounds also apply to semi-random instances.

1.5 Results

In this thesis, we focus on the Sum of Squares (also known as 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 SA_+ and LS_+ proof systems. In the context of $\mathrm{CSP}(P)$ over domain $\{0,1\}$, it is often also (approximately) automatizable in $n^{O(d)}$ time using semidefinite programming [RW17]. 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 in the worst case at least as well as *any* polynomial-size family of SDP relaxations. See, e.g., [OZ13, BS14, Lau09] for surveys concerning SOS.

For every predicate P, we provide a *full three-way tradeoff between constraint density, SOS degree, and strength of refutation.* To state our result, we need a definition. For a predicate $P : \{0,1\}^k \to \{0,1\}$ 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,

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

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

First, we provide a framework reducing δ-refutation of any CSP to strong refutation of XOR. Plugging in existing strong XOR refutation results, we obtain algorithms for refuting any CSP. Work of Barak and Moitra [BM16] and this thesis independently showed that random k-XOR can be strongly refuted for $m = \widetilde{O}(n^{k/2})$ via constant-degree SOS in polynomial time. Using this result, we proved upper bounds on the number of constraints needed to refute any CSP using constant-degree SOS.

Theorem 1.5.1 ([AOW15], Chapter 3)**.** *Let P be a k-ary Boolean predicate and let $1 < t \leq k$. Let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with $m = \widetilde{O}(n^{t/2})$ constraints. Then with high probability, constant-degree SOS $(\delta_P(t) - o(1))$-refutes $\mathcal{I}$. Furthermore, a refutation can be found in polynomial time.*

When t and k are constant, Corollary 3.1.10 in Chapter 3 states that if $\delta_P(t) \neq 0$, then $\delta_P(t) = \Omega(1)$. This gives the following corollary.

Corollary 1.5.2 ([AOW15], Chapter 3)**.** *Let P be a k-ary Boolean predicate for which there exists no t-wise uniform distribution supported on satisfying assignments. Let $\mathcal{I}$ be a random instance of* CSP(P) *with $m = \widetilde{O}(n^{t/2})$ constraints. Then with high probability, constant-degree SOS (weakly) refutes $\mathcal{I}$. Furthermore, a refutation can be found in polynomial time.*

This result subsumes almost all previous results on polynomial-time refutation of CSPs. Let $\mathcal{C}(P)$ be the minimum t for which P fails to support a t-wise uniform distribution on satisfying assignments. That is, $\mathcal{C}(P) = \min\{t : \delta_P(t) > 0\}$. When $m = \widetilde{O}(n^{\mathcal{C}(P)/2})$, random CSP$(P)$ can be weakly refuted in polynomial time. When $m = \widetilde{O}(n^{k/2})$, random CSP$(P)$ can be strongly refuted in polynomial time.

Both of these results also extend to CSPs in which variables take values from a alphabet of constant size greater than 2.

Extending the Barak–Moitra XOR refutation result [BM16] to higher-degree SOS, Raghavendra, Rao, and Schramm [RRS17] proved that random k-XOR with Δn constraints can be strongly refuted by degree-$\widetilde{O}\left(\frac{n}{\Delta^{2/(t-2)}}\right)$ SOS.

Corollary 1.5.3 ([RRS17])**.** *Let P be a k-ary Boolean predicate and let $2 < t \leq k$. Let $\mathcal{I}$ be a random instance of* CSP(P) *with $m = \Delta n$ constraints. Then with high probability, degree-$\widetilde{O}\left(\frac{n}{\Delta^{2/(t-2)}}\right)$ SOS $(\delta_P(t) - o(1))$-refutes $\mathcal{I}$. 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).

Second, we show that the full three-way tradeoff in Corollary 1.5.3 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.

Theorem 1.5.4 ([KMOW17], Chapter 4)**.** *Let P be a k-ary Boolean predicate and let $1 < t \leq k$. Let $\mathcal{I}$ be a random instance of* CSP(P) *with $m = \Delta n$ constraints. Then with high probability, degree-$\widetilde{\Omega}\left(\frac{n}{\Delta^{2/(t-1)}}\right)$ 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 1.5.5 ([KMOW17], Chapter 4)**.** *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* CSP(P) *with $m = \Delta n$ constraints, with high probability degree-$\widetilde{\Omega}\left(\frac{n}{\Delta^{2/(\mathcal{C}(P)-2)}}\right)$ SOS fails to (weakly) refute $\mathcal{I}$.*

Our lower bound subsumes all of the hardness results for semialgebraic proof systems mentioned in the previous section. 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 Theorems 1.5.4 and 1.5.5 imply the following corollaries.

Corollary 1.5.6 ([KMOW17])**.** *Let P be a k-ary Boolean predicate and let $1 < t \leq k$. Let $\mathcal{I}$ be a random instance of* CSP(P) *with $m = \widetilde{\Omega}(n^{(t+1)/2})$ constraints. Then with high probability, constant-degree SOS fails to $(\delta_P(t) + o(1))$-refute $\mathcal{I}$.*

Corollary 1.5.7 ([KMOW17]). *Let P be a k-ary predicate and let $\mathcal{C}(P)$ be the minimum integer $3 \le \tau \le k$ for which P fails to support a τ-wise uniform distribution. Then if $\mathcal{I}$ is a random instance of $\mathrm{CSP}(P)$ with $m = \widetilde{\Omega}(n^{\mathcal{C}(P)/2})$ constraints, with high probability constant-degree SOS fails to (weakly) refute $\mathcal{I}$.*

All of our lower bound results have no dependence on alphabet size and therefore hold with no change for CSPs over any alphabet.

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 if its three inputs is true. The resulting 3-SAT variant $\mathrm{CSP}(P)$ 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 $\delta_{\text{1-in-3-SAT}}(2) = \frac{1}{4}$. 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 $\delta_{\text{1-in-3-SAT}}(3) = \frac{5}{8}$.

Let us also note that as soon as Δ is a large enough constant, $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{8} + o(1)$ (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 $\mathrm{Opt}(\mathcal{I}) < 1$. But what can be said about stronger refutation? Let us see what our Theorem 1.5.4 and its counterpart Corollary 1.5.3 tell us.

Suppose first that there are $m = n\,\mathrm{polylog}(n)$ constraints. Corollary 1.5.3 tells us that constant-degree SOS certifies $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{4} + o(1)$. However our result, Theorem 1.5.4, says this $\frac{3}{4}$ cannot be improved: SOS cannot certify $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{4} - o(1)$ until the degree is as large as $\widetilde{\Omega}(n)$. (Of course at degree n, SOS can certify the exact value of $\mathrm{Opt}(\mathcal{I})$.)

What if there are $m = n^{1.1}$ constraints, meaning $\Delta = n^{.1}$? Our result says SOS still cannot certify $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{4} - o(1)$ 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 $\widetilde{O}(n^{.8})$, SOS *does* certify $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{4} - o(1)$; in fact, it certifies $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{8} + o(1)$.

Similarly (dropping lower-order terms for brevity), if there are $m = n^{1.2}$ constraints, SOS is stuck at certifying just $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{4}$ up until degree $n^{.6}$, at which point it jumps to being able to certify the truth, $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{8} + o(1)$. If there are $n^{1.49}$ constraints, SOS remains stuck at certifying just $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{4}$ up until degree $n^{.02}$. Finally, Theorem 1.5.1 shows that, once $m = n^{1.5}\,\mathrm{polylog}(n)$, constant-degree SOS can certify $\mathrm{Opt}(\mathcal{I}) \le \frac{3}{8} + o(1)$. **(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.

1.6 Organization

In Chapter 2, we more formally introduce the uniform random CSP model we study and the SOS proof system. In Chapter 3, we describe our framework for reducing refutation of arbitrary CSPs to

strong refutation of XOR and prove Theorem 1.5.1. In Chapter 4, we prove our SOS lower bounds for refutation (Theorem 1.5.4). Finally, we conclude in Chapter 5 with a discussion of directions for future work on refutation.

Chapter 2

Preliminaries

2.1 Constraint satisfaction problems

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

Definition 2.1.1. Given a predicate $P : \{-1, 1\}^k \to \{0, 1\}$, an instance $\mathcal{I}$ of the CSP(P) problem over variables $x_1, \ldots, x_n$ is a multiset of *P-constraints*. Each P-constraint consists of a pair (c, S), where $S \in [n]^k$ is the scope and $c \in \{-1, 1\}^k$ is the negation pattern; this represents the constraint $P(c \circ x_S) = 1$. We typically write $m = |\mathcal{I}|$. Let $\mathrm{Val}_{\mathcal{I}}(x)$ be the fraction of constraints satisfied by assignment $x \in \{-1, 1\}^n$, i.e., $\mathrm{Val}_{\mathcal{I}}(x) = \frac{1}{m} \sum_{(c,S) \in \mathcal{I}} P(c \circ x_S)$. The objective is to find an assignment x maximizing $\mathrm{Val}_{\mathcal{I}}(x)$. The *optimum* of $\mathcal{I}$, denoted by $\mathrm{Opt}(\mathcal{I})$, is $\max_{x \in \{-1,1\}^k} \mathrm{Val}_{\mathcal{I}}(x)$. If $\mathrm{Opt}(\mathcal{I}) = 1$, we say that $\mathcal{I}$ is *satisfiable*. We also write $\mathbf{E}[P]$ for the quantity $\mathbf{E}_{z \sim \{-1,1\}^k}[P(z)]$; i.e., the fraction of assignments that satisfy P.

Two random models for CSPs. We next define two standard random model for CSPs.

1. For $P : \{-1, 1\}^k \to \{0, 1\}$, let $\mathcal{F}_P(n, m)$ be the distribution in which we choose m constraints uniformly at random from among the set all $2^k \binom{n}{k}$ possible constraints on k distinct variables. Note that we may include constraints on different permutations of the same set of variables and constraints on the same tuple of variables with different negations c. It is reasonable to include such constraints in the case that the predicate P is not a symmetric function. We prove our lower bounds in Chapter 4 in this model.
2. Let $\mathcal{F}_P(n, p)$ be the distribution over CSP instances given by including each of the $2^k n^k$ possible constraints on k not necessarily distinct variables independently with probability p.

   Our upper bounds in Chapter 3 are most easily proven in this model, but it is not hard to see that they also hold in the $\mathcal{F}_P(n, m)$ model. We use $\overline{m}$ to denote the expected number of constraints, namely $2^k n^k p$. As noted in Fact 2.1.6 below, the number of constraints m in a draw from $\mathcal{F}_P(n, p)$ is very tightly concentrated around $\overline{m}$, and we often blur the distinction between these parameters. Section 3.8 explicitly describes a method for simulating an instance drawn from $\mathcal{F}_P(n, p)$ when the number of constraints is fixed. In addition, note that we may include constraints with the same variable occurring as more than one argument. There are only $o(m)$ such constraints with high probability, so they do not significantly affect our upper bounds.

On the other hand, it is not immediately clear to us how to extend our lower bounds in Chapter 4 to the $\mathcal{F}_P(n,p)$ model.

Quasirandomness. We now introduce an important notion for this paper: that of a CSP instance being *quasirandom*. Versions of this notion originate in the works of Goerdt and Lanka [GL03] (under the name "discrepancy"), Khot [Kho06] ("quasi-randomness"), Austrin and Håstad [AH13] ("adaptive uselessness"), and Chan [Cha13] ("low correlation"), among other places. To define it, we first need to define the induced distribution of an instance and an assignment.

Definition 2.1.2. Given a CSP instance $\mathcal{I}$ and and an assignment $x \in \{-1,1\}^n$, the *induced distribution*, denoted $\mathcal{D}_{\mathcal{I},x}$, is the probability distribution on $\{-1,1\}^k$ where the probability mass on $\alpha \in \{-1,1\}^k$ is given by $\mathcal{D}_{\mathcal{I},x}(\alpha) = \frac{1}{|\mathcal{I}|} \cdot \#\{(c,S) \in \mathcal{I} \mid c \circ x_S = \alpha\}$. In other words, it is the empirical distribution on inputs to P generated by the constraint scopes/negations on assignment x. Note that the predicate P itself is irrelevant to this notion. We will drop the subscript $\mathcal{I}$ when it is clear from the context. We define $D_{\mathcal{I},x} = 2^k \cdot \mathcal{D}_{\mathcal{I},x}$ to be the density function associated with $\mathcal{D}_{\mathcal{I},x}$.

We can now define quasirandomness.

Definition 2.1.3. A CSP instance $\mathcal{I}$ is *ϵ-quasirandom* if $\mathcal{D}_{\mathcal{I},x}$ is ϵ-close to the uniform distribution for all $x \in \{-1,1\}^n$; i.e., if $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x}, U^k) \le \epsilon$ for all $x \in \{-1,1\}^n$.

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

Definition 2.1.4. If $\mathcal{D}$ and $\mathcal{D}'$ are probability distributions on the same finite set A then $d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}')$ denotes their *total variation distance*, $\frac{1}{2}\sum_{\alpha \in A} |\mathcal{D}(\alpha) - \mathcal{D}'(\alpha)|$. If $d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}') \le \epsilon$ we say that $\mathcal{D}$ and $\mathcal{D}'$ are *ϵ-close*. If $d_{\mathrm{TV}}(\mathcal{D},\mathcal{D}') \ge \epsilon$ we say they are *ϵ-far*. (As neither inequality is strict, these notions are not quite opposites.)

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

Fact 2.1.5. *If $\mathcal{I}$ is ϵ-quasirandom, then* $\mathrm{Opt}(\mathcal{I}) \le \mathbf{E}[P] + \epsilon$ *(and in fact,* $|\mathrm{Opt}(\mathcal{I}) - \mathbf{E}[P]| \le \epsilon$*).*

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

Fact 2.1.6. *Let $\mathcal{I} \sim \mathcal{F}_P(n,p)$. Then the following statements hold with high probability.*

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

2.1.1 Algorithms and refutations on random CSPs

Definition 2.1.7. Let P be a Boolean predicate. We say that $\mathcal{A}$ is *δ-refutation algorithm for random* $\mathrm{CSP}(P)$ *with $\overline{m}$ constraints* if $\mathcal{A}$ has the following properties. First, on all instances $\mathcal{I}$ the output of $\mathcal{A}$ is either the statement "$\mathrm{Opt}(\mathcal{I}) \le 1-\delta$" or is "fail". Second, $\mathcal{A}$ is *never* allowed to *err*, where erring means outputting "$\mathrm{Opt}(\mathcal{I}) \le 1-\delta$" on an instance which actually has $\mathrm{Opt}(\mathcal{I}) > 1-\delta$. Finally, $\mathcal{A}$ must satisfy

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

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

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

Remark 2.1.8. In Section 3.3 we will encounter a "two-sided error" variant of this definition. This is the slightly easier algorithmic task in which we relax the condition on erring: it is only required that for each instance $\mathcal{I}$ with $\mathrm{Opt}(\mathcal{I}) > 1 - \delta$, it holds that $\mathbf{Pr}[\mathcal{A}(\mathcal{I}) = \text{"}\mathrm{Opt}(\mathcal{I}) \leq 1 - \delta\text{"}] \leq 1/4$, where the probability is just over the random coins of $\mathcal{A}$.

Remark 2.1.9. We will also use the analogous definition for certification of related properties; e.g., we will discuss *ϵ-quasirandomness certification algorithms* in which the statement "$\mathrm{Opt}(\mathcal{I}) \leq 1 - \delta$" is replaced by the statement "$\mathcal{I}$ is ϵ-quasirandom".

2.1.2 t-wise uniformity

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

Definition 2.1.10. Probability distribution $\mathcal{D}$ on $\{-1,1\}^k$ is said to be *t-wise uniform*, $1 \leq t \leq k$, if for all $S \subseteq [k]$ with $|S| = t$ the random variable x_S is uniform on $\{-1,1\}^t$ when $x \sim \mathcal{D}$. (We remark that this condition is sometimes inaccurately called "t-wise independence" in the literature.)

This definition generalizes naturally to distributions with larger alphabets Ω: A probability distribution on Ω^k is said to be t-wise uniform if its marginal on every subset of t coordinates is uniform.

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

Definition 2.1.11. Probability distribution $\mathcal{D}$ on $\{-1,1\}^k$ is said to be *(ϵ, t)-wise uniform* if $|\widehat{D}(S)| \leq \epsilon$ for all $S \subseteq [k]$ with $0 < |S| \leq t$, where $D = 2^k \cdot \mathcal{D}$ is the probability density associated with distribution $\mathcal{D}$.

Here we are using standard notation from Fourier analysis of Boolean functions [O'D14]. In particular, for any $f : \{-1,1\}^k \to \mathbb{R}$ we write $f(x) = \sum_{S \subseteq [k]} \widehat{f}(S) x^S$ for its expansion as a multilinear polynomial over $\mathbb{R}$, with x^S denoting $\prod_{i \in S} x_i$ (not to be confused with the projection $x_S \in \mathbb{R}^{|S|}$).

Remark 2.1.12. It is a simple fact (and it follows from Lemma 2.1.13 below) that $(0, t)$-wise uniformity is equivalent to t-wise uniformity.

Also important for us is a related but distinct notion, that of being ϵ-close to a t-wise uniform distribution. It's easy to show that if $\mathcal{D}$ is ϵ-close to a t-wise uniform distribution then $\mathcal{D}$ is $(2\epsilon, t)$-wise uniform. In the other direction, we have the following (see also [AAK+07] for some quantitative improvement):

Lemma 2.1.13. (Alon–Goldreich-Mansour [AGM03, Theorem 2.1]). *If $\mathcal{D}$ is an (ϵ, t)-wise uniform distribution on $\{-1,1\}^k$, then there exists a t-wise uniform distribution $\mathcal{D}'$ on $\{-1,1\}^k$ with*

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

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

Finally, we make a crucial definition:

Definition 2.1.14. A predicate $P : \{-1,1\}^k \to \{0,1\}$ is said to be *t-wise supporting* if there is a t-wise uniform distribution $\mathcal{D}$ whose support is contained in $P^{-1}(1)$. In other words, P is t-wise supporting if $\delta_P(t) = 0$. We say P is *δ-far from t-wise supporting* if every t-wise uniform distribution $\mathcal{D}$ is δ-far from being supported on P; i.e., has probability mass at least δ on $P^{-1}(0)$. That is, P is δ-far from t-wise supporting if $\delta_P(t) \geq \delta$.

2.2 The sum of squares proof system and hierarchy

We give a brief overview of the sum of squares proof system and SDP hierarchy here. First, we define SOS as a proof system. Then we describe the dual view of SOS as a hierarchy of SDP relaxations optimizing over objects called pseudoexpectations. For more background see, e.g., [OZ13, Lau09, BS].

2.2.1 The sum of squares proof system

The SOS proof system, introduced in [GV01], certifies bounds on the values of polynomial optimization problems; i.e., maximizing an n-variate polynomial subject to polynomial inequality and equality constraints. Call a polynomial $q \in \mathbb{R}[X_1, \ldots, X_n]$ sum-of-squares (SOS) if there exist $q_1, \ldots, q_\ell \in \mathbb{R}[X_1, \ldots, X_n]$ such that $q = q_1^2 + \cdots + q_\ell^2$.

Definition 2.2.1. Let $X = (X_1, \ldots, X_n)$ be indeterminates. Let $q_1, \ldots, q_m, r_1, \ldots, r_{m'} \in \mathbb{R}[X]$ and let $A = \{q_1 \geq 0, \ldots, q_m \geq 0\} \cup \{r_1 = 0, \ldots, r_{m'}\}$. SOS is parameterized by degree d: the larger d is, the more powerful the proof system. There is a degree-d SOS proof that A implies $s \geq 0$, written as

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

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

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

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

It is well-known that a degree-d SOS proof can be found by solving an SDP of size $n^{O(d)}$ if it exists [Sho87, Par00, Las00, Las01]. It is not always clear that these SOS SDPs can be solved in polynomial time [O'D17, RW17]. However, our SOS upper bound in Chapter 3 does lead to an efficient algorithm, as the corresponding SOS proof can be found simply by computing the spectral norm of a matrix.

We can think of a CSP $\mathcal{I} = \{(c, S)\}$ over n Boolean variables $x_1, \ldots, x_n$ as a polynomial feasibility problem, with (the arithmetization of) the constraints $P(x_S \oplus c) = 1$ as polynomial identities. An SOS refutation of degree-d is then a degree-d SOS proof that $\{P(x_S \oplus c) = 1\}$ together with the Boolean constraints $\{x_i^2 = x_i\}$ implies $-1 \geq 0$, or, in symbols, that $\{P(x_S \oplus c) = 1\} \cup \{x_i^2 = x_i\} \vdash_d -1 \geq 0$. For δ-refutation, we consider the polynomial $\frac{1}{m}\sum_{(c,S)\in\mathcal{I}} P(x_S \oplus c)$ capturing the fraction of satisfied constraints. An SOS δ-refutation of degree-d is a degree-d SOS proof that $\{\frac{1}{m}\sum_{(c,S)\in\mathcal{I}} P(x_S \oplus c) \geq 1 - \delta\} \cup \{x_i^2 = x_i\} \vdash_d -1 \geq 0$.

2.2.2 The dual view: Pseudoexpectations

In the dual view, SOS is a hierarchy of SDP-based relaxations of polynomial optimization problems, again parameterized by degree. Central to the algorithm is the concept of *pseudoexpectations* that describe the feasible points of the SOS algorithm of degree d.

Definition 2.2.2 (Pseudoexpectations)**.** Given n indeterminates, a degree-d pseudoexpectation is a linear operator $\widetilde{\mathbf{E}}$ on the space of real polynomials of degree at most d in those indeterminates, such that

1. $\widetilde{\mathbf{E}}[1] = 1$. (Normalization)
2. $\widetilde{\mathbf{E}}[p^2] \geq 0$ for every polynomial p of degree at most $d/2$. (Positive semidefiniteness)

Definition 2.2.3 (Pseudoexpectations satisfying an identity)**.** A degree-d pseudoexpectation $\widetilde{\mathbf{E}}$ is said to *satisfy a polynomial identity* "$p = 0$" if, for every polynomial q with $\deg(p) + \deg(q) \leq d$, we have $\widetilde{\mathbf{E}}[pq] = 0$.

Given a polynomial optimization problem — say, maximizing a polynomial p_1 subject to constraints $\{q_i = 0 \mid i \in [m]\}$ — the degree-d SOS relaxation maximizes $\widetilde{\mathbf{E}}[p_1]$ over all degree-d pseudoexpectations $\widetilde{\mathbf{E}}$ that satisfy the identities $\{q_i = 0 \mid i \in [m]\}$. A feasibility problem, in particular, would ask if there is a degree-d pseudoexpectation satisfying certain polynomial equality constraints. As in the dual proof system, these SOS relaxations can be expressed using a SDP of size $n^{O(d)}$.

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.

To show that the degree-d SOS refutation algorithm cannot refute a CSP 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. Similarly, to show that degree-d SOS cannot δ-refute a CSP, we need to show that there exists a degree-d pseudoexpectation that satisfies at least a $1 - \delta$ fraction of the constraints.

Chapter 3

A framework for refuting random CSPs using sum of squares

3.1 Our results and techniques

Here we describe our main results and techniques at a high level. Precise theorem statements appear later in the work and the definitions of the terminology we use is given in Chapter 2. We also mention that all of our results can be generalized to the case of larger alphabets, but we discuss Boolean predicates $P : \{0,1\}^k \to \{0,1\}$ for simplicity. Our main result gives a bound on the number of constraints needed to refute random CSP(P) instances. Before getting to it, we first describe some more concrete results that go into the main proof. All of our results rely on a strong refutation algorithm for k-XOR (actually, a slight generalization thereof). For $m \geq \widetilde{O}(n^{\lceil k/2 \rceil})$, such a result follows from [COCF10]; however, the exponent $\lceil k/2 \rceil$ can be improved to $k/2$. We give a demonstration of this fact herein; as mentioned earlier, it was published very recently by Barak and Moitra [BM16, Theorem 14].

Theorem 3.1.1. *There is an efficient algorithm that (whp) strongly refutes random k-XOR instances with at least $\widetilde{O}(n^{k/2})$ constraints; i.e., it certifies* $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + o(1)$.

The proof of Theorem 3.1.1 follows ideas from [COGL07] and earlier works on "discrepancy" of random k-SAT instances. The case of even k is notably easier, and we present two "folklore" arguments for it. The case of odd k is trickier. Roughly speaking we view the instance as a homogeneous degree-k multilinear polynomial, which we want to certify takes on only small values on inputs in $\{-1,1\}^n$. Considering separately the contributions based on the "last" of the k variables in each constraint, and then using Cauchy–Schwarz, it suffices to bound the norm of a carefully designed quadratic form of dimension n^{k-1}, indexed by tuples of $k-1$ variables. This is done using the trace method [Wig55, FK81]. Similar techniques, including the use of the trace method, date back to the 2001 Friedman–Goerdt work [FG01] refuting random 3-SAT given $m = n^{3/2+\epsilon}$ constraints.

With Theorem 3.1.1 in hand, the next step is certifying *quasirandomness* of random k-ary CSP instances having $m \geq \widetilde{O}(n^{k/2})$ constraints. Roughly speaking we say that a CSP instance is quasirandom if, for every assignment $x \in \{0,1\}^n$, the m induced k-tuples of literal values are close to being uniformly distributed over $\{0,1\}^k$. (Note that this is only a property of the instances' constraint scopes/negations, and has nothing to do with P.) Since the "Vazirani XOR Lemma" implies that a distribution on $\{-1,1\}^k$ is uniform if and only if all its 2^k XORs have bias 0, we are able to leverage Theorem 3.1.1 to prove:

Theorem 3.1.2. *There is an efficient algorithm that (whp) certifies that a random instance of* $\mathrm{CSP}(P)$ *is quasirandom, provided the number of constraints is at least* $\widetilde{O}(n^{k/2})$.

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

Theorem 3.1.3. *For any* k*-ary predicate* P*, there is an efficient algorithm that (whp) strongly refutes random* $\mathrm{CSP}(P)$ *instances when the number of constraints is at least* $\widetilde{O}(n^{k/2})$.

In particular, this theorem improves upon [COCF10] by a factor of $\sqrt{n}$ whenever k is odd; this savings is new even in the well-studied case of k-SAT.

The above result does not make use of any properties of the predicate P other than its arity, k. We now come to our main result, which shows that for many interesting P, random $\mathrm{CSP}(P)$ instances can be refuted with many fewer constraints than $n^{k/2}$.

Theorem 1.5.1 ([AOW15])**.** Let P be a k-ary Boolean predicate and let $1 < t \leq k$. Let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with $m = \widetilde{O}(n^{t/2})$ constraints. Then with high probability, constant-degree SOS $(\delta_P(t) - o(1))$-refutes $\mathcal{I}$. Furthermore, a refutation can be found in polynomial time.

In Corollary 3.1.10 below, we show that if $\delta_P(t) \neq 0$, then $\delta_P(t) = \Omega(1)$. This gives the following corollary.

Corollary 1.5.2 ([AOW15])**.** Let P be a k-ary Boolean predicate for which there exists no t-wise uniform distribution supported on satisfying assignments. Let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with $m = \widetilde{O}(n^{t/2})$ constraints. Then with high probability, constant-degree SOS (weakly) refutes $\mathcal{I}$. Furthermore, a refutation can be found in polynomial time.

In Section 3.2, we prove that we can $(\delta_P(t) - o(1))$-refute in polynomial time, and, in Section 3.4, we show the following:

Theorem 3.1.4. *All of our refutation algorithms for* k*-ary predicates can be extended to produce degree-*$2k$ *SOS proofs.*

The idea behind the proof of Theorem 1.5.1 is that with $\widetilde{O}(n^{t/2})$ constraints we can use the algorithm of Theorem 3.1.2 to certify quasirandomness (closeness to uniformity) for all subsets of t out of k coordinates. Thus for every assignment $x \in \{0,1\}^n$, the induced distribution on constraint k-tuples is ($o(1)$-close to) t-wise uniform. Since P is $\delta_P(t)$-far from supporting a t-wise uniform distribution, this essentially shows that no x can induce a distribution on constraint inputs with more that $1 - \delta_P(t)$ weight on satisfying assignments.

Example 3.1.5. To briefly illustrate the result, consider the Exactly-k-out-of-$2k$-SAT CSP, studied previously in [BB02, GJ03]. The associated predicate supports a 1-wise uniform distribution, namely the uniform distribution over strings in $\{0,1\}^{2k}$ of Hamming weight k. However, it is not hard to show that it does not support any pairwise uniform distribution. As a consequence, random instances of this CSP can be refuted with only $\widetilde{O}(n)$ constraints, independent of k.

3.1.1 An application from learning theory

Recently, an exciting approach to proving hardness-of-learning results has been developed by Daniely, Linial, and Shalev-Shwartz [DLSS13, DLSS14, DS14, Dan15]. The most general results appear in [DLSS14]. In this work, Daniely et al. prove computational hardness of several central

learning theory problems, based on two assumptions concerning the hardness of random CSP refutation. While the assumptions made in [DS14, Dan15] appear to be plausible, our work unfortunately shows that the more general assumptions made in [DLSS14] are false.

Below we state the (admittedly strong) assumptions from [DLSS14] (up to some very minor technical details which are discussed and treated in Section 3.3). We will need one piece of terminology: the 0-*variability* $\mathrm{VAR}_0(P)$ of a predicate $P : \{-1,1\}^k \to \{0,1\}$ is the least c such that there exists a restriction to some c input coordinates forcing P to be 0. Essentially, the assumptions state that one can obtain hardness-of-refutation with an arbitrarily large polynomial number of constraints by using a family of predicates (P_k) that: a) have unbounded 0-variability; b) support pairwise uniformity. However, our work shows that supporting t-wise uniformity for unbounded t is also necessary.

SRCSP Assumption 1. *([DLSS14].) For all $d \in \mathbb{N}$ there is a large enough C such that the following holds: If $P : \{-1,1\}^k \to \{0,1\}$ has $VAR_0(P) \geq C$ and is hereditarily approximation resistant on satisfiable instances, then there is no polynomial-time algorithm refuting (whp) random instances of CSP(P) with $m = n^d$ constraints.*

SRCSP Assumption 2. *([DLSS14], generalizing the "RCSP Assumption" of [BKS13] to super-linearly many constraints.) For all $d \in \mathbb{N}$ there is a large enough C such that the following holds: If $P : \{-1,1\}^k \to \{0,1\}$ has $VAR_0(P) \geq C$ and is δ-close to supporting a pairwise uniform distribution, then for all $\epsilon > 0$ there is no polynomial-time algorithm that $(\delta + \epsilon)$-refutes (whp) random instances of CSP(P) with $m = n^d$ constraints.*

In [DLSS14] it is shown how to obtain three very notable hardness-of-learning results from these assumptions. However as stated, our work falsifies the SRCSP Assumptions. Indeed, the assumptions are false even in the three specific cases used by [DLSS14] to obtain hardness-of-learning results. We now describe these cases.

Case 1. The Huang predicates (H_κ) are arity-$\Theta(\kappa^3)$ predicates introduced in [Hua13]; they are hereditarily approximation resistant on satisfiable instances and have 0-variability $\Omega(\kappa)$. In [DLSS14] they are used in SRCSP Assumption 1 to deduce hardness of PAC-learning DNFs with $\omega(1)$ terms. However:

Theorem 3.1.6. *For all $\kappa \geq 9$, the predicate H_κ does not support a 4-wise uniform distribution.*

Thus by Corollary 1.5.2 we can efficiently refute random instances of CSP(H_κ) with just $\widetilde{O}(n^2)$ constraints, independent of κ. This contradicts SRCSP Assumption 1.

Case 2. The majority predicate Maj_k has 0-variability $\lceil k/2 \rceil$ and is shown in [DLSS14] to be $\frac{1}{k+1}$-far from supporting a pairwise uniform distribution. In [DLSS14] these predicates are used in SRCSP Assumption 2 to deduce hardness of agnsotically learning Boolean halfspaces to within any constant factor. However:

Theorem 3.1.7. *For odd $k \geq 25$, the predicate Maj_k does not support a 4-wise uniform distribution; in fact, it is .1-far from supporting a 4-wise uniform distribution.*

Theorem 1.5.1 then implies we can efficiently δ-refute random instances of CSP(Maj_k) with $\widetilde{O}(n^2)$ constraints, where $\delta = .1 \gg \frac{1}{k+1}$. This contradicts SRCSP Assumption 2.

Case 3. Finally, we also prove that SRCSP Assumption 1 is false for another family of predicates (T_k) used by [DLSS14] to show hardness of PAC-learning intersections of 4 Boolean halfspaces.

Our results described in these three cases all use linear programming duality. Specifically, in Lemma 3.1.9 we show that P is δ-far from supporting a t-wise uniform distribution if and only if there exists a k-variable multilinear polynomial Q that satisfies certain properties involving P and δ. We then explicitly construct these dual polynomials for the Huang, Majority, and T_k predicates.

We conclude this section by emphasizing the importance of the Daniely–Linial–Shalev-Shwartz hardness-of-learning program, despite the above results. Indeed, subsequently to [DLSS14], Daniely and Shalev-Shwartz [DS14] showed hardness of improperly learning DNF formulas with $\omega(\log n)$ terms under a much weaker assumption than SRCSP Assumption 1. Specifically, their work only assumes that for all d there is a large enough k such that refuting random k-SAT instances is hard when there are $m = n^d$ constraints. This assumption looks quite plausible to us, and may even be true with k not much larger than $2d$. Most recently, Daniely showed hardness of approximately agnostically learning halfspaces using the XOR predicate rather than majority [Dan15]. This result shows that there is no efficient algorithm that agnostically learns halfspaces to within a constant approximation ratio under the assumption that refuting random k-XOR instances is hard when $m = n^{c\sqrt{k}\log k}$ for some $c > 0$. He also shows hardness of learning halfspaces to within an approximation factor of $2^{\log^{1-\nu} n}$ for any $\nu > 0$ assuming that there exists some constant $c > 0$ such that for all s, refuting random k-XOR instances with $k = \log^s n$ is hard when $m = n^{ck}$.

3.1.2 A dual characterization of limited uniformity

It is known that the condition of P supporting a t-wise uniform distribution is equivalent to the feasibility of a certain linear program; hence one can show that P is *not* t-wise supporting by exhibiting a certain dual object, namely a polynomial. This appears, e.g., in work of Austrin and Håstad [AH09, Theorem 3.1]. Herein we will extend this fact by giving a dual characterization of being far from t-wise supporting.

Definition 3.1.8. Let $0 < \delta < 1$. For a multilinear polynomial $Q : \{-1,1\}^k \to \mathbb{R}$, we say that Q *δ-separates* $P : \{-1,1\}^k \to \{0,1\}$ if the following conditions hold:

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

We now provide the quantitative version of the aforementioned [AH09, Theorem 3.1]:

Lemma 3.1.9. *Let $P : \{-1,1\}^k \to \{0,1\}$ and let $0 < \delta < 1$. Then P is δ-far from t-wise supporting if and only if there is a δ-separating polynomial for P of degree at most t.*

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

$$\begin{aligned}
\text{minimize} \quad & \sum_{z\in\{-1,1\}^k} (1-P(z))\mathcal{D}(z) && && (3.1)\\
\text{s.t.} \quad & \sum_{z\in\{-1,1\}^k} \mathcal{D}(z)z^S = 2^k\cdot\widehat{\mathcal{D}}(S) = 0 && \forall S\subseteq[k] \quad 0<|S|\le t && (3.2)\\
& \sum_{z\in\{-1,1\}^k} \mathcal{D}(z) = 1 && && (3.3)\\
& \mathcal{D}(z)\ge 0 && \forall z\in\{-1,1\}^k &&
\end{aligned}$$

Constraint (3.3) and the nonnegativity constraint ensure that $\mathcal{D}$ is a probability distribution on $\{-1,1\}^k$. Constraint (3.2) expresses that $\mathcal{D}$ is t-wise uniform (see Remark 2.1.12). The objective (3.1) is minimizing the probability mass that $\mathcal{D}$ puts on assignments in $P^{-1}(0)$. Thus the optimal value of the LP is equal to the smallest γ such that P is γ-close to t-wise supporting; equivalently, the largest γ such that P is γ-far from t-wise supporting.

The following is the dual of the above LP. It has a variable $c(S)$ for each $0<|S|\le t$ as well as a variable ξ corresponding to constraint (3.3).

$$\begin{aligned}
\text{maximize} \quad & \xi && && (3.4)\\
\text{s.t.} \quad & \sum_{\substack{S\subseteq[k]\\ 0<|S|\le t}} c(S)z^S \le 1-P(z)-\xi && \forall z\in\{-1,1\}^k. && (3.5)
\end{aligned}$$

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

Thus P is δ-far from t-wise supporting if and only if the LP's objective (3.1) is at least δ, if and only if the dual's objective (3.4) is at least δ, if and only if there is a δ-separating polynomial for P of degree at most t. □

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

Corollary 3.1.10. *Suppose $P:\{-1,1\}^k\to\{0,1\}$ is not t-wise supporting. Then it is in fact δ-far from t-wise supporting for $\delta=2^{-\widetilde{O}(k^t)}$ (or $\delta=2^{-\widetilde{O}(2^k)}$ when $t=k$).*

Proof. Let $K=1+\sum_{i=1}^t\binom{k}{t}$ be the number of variables in the dual LP from Lemma 3.1.9, so $K\le k^t+1$ in general, with $K\le 2^k$ when $t=k$. By assumption, the objective (3.4) of the dual LP's optimal solution is strictly positive. This optimum occurs at a vertex, which is the solution of a linear system given by a $K\times K$ matrix of ±1 entries and a "right-hand side" vector with $0,1$ entries. By Cramer's rule, the solution's entries are ratios of determinants of integer matrices with entries in $\{-1,0,1\}$. Thus any strictly positive entry is at least $1/N$, where N is the maximum possible such determinant. By Hadamard's inequality, $N=K^{K/2}$ and the claimed result follows. □

Using Corollary 3.1.10, we obtain the following corollary of Theorem 1.5.1.

Corollary 1.5.2 ([AOW15])**.** Let P be a k-ary Boolean predicate for which there exists no t-wise uniform distribution supported on satisfying assignments. Let $\mathcal{I}$ be a random instance of CSP(P) with $m=\widetilde{O}(n^{t/2})$ constraints. Then with high probability, constant-degree SOS (weakly) refutes $\mathcal{I}$. Furthermore, a refutation can be found in polynomial time.

3.1.3 Certifying independence number and chromatic number of random hypergraphs

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

Theorem 3.1.11. *For a random k-uniform hypergraph H, there is a polynomial time algorithm certifying that the independence number of H is at most β with high probability when H has at least $\widetilde{O}\left(\frac{n^{3k/2}}{\beta^k}\right)$ hyperedges.*

Theorem 3.1.12. *For a random k-uniform hypergraph H, there is a polynomial time algorithm certifying that the chromatic number of H is at least ξ with high probability when H has at least $\widetilde{O}\left(\xi^k n^{k/2}\right)$ hyperedges.*

The proofs of these theorems follow the outline of [COGL07]. We show Theorem 3.1.11 using a slightly more general form of Theorem 3.1.1. Theorem 3.1.12 follows almost directly from Theorem 3.1.11 using the fact that every color class of a valid coloring is an independent set. Details are given in Section 3.6.

3.2 Quasirandomness and its implications for refutation

3.2.1 Strong refutation of k-XOR

In this section, we state our result on strong refutation of random k-XOR instances with $m = \tilde{O}\left(n^{k/2}\right)$ constraints. (Recall that essentially this result was very recently obtained by Barak and Moitra [BM16].) We actually have a slightly more general result, allowing variables and coefficients to take values in $[-1, 1]$ and not just in $\{-1, 1\}$.

Theorem 3.2.1. *For $k \geq 2$ and $p \geq n^{-k/2}$, let $\{w(T)\}_{T\in[n]^k}$ be independent random variables such that for each $T \in [n]^k$,*

$$\mathbf{E}[w(T)] = 0 \tag{3.6}$$
$$\mathbf{Pr}[w(T) \neq 0] \leq p \tag{3.7}$$
$$|w(T)| \leq 1. \tag{3.8}$$

Then there is an efficient algorithm certifying that

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

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

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

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

Corollary 3.2.2. *For $k \geq 2$, let $\mathcal{I} \sim \mathcal{F}_{k\text{-XOR}}(n, p)$. Then, with high probability, there is a degree-$2k$ SOS proof that $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + \gamma$ when $m \geq \frac{2^{O(k)}n^{k/2}\log^3 n}{\gamma^2}$.*

Proof. We can write the k-XOR predicate as

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

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

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

where $w(T) = -2^{-k} \sum_{c \in \{\pm 1\}^k} 1_{\{(T,c) \in \mathcal{I}\}} \prod_{i \in [k]} c_i$. The $w(T)$'s are random variables depending on the choice of $\mathcal{I}$; observe that $\mathbf{E}[w(T)] = 0$, $\mathbf{Pr}[w(T) \neq 0] \leq 2^k p$, and $|w(T)| \leq 1$ for all $T \in [n]^k$. By Theorem 3.2.1, there is an algorithm certifying that

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

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

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

3.2.2 Quasirandomness and strong refutation of any k-CSP

Next, we will use the algorithm of Theorem 3.2.1 to certify that an instance of CSP(P) is quasirandom. This will immediately give us a strong refutation algorithm.

In order to certify quasirandomness, Lemma 2.1.13 implies that it suffices to certify each Fourier coefficient of $D_{\mathcal{I},x}$ has small magnitude.

Lemma 3.2.3. *Let $\emptyset \neq S \subseteq [k]$ with $|S| = s$. There is an algorithm that, with high probability, certifies that*

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

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

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

Lemma 3.2.4. *Let $S \subseteq [k]$ with $|S| = s > 0$. Let $\tau \in \mathbb{N}$ and let $\{w_U(i)\}_{U \in [n]^s, i \in [\tau]}$ be independent random variables satisfying conditions (3.6), (3.7), and (3.8) for some $p \geq \frac{1}{\tau n^{s/2}}$. Then there is an algorithm that certifies with high probability that*

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

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

The proof is straightforward and we defer it to Section 3.2.4.

Proof of Lemma 3.2.3. Without loss of generality, assume $1 \in S$. Applying definitions, we see that

$$\widehat{D_{\mathcal{I},x}}(S) = \mathop{\mathbf{E}}_{z\sim\mathcal{D}_{\mathcal{I},x}}\left[z^S\right] = \frac{1}{m}\sum_{U\in[n]^s}\sum_{\substack{T\in[n]^k\\ T_S=U}}\sum_{c\in\{\pm1\}^k} 1_{\{(T,c)\in\mathcal{I}\}}c^S x^U = \frac{1}{m}\sum_{U\in[n]^s} x^U \sum_{\substack{T\in[n]^k\\ T_S=U}}\sum_{c'\in\{\pm1\}^{k-1}} w_S(T,c'). \tag{3.9}$$

where we define $w_S(T,c') = 1_{\{(T,(1,c'))\in\mathcal{I}\}}(c')^{S\setminus\{1\}} - 1_{\{(T,(-1,c'))\in\mathcal{I}\}}(c')^{S\setminus\{1\}}$ and recall that T_S is the projection of T onto the the coordinates in S. It is clear that $\mathbf{E}[w_S(T,c')] = 0$, $\mathbf{Pr}[w_S(T,c') \neq 0] \leq p$, and $|w_S(T,c')| \leq 1$. There are $\tau = 2^{k-1}n^{k-s}$ terms in each sum of $w_S(T,c')$'s and we can apply Lemma 3.2.4. When $s = 2$, we plug in these values and see that we can certify that $\widehat{D_{\mathcal{I},x}}(S) \leq \frac{2^{O(s)}n^{s/4}\log^{5/2} n}{\overline{m}^{1/2}}$. When $s = 1$, $\overline{m} \geq n$ implies that $\tau p \geq \frac{1}{2}$ and we can certify that $\widehat{D_{\mathcal{I},x}}(S) \leq \frac{2^{O(s)}\sqrt{n}\log n}{\overline{m}^{1/2}}$. The lower bound can be proved in exactly the same way by considering the random variables $-w_S(T,c')$. □

The existence of an algorithm for certifying quasirandomness follows from Lemmas 2.1.13 and 3.2.3.

Theorem 3.2.5. *There is an efficient algorithm that certifies that an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P) *is γ-quasirandom with high probability when $\overline{m} \geq \frac{2^{O(k)}n^{k/2}\log^5 n}{\gamma^2}$.*

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

Theorem 3.2.6. *There is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P)*, certifies that* $\mathrm{Opt}(\mathcal{I}) \leq \mathbf{E}[P] + \gamma$ *with high probability when $\overline{m} \geq \frac{2^{O(k)}n^{k/2}\log^5 n}{\gamma^2}$.*

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

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

Definition 3.2.7. An instance $\mathcal{I}$ of CSP(P) is (ϵ,t)-quasirandom if $\mathcal{D}_{\mathcal{I},x}$ is (ϵ,t)-wise uniform for every $x \in \{-1,1\}^n$.

Fact 2.1.6 shows that random instances with $\widetilde{O}(n)$ constraints are $(o(1),t)$-quasirandom for all $t \leq k$ with high probability. Lemma 3.2.3 directly implies that we can certify (ϵ,t)-quasirandomness when $m \geq \widetilde{O}(n^{t/2})$.

Theorem 3.2.8. *There is an efficient algorithm that certifies that an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P) *is (γ,t)-quasirandom with high probability when $\overline{m} \geq \frac{2^{O(t)}n^{t/2}\log^5 n}{\gamma^2}$ and $t \geq 2$.*

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

Theorem 3.2.9. *Let P be δ-far from t-wise supporting. There is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P)*, certifies that* $\mathrm{Opt}(\mathcal{I}) \leq 1-\delta+\gamma$ *w.h.. when $\overline{m} \geq \frac{k^{O(t)}n^{t/2}\log^5 n}{\gamma^2}$ and $t \geq 2$.*

We give two proofs of this theorem. In Proof 1, the theorem follows directly from certification of (γ,t)-quasirandomness and Lemma 2.1.13.

Proof 1. Run the algorithm of Theorem 3.2.8 to certify that $\mathcal{I}$ is $(\gamma/k^t, t)$-quasirandom with high probability. By definition, we have certified that $\mathcal{D}_{\mathcal{I},x}$ is $(\gamma/k^t, t)$-wise uniform for all $x \in \{-1,1\}^n$,. Lemma 2.1.13 then implies that for all x there exists a t-wise uniform distribution $\mathcal{D}'_x$ such that $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x}, \mathcal{D}'_x) \leq \gamma$. Now define $\mathcal{D}_{\mathrm{sat}}$ to be an arbitrary distribution over satisfying assignments to P. Since P is δ-far from being t-wise supporting, we know that $d_{\mathrm{TV}}(\mathcal{D}', \mathcal{D}_{\mathrm{sat}}) \geq \delta$ for any t-wise uniform distribution $\mathcal{D}'$. The triangle inequality then implies that $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x}, \mathcal{D}_{\mathrm{sat}}) \geq \delta - \gamma$ for all $x \in \{-1,1\}^n$ and the theorem follows. □

Proof 2 gives a slightly weaker version of Theorem 3.2.9, requiring the stronger assumption that $\overline{m} \geq \frac{2^{O(k)} n^{t/2} \log^5 n}{\gamma^2}$. It is based on the dual polynomial characterization of being δ-far from t-wise supporting. While perhaps less intuitive than Proof 1, Proof 2 is more direct. It only uses the XOR refutation algorithm and bypasses [AGM03]'s connection between (ϵ, t)-wise uniformity and ϵ-closeness to a t-wise uniform distribution. We were able to convert Proof 2 into an SOS proof (see Section 3.4.3), but we did not see how to translate Proof 1 into an SOS version. Proof 2 requires Plancherel's Theorem, a fundamental result in Fourier analysis.

Theorem 3.2.10 (Plancherel's Theorem)**.** *For any $f, g : \{-1,1\}^k \to \mathbb{R}$,*

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

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

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

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

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

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

where the second equality follows from Plancherel's Theorem. Since $Q \geq -1$ and $\mathbf{E}[Q] = 0$, $Q \leq 2^k$ and hence $|\widehat{Q}(S)| \leq 2^k$ for all S. To finish the proof, we apply Lemma 3.2.3 to certify that $\left|\widehat{D_{\mathcal{I},x}}(S)\right| \leq \frac{\gamma}{2^{2k}}$ for all S. □

With Corollary 3.1.10, Theorem 3.2.9 implies that we can $\Omega_k(1)$-refute instances of $\mathrm{CSP}(P)$ with $\widetilde{O}_k(n^{t/2})$ constraints when P is not t-wise supporting.

Corollary 3.2.11. *Let P be a predicate that does not support any t-wise uniform distribution. Then there is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq 1 - 2^{-\widetilde{O}(k^t)}$ with high probability when $\overline{m} \geq 2^{\widetilde{O}(k^t)} n^{t/2} \log^5 n$ and $t \geq 2$.*

3.2.4 Proof of Lemma 3.2.4

Recall the statement of the lemma.

Lemma 3.2.4. Let $S \subseteq [k]$ with $|S| = s > 0$. Let $\tau \in \mathbb{N}$ and let $\{w_U(i)\}_{U \in [n]^s, i \in [\tau]}$ be independent random variables satisfying conditions (3.6), (3.7), and (3.8) for some $p \geq \frac{1}{\tau n^{s/2}}$. Then there is an algorithm that certifies with high probability that

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

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

The proof uses Bernstein's Inequality.

Theorem 3.2.12 (Bernstein's Inequality)**.** *Let $X_1, \ldots, X_M$ be independent 0-mean random variables such that $|X_i| \leq B$. Then, for $a > 0$,*

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

Proof of Lemma 3.2.4. First, we define

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

Observe that the v_U's are independent and that each one is the sum of τ mean-0, i.i.d. random variables with magnitude at most 1. Noting that $\sum_{i=1}^{\tau} \mathbf{E}[w_U(i)^2] \leq \tau p$, we can use Bernstein's Inequality to show that the $|v_U|$'s are not too big with high probability. If $s \geq 2$, Theorem 3.2.1 then implies that the desired algorithm exists. If $s = 1$, we are simply bounding a linear function over ± 1 variables. We consider two cases: Small p and large p.

Case 1: $p \leq \frac{1}{4\tau}$**.** Choosing $a = 2s \log n$ in Bernstein's Inequality, we see that $\mathbf{Pr}[|v_U| \geq 2s \log n] \leq n^{-2s}$. A union bound over all U then implies that $\mathbf{Pr}[\text{any } |v_U| > 2s \log n] \leq n^{-s}$. If $s \geq 2$, we observe that $\mathbf{Pr}[v_U \neq 0] \leq \tau p$, scale the v_U's down by $2s \log n$, and apply Theorem 3.2.1 to get the stated result. If $s = 1$, we obtain the second bound by observing that

$$\sum_{i \in [n]} v_i x_i \leq \sum_{i \in [n]} |v_i| \leq 2n \log n. \tag{3.10}$$

Case 2: $p > \frac{1}{4\tau}$**.** We set $a = 4s\sqrt{\tau p} \log n$ and get that $\mathbf{Pr}[\text{any } |v_U| > 4s\sqrt{\tau p} \log n] \leq n^{-s}$ as above. If $s \geq 2$, we can then divide the v_U's by $4s\sqrt{\tau p} \log n$ and apply Theorem 3.2.1. If $s = 1$, we get a bound of $4\sqrt{\tau p} \cdot n \log n$ in the same way as (3.10). □

3.3 Hardness of learning implications

Recent work by Daniely et al. [DLSS14] reduces the problem of refuting specific instances of CSP(P) to the problem of improperly learning certain hypothesis classes in the Probably Approximately Correct (PAC) model [Val84]. In this model, the learner is given m labeled training examples $(x_1, \ell(x_1)), \ldots, (x_m, \ell(x_m))$, where each $x_i \in \{-1, 1\}^n$, each $\ell(x_i) \in \{0, 1\}$, and the examples are drawn from some unknown distribution $\mathcal{D}$ on $\{-1, 1\}^n \times \{0, 1\}$. For some hypothesis class $\mathcal{H} \subseteq \{0, 1\}^{\{-1,1\}^n}$ we say that $\mathcal{D}$ can be *realized* by $\mathcal{H}$ if there exists some $h \in \mathcal{H}$ such that

$\mathbf{Pr}_{(x,\ell(x))\sim\mathcal{D}}[h(x) \neq \ell(x)] = 0$. In improper PAC learning, on an input of m training examples drawn from $\mathcal{D}$ such that $\mathcal{D}$ can be realized by some $h \in \mathcal{H}$, and an error parameter ϵ, the algorithm outputs some hypothesis function $f_h : \{-1,1\}^n \to \{0,1\}$ (not necessarily in $\mathcal{H}$) such that $\mathbf{Pr}_{(x,\ell(x))\sim\mathcal{D}}[f_h(x) \neq \ell(x)] \leq \epsilon$. In improper *agnostic* PAC learning, the assumption that $\mathcal{D}$ can be realized by some $h \in \mathcal{H}$ is removed and the algorithm must output a hypothesis that performs almost as well as the best hypothesis in $\mathcal{H}$. More formally, the hypothesis f_h must satisfy the following: $\mathbf{Pr}_{(x,\ell(x))\sim\mathcal{D}}[f_h(x) \neq \ell(x)] \leq \min_{h\in\mathcal{H}} \mathbf{Pr}_{(x,\ell(x))\sim\mathcal{D}}[h(x) \neq \ell(x)] + \epsilon$. In improper approximate agnostic PAC learning, the learner is also given an approximation factor $a \geq 1$ and must output a hypothesis f_h such that $\mathbf{Pr}_{(x,\ell(x))\sim\mathcal{D}}[f_h(x) \neq \ell(x)] \leq a \cdot \min_{h\in\mathcal{H}} \mathbf{Pr}_{(x,\ell(x))\sim\mathcal{D}}[h(x) \neq \ell(x)] + \epsilon$.

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

3.3.1 Hardness assumptions

The hardness assumptions made in [DLSS14] are the same as those presented in Section 3.1.1, except for a few minor differences. First, their model fixes the number of constraints rather than the probability with which each constraint is included in the instance. It is well-known that results in one model easily translate to the other. We include a proof in Section 3.8 for completeness. Additionally, SRCSP Assumptions 1 and 2 purport hardness of distinguishing random instances of CSP(P) from satisfiable instances, even when the algorithm is allowed to err with probability $\frac{1}{4}$ over its internal coins. The algorithms presented in the preceding sections never err on satisfiable instances; further, they only fail to certify random instances with probability $o(1)$. As a result, our refutation algorithms also falsify weaker versions of both SRCSP Assumptions, wherein the allowed probability of error is both lower and one-sided. For each predicate presented in [DLSS14], we falsify the appropriate SRCSP assumption using the following approach. For each predicate P and corresponding $\delta > 0$, we define a degree-t polynomial that δ-separates P. Using the refutation techniques presented in the preceding sections, we deduce that $\widetilde{O}(n^{t/2})$ constraints are sufficient to distinguish random instances of CSP(P) from those that are satisfiable (or have value at least α). In order to simplify the presentation, we begin with simpler versions of the polynomials and then scale them to attain the appropriate values of δ. The following lemma will be of use for this scaling.

Lemma 3.3.1. *For predicate $P : \{-1,1\}^k \to \{0,1\}$, let $Q : \{-1,1\}^k \to \mathbb{R}$ be an unbiased multilinear polynomial of degree t such that there exist $\theta_1 > 0, \theta_0 < 0$ not dependent on z for which the following holds: $Q(z) \geq \theta_1$ for all $z \in P^{-1}(1)$ and $Q(z) \geq \theta_0$ for all $z \in \{-1,1\}^k$. Then there exists a degree-t polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that $\frac{\theta_1}{\theta_1-\theta_0}$-separates P.*

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

We now demonstrate that the above can be applied to the predicates suggested in [DLSS14] by defining separating polynomials and applying Theorem 3.2.9

3.3.2 Huang's predicate and hardness of learning DNF formulas

In order to obtain hardness of improperly learning DNF formulas with $\omega(1)$ terms, Daniely et al. use the following predicate, introduced by Huang [Hua13]. Huang showed that it is hereditarily approximation resistant; Daniely et al. also observed that its 0-variability is $\Omega(k^{1/3})$ [DLSS14].

Definition 3.3.2. Let $k = \kappa + \binom{\kappa}{3}$ for some integer $\kappa \geq 3$. For $z \in \{-1,1\}^k$, index z as follows. Label the first κ bits of z as $z_1, \ldots, z_\kappa$. The remaining $\binom{\kappa}{3}$ bits are indexed by unordered triples of integers between 1 and κ. Each $T \subseteq [\kappa]$ with $|T| = 3$ is associated with a distinct bit of the remaining $\binom{\kappa}{3}$ bits, which is indexed by z_T. We say that z *strongly satisfies* the Huang predicate iff for every $T = \{z_i, z_j, z_\ell\}$ such that z_i, z_j, z_ℓ are distinct elements of $[\kappa]$, $z_i z_j z_\ell = z_{\{i,j,\ell\}}$. Additionally, we say that z *satisfies* the Huang predicate iff there exists some $z' \in \{-1,1\}^k$ such that z has Hamming distance at most κ from z' and z' strongly satisfies the Huang predicate. Define $H_\kappa : \{-1,1\}^k \to \{0,1\}$ as follows: $H_\kappa(z) = 1$ if z satisfies the Huang predicate and $H_\kappa(z) = 0$ otherwise.

Daniely et al. reduce the problem of distinguishing between random instances of $\mathrm{CSP}(H_\kappa)$ with $2n^d$ constraints and satisfiable instances to the problem of improperly PAC learning the class of DNF formulas with $\omega(1)$ terms on a sample of $O(n^d)$ training examples with error $\epsilon = 1/5$ with probability at least $\frac{3}{4}$. Here we show that there exists a polynomial time algorithm that refutes random instances of $\mathrm{CSP}(H_\kappa)$ by demonstrating that H_k does not support a 4-wise uniform distribution and applying Theorem 3.2.9.

Theorem 3.3.3. *Assume $\kappa \geq 9$. There exists a degree-4 polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that $\frac{1}{8}$-separates H_κ. Consequently, H_κ is $\frac{1}{8}$-far from supporting a 4-wise uniform distribution.*

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

$$\begin{aligned}\zeta(i_1,i_2,i_3,i_4,i_5,i_6,z) = z_{126}z_{134}z_{235}z_{456} + z_{256}z_{146}z_{345}z_{123} + z_{136}z_{236}z_{145}z_{245} \\ + z_{124}z_{234}z_{356}z_{156} + z_{125}z_{135}z_{346}z_{246}.\end{aligned} \tag{3.11}$$

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

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

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

Observe that Q does not depend on any of $z_{\{1\}}, \ldots z_{\{\kappa\}}$. By construction, Q contains no constant term, so $\widehat{Q}(\emptyset) = 0$. Clearly $Q(z) \geq -5$ for all z because (3.11) is always at least -5.

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

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

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

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

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

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

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

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

$$Q(z) \geq 5 - \frac{240\kappa}{|\mathcal{Z}_6|}(\kappa - 3)(\kappa - 4)(\kappa - 5) = 5 - \frac{240}{(\kappa - 1)(\kappa - 2)}. \tag{3.12}$$

For $\kappa \geq 9$, (3.12) is at least $5 - \frac{30}{7}$. Applying Lemma 3.3.1, there exists $\mathcal{Q} : \{-1, 1\}^k \to \mathbb{R}$ that $\frac{1}{8}$-separates H_κ. □

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

Corollary 3.3.4. *For sufficiently large n and $k \geq 93$, there exists an efficient algorithm that refutes random instances of $\mathrm{CSP}(H_\kappa)$ with $\widetilde{O}(n^2)$ constraints with high probability. This falsifies Assumption 3.1.1 in the case of the Huang predicate.*

Remark 3.3.5. If we instead choose to scale Q by a factor of $\frac{1}{5} \cdot \frac{\kappa^2 - 3\kappa + 2}{2\kappa^2 - 6\kappa - 44}$ rather than substituting $\kappa = 9$ into (3.12), we can achieve a better separation of $\delta = \frac{\kappa^2 - 3\kappa - 46}{2\kappa^2 - 6\kappa - 44}$. For $\kappa \geq 9$, this expression is strictly increasing and it approaches $\frac{1}{2}$ as κ grows.

3.3.3 Hamming weight predicates

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

Definition 3.3.6. For $z \in \{-1, 1\}^k$ where $z = z_1, \ldots, z_k$, define $S_z = \sum_{i=1}^{k} z_i$ and call S_z the Hamming weight of z.

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

Definition 3.3.7. For all odd k and any $\theta \in \{-k, -k + 2, \ldots, k - 2, k\}$, define the predicate $\mathrm{Thr}_k^\theta : \{-1, 1\}^k \to \{0, 1\}$ as follows:

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

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

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

Lemma 3.3.8. *Let $Q : \{-1,1\}^k \to \mathbb{R}$ be of the following form for some $a, b, c, d \in \mathbb{R}$:*

$$Q(z) = a \sum_{\substack{T \subseteq [n] \\ |T|=1}} z^T + b \sum_{\substack{T \subseteq [n] \\ |T|=2}} z^T + c \sum_{\substack{T \subseteq [n] \\ |T|=3}} z^T + d \sum_{\substack{T \subseteq [n] \\ |T|=4}} z^T. \tag{3.13}$$

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

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

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

Proof. We can write (3.13) as follows:

$$Q(z) = a\mathscr{K}_1\left(\frac{k-S_z}{2}; k\right) + b\mathscr{K}_2\left(\frac{k-S_z}{2}; k\right) + c\mathscr{K}_3\left(\frac{k-S_z}{2}; k\right) + d\mathscr{K}_4\left(\frac{k-S_z}{2}; k\right), \tag{3.14}$$

where $\mathscr{K}_i(\nu; k) = \sum_{j=0}^{i} (-1)^j \binom{\nu}{i} \binom{k-\nu}{i-j}$ denotes the Krawtchouk polynomial of degree i [Kra29, KL96]. Substituting $\nu = \frac{k-S_z}{2}$, yields the following expressions. In [KL96] the first three expressions are given explicitly and the fourth can be easily obtained by applying their recursive formula.

$$\mathscr{K}_1\left(\frac{k-S_z}{2}; k\right) = S_z, \qquad \mathscr{K}_3\left(\frac{k-S_z}{2}; k\right) = \frac{S_z^3 - (3k-2)S_z}{6},$$
$$\mathscr{K}_2\left(\frac{k-S_z}{2}; k\right) = \frac{S_z^2 - k}{2}, \qquad \mathscr{K}_4\left(\frac{k-S_z}{2}; k\right) = \frac{S_z^4 + (8-6k)S_z^2 + 3k^2 - 6k}{24}.$$

Finally, substituting these expressions into (3.14) and by some algebra,

$$Q(z) = \frac{d}{24} S_z^4 + \frac{c}{6} S_z^3 + \left(\frac{b}{2} + \frac{d}{3} - \frac{dk}{4}\right) S_z^2 + \left(a + \frac{c}{6} \cdot (-3k+2)\right) S_z - \frac{bk}{2} + \frac{dk}{24}(3k-6). \tag{3.15}$$

□

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

Almost-Majority and hardness of learning intersections of halfspaces

Definition 3.3.9. Daniely et al. define the following predicate in order to show hardness of improperly learning intersections of four halfspaces.

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

The reduction relies on the assumption that for all $d > 0$, it is hard to distinguish random instances of $\mathrm{CSP}(I_{8k})$ with n^d constraints from satisfiable instances. Because the input variables to each instance of Thr_k^{-1} above are disjoint, it is sufficient to show that each of the first four groups of k variables cannot support a 3-wise uniform distribution and consequently neither can I_{8k}; therefore, from Theorem 3.2.9 we deduce that there exists an efficient algorithm that refutes random instances of $\mathrm{CSP}(I_{8k})$ with $\widetilde{O}(n^{3/2})$ constraints with high probability. Daniely et al. define a pairwise uniform distribution supported on I_{8k} as well as a pairwise uniform distribution supported on Thr_k^{-1}, so $t = 3$ is optimal.

Theorem 3.3.10. *Assume $k \geq 5$ and k is odd. There exist $\delta = \delta(k) > 0$ where δ is $\Omega(k^{-4})$ and a degree-3 multilinear polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that δ-separates Thr_k^{-1}. Consequently, Thr_k^{-1} does not support a 3-wise uniform distribution.*

Proof. Let

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

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

$$\begin{aligned} Q_u(s) &= \frac{1+k}{6} s^3 + \left(\frac{1-k}{2}\right) s^2 + \left(k^2 - k - 1 + \frac{1+k}{6} \cdot (-3k+2)\right) s - \frac{(1-k)k}{2} \\ &= \frac{1+k}{6} s^3 + \left(\frac{1-k}{2}\right) s^2 + \left(\frac{3k^2 - 7k - 4}{6}\right) s - \frac{(1-k)k}{2}. \end{aligned}$$

Then by Lemma 3.3.8, for all $z \in \{-1,1\}^k$, $Q(z) = Q_u(S_z)$. It therefore suffices to lower bound $Q_u(s)$ both when $s \geq -1$ and for all $s \in [-k, k]$.

First we show that Q_u is monotonically increasing in s.

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

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

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

$$\begin{aligned} Q_u(-k) &= \frac{-k-1}{6} k^3 + \left(\frac{1-k}{2}\right) k^2 - \left(\frac{3k^2 - 7k - 4}{6}\right) k - \frac{(1-k)k}{2} \\ &= -\tfrac{1}{6}\left[k^4 + 7k^3 - 13k^2 - k\right], \qquad (3.16) \\ &= -\tfrac{1}{6}\left[k(k-2)(k^2 + 9k + 5) + 9k\right], \qquad (3.17) \end{aligned}$$

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

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

By applying Lemma 3.3.1, there exists an unbiased multilinear polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ of degree 3 that $\frac{6}{k^4+7k^3-13k^2-k+6}$-separates Thr_k^{-1}. □

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

Corollary 3.3.11. *For odd $k \geq 5$ and sufficiently large n, there exists an efficient algorithm that distinguishes between random instances of CSP(I_{8k}) with $\widetilde{O}(n^{3/2})$ constraints and satisfiable instances with high probability.*

Remark 3.3.12. Thr_3^{-1} is the same as 3-OR and Thr_5^{-1} is the same as is the same as 2-out-of-5-SAT, so this approach can be used to $\Omega_k(1)$-refute 3-SAT instances and 2-out-of-5-SAT instances with $\widetilde{O}_k(n^{3/2})$ constraints, which improves upon the $O(n^{3/2+\epsilon})$ constraints required for refutation of 2-out-of-5-SAT in [GJ02, GJ03].

3.3.4 Majority and hardness of approximately agnostically learning halfspaces

Daniely et al. show that approximate agnostic improper learning of halfspaces is hard for all approximation factors $\phi \geq 1$ based on the assumption that for all $d > 0$ and for sufficiently large odd k, it is hard to distinguish between random instances of CSP(Thr_k^1) with n^d constraints and instances with value at least $1 - \frac{1}{10\phi}$. This is based on the fact that $\max_{\mathcal{D}} \mathbf{E}_{z\sim\mathcal{D}}[\mathrm{Thr}_1(z)] = 1 - \frac{1}{k+1}$, where $\mathcal{D}$ is a pairwise independent distribution on $\{-1,1\}^k$, and applying SRCSP Assumption 2. Here we show that for odd $k \geq 25$, Thr_k^1 is 0.1-far from supporting a 4-wise uniform distribution. The value 0.1 is not sharp, but is chosen as a compromise between a reasonably large value and a reasonably simple proof.

Theorem 3.3.13. *There exists a degree-4 multilinear polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that 0.1-separates Thr_k^1 for all odd $k \geq 25$.*

Proof. Let

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

and let

$$\begin{aligned} Q_u(s) &= \frac{1}{18k^2}s^4 - \frac{5}{54k^{3/2}}s^3 + \left(-\frac{1}{3k} + \frac{4}{9k^2}\right)s^2 + \left(\frac{31}{54\sqrt{k}} - \frac{5}{27k^{3/2}}\right)s + \frac{1}{6} - \frac{1}{3k} \\ &= \frac{1}{54}\left[\frac{3}{k^2}s^4 - \frac{5}{k^{3/2}}s^3 + \left(-\frac{18}{k} + \frac{24}{k^2}\right)s^2 + \left(\frac{31}{\sqrt{k}} - \frac{10}{k^{3/2}}\right)s + 9 - \frac{18}{k}\right]. \end{aligned} \tag{3.18}$$

Then by Lemma 3.3.8, for all $z \in \{-1,1\}^k$, $Q(z) = Q_u(S_z)$. To simplify Q, let $\sigma = sk^{-1/2}$. Then we can rewrite (3.18) as follows:

$$Q_u(s) = \tfrac{1}{54}\left[3\sigma^4 - 5\sigma^3 + \left(-18 + \tfrac{24}{k}\right)\sigma^2 + \left(31 - \tfrac{10}{k}\right)\sigma + 9 - \tfrac{18}{k}\right]. \tag{3.19}$$

First we lower-bound $Q_u(s)$ for all $\sigma \in \mathbb{R}$ using the following expression, which is equivalent to (3.19) by some algebra.

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

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

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

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

Applying Lemma 3.3.1, there exists $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ such that $\mathcal{Q}$ has degree 4 and $\mathcal{Q}$ $\frac{9}{73}$-separates Thr_k^1. □

Corollary 3.3.14. *For sufficiently large n and k, there exists an efficient algorithm that distinguishes between random instances of CSP(Thr_k^1) with $\widetilde{O}(n^2)$ constraints and instances with value at least 0.9 with high probability.*

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

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

We also consider the predicate $\mathrm{Thr}_k^{-\frac{1}{2}\sqrt{k}}$. While it is not used in [DLSS14], we show that it does not support a 4-wise uniform distribution in the interest of obtaining a tighter bound for the Hamming weight above which an unbiased, symmetric predicate is not 4-wise supporting. The threshold of $-\frac{1}{2}\sqrt{k}$ is particularly interesting in that it asymptotically matches the threshold θ below which Thr_k^θ is 4-wise supporting.

Theorem 3.3.15. *Assume $k \geq 99$ and k is odd. Then there exists a degree-4 polynomial $\mathcal{Q} : \{-1,1\}^k \to \mathbb{R}$ that $\frac{1}{225}$-separates $\mathrm{Thr}_k^{-\frac{1}{2}\sqrt{k}}$. Consequently, $\mathrm{Thr}_k^{-\frac{1}{2}\sqrt{k}}$ is $\frac{1}{255}$-far from 4-wise supporting.*

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

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

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

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

$$Q_u(s) = \tfrac{1}{3}\sigma^4 + \tfrac{1}{3}\sigma^3 - \tfrac{7}{4}\sigma^2 + \tfrac{1}{2}\sigma + \tfrac{3}{4} + \tfrac{1}{k}\left(\tfrac{8}{3}\sigma^2 + \tfrac{2}{3}\sigma - 2\right). \tag{3.20}$$

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

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

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

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

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

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

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

Applying Lemma 3.3.1, $T_k^{-\frac{1}{2}\sqrt{k}}$ is $\frac{1}{255}$-far from supporting a 4-wise uniform distribution. □

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

Claim 3.3.16. *Assume $k = 2^m - 1$ for some integer $m \geq 3$. Then there exists a 4-wise uniform distribution supported only on $z \in \{-1, 1\}^k$ such that $S_z \geq 1 - 2\sqrt{k+1}$.*

Proof. Let $\mathcal{C}$ be a binary BCH code of length k with designed distance $2\iota + 1$ and let $\mathcal{C}^\perp$ be its dual. Then the uniform distribution on the codewords of $\mathcal{C}$ is 2ι-wise uniform [ABI86, MS77]; see also [AS04, Chapter 16.2].

Let $c = c_1 \dots c_k$ be a codeword of $\mathcal{C}^\perp$, where each $c_i \in \{-1, 1\}$. The Carlitz-Uchiyama bound [MS77, page 280] states that for all $c \in \mathcal{C}^\perp$,

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

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

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

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

Remark 3.3.17. In order to construct a 4-wise uniform distribution for any value of k, one could simply express k as a sum of powers of 2, construct separate distributions on disjoint variables as described above for each power of 2 (down to the minimum length for which we can achieve distance at least 5, after which point we use the uniform distribution, and obtain a 4-wise uniform distribution. The total Hamming weight of a vector supported by this distribution would then be at least $-O(\sqrt{k})$.

3.4 SOS refutation proofs

3.4.1 SOS certification of quasirandomness

All of our SOS results rely on the following theorem, which is the SOS version of Theorem 3.2.1.

Theorem 3.4.1. *For $k \geq 2$ and $p \geq n^{-k/2}$, let $\{w(T)\}_{T\in[n]^k}$ be independent random variables such that for each $T \in [n]^k$,*

$$\mathbf{E}[w(T)] = 0 \tag{3.21}$$
$$\mathbf{Pr}[w(T) \neq 0] \leq p \tag{3.22}$$
$$|w(T)| \leq 1. \tag{3.23}$$

Then, with high probability,

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

This theorem was essentially proven by Barak and Moitra [BM16]. We give a proof in Appendix 3.5.3. We first use this theorem to show that an SOS version of Lemma 3.2.4 holds.

Lemma 3.4.2. *Let $S \subseteq [k]$ with $|S| = s > 0$. Let $\tau \in \mathbb{N}$ and let $\{w_U(i)\}_{U\in[n]^s, i\in[\tau]}$ be independent random variables satisfying conditions (3.6), (3.7), and (3.8) for some $p \geq \frac{1}{\tau n^{s/2}}$. Then, with high probability,*

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

Proof. We sketch the differences from the proof of Lemma 3.2.4 given in Section 3.2.4. For $s \geq 2$, the lemma follows by using Theorem 3.4.1 instead of Theorem 3.2.1. If $s = 1$, it suffices to show that

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

for any v since summing over all i as in (3.10) finishes the proof. If $v_i \geq 0$, observe that

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

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

The lemma implies an SOS version of Lemma 3.2.3. To make this precise, we define a specific polynomial representation of $\widehat{D_{\mathcal{I},x}}(S)$:

$$\widehat{D_{\mathcal{I},x}}(S)^{\text{poly}} = \frac{1}{m}\sum_{T\in[n]^k}\sum_{c\in\{\pm1\}^k} 1_{\{(T,c)\in\mathcal{I}\}}c^S x_T^S,$$

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

We can show these polynomials are not too large.

Lemma 3.4.3. *Let $\emptyset \neq S \subseteq [k]$ with $|S| = s$. Then*

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

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

Proof. The proof is essentially the same as that of Lemma 3.2.3. The expression we bound in that proof is exactly $\widehat{D_{\mathcal{I},x}}(S)^{\text{poly}}$. We use Lemma 3.4.2 instead of Lemma 3.2.4 to show that this can be done in degree-$2s$ SOS. □

Based on Lemma 2.1.13, we will think of Lemma 3.4.3 as giving an SOS proof of quasirandomness. Below, we use it to prove SOS versions of Theorems 3.2.6 and 3.2.9.

3.4.2 Strong refutation of any k-CSP

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

$$\text{Val}_{\mathcal{I}}^{\text{poly}}(x) = \frac{1}{m} \sum_{T\in[n]^k} \sum_{c\in\{\pm1\}^k} 1_{\{(T,c)\in\mathcal{I}\}} \left(\sum_{S\subseteq[k]} \widehat{P}(S) c^S x_T^S \right),$$

where x_T^S is as above.

We can then give an SOS proof strongly refuting $\text{CSP}(P)$.

Theorem 3.4.4. *Given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\text{CSP}(P)$,*

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

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

Proof. By rearranging terms, we see that

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

Note that this is just Plancherel's Theorem in SOS. The theorem then follows from Lemma 3.4.3 and the observation that $\sum_{S\subseteq[k]} |\widehat{P}(S)| \leq 2^{O(k)}$. □

3.4.3 $\Omega(1)$-refutation of non-t-wise supporting CSPs

Theorem 3.4.5. *Let P be δ-far from being t-wise supporting. Given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of $\text{CSP}(P)$,*

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

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

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

Claim 3.4.6. *Let $f : \{-1,1\}^k \to \mathbb{R}$ such that $f(z) \geq 0$ for all $z \in \{-1,1\}^k$ and let f^{poly} be the unique multilinear polynomial such that $f(z) = f^{\mathrm{poly}}(z)$ for all $z \in \{-1,1\}^k$. Then*

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

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

Proof of Theorem 3.4.5. The proof is an SOS version of Proof 2 of Theorem 3.2.9 above. Claim 3.4.6 implies that for Q of degree at most t that δ-separates P,

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

Summing over all constraints, we get

$$\{x_i^2 = 1\}_{i\in[n]} \vdash_k \mathrm{Val}_{\mathcal{I}}^{\mathrm{poly}}(x) - (1-\delta) \leq \frac{1}{m} \sum_{T\in[n]^k} \sum_{c\in\{\pm1\}^k} 1_{\{(T,c)\in\mathcal{I}\}} \left(\sum_{S\subseteq[k]} \widehat{Q}(S) c^S x_T^S \right),$$

Rearranging terms as in the proof of Theorem 3.4.4, we see that the right hand side is equal to

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

Since Q has mean 0, $|Q| \leq 2^k$ and $\sum_{S\subseteq[k]} |\widehat{P}(S)| \leq 2^{O(k)}$. The theorem then follows from Lemma 3.4.3. □

With Corollary 3.1.10, the theorem implies that we can $\Omega_k(1)$-refute any CSP(P) in SOS when P is not t-wise supporting.

Corollary 3.4.7. *Let P be a predicate that does not support any t-wise uniform distribution. Given an instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$ of* CSP(P),

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

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

3.5 Proof of Theorem 3.2.1

We restate Theorem 3.2.1:

Theorem 3.2.1. For $k \geq 2$ and $p \geq n^{-k/2}$, let $\{w(T)\}_{T\in[n]^k}$ be independent random variables such that for each $T \in [n]^k$,

$$\mathbf{E}[w(T)] = 0 \tag{3.24}$$
$$\mathbf{Pr}[w(T) \neq 0] \leq p \tag{3.25}$$
$$|w(T)| \leq 1. \tag{3.26}$$

Then there is an efficient algorithm certifying that

$$\sum_{T\in[n]^k} w(T)x^T \le 2^{O(k)}\sqrt{p}n^{3k/4}\log^{3/2} n. \tag{3.27}$$

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

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

3.5.1 The even arity case

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

$$\sum_{T\in[n]^k} w(T)x^T = \sum_{T_1,T_2\in[n]^{k/2}} w(T_1, T_2)y_{T_1}y_{T_2}, \tag{3.28}$$

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

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

Theorem 3.5.1. [CW04, Theorem 1] *Let M be any $n \times n$ matrix with all diagonal elements 0. There exists an efficient randomized algorithm that finds $y \in \{-1, 1\}^n$ such that*

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

By Markov's Inequality, this statement holds with probability at least $1/2$. We can run the algorithm $O(\log n)$ times to get a high probability result. To apply Theorem 5.2.3, we separate out the diagonal terms of (3.28), rewriting it as

$$\sum_{T_1\neq T_2\in[n]^{k/2}} w(T_1, T_2)y_{T_1}y_{T_2} + \sum_{U\in[n]^{k/2}} w(U, U)y_U^2. \tag{3.29}$$

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

Claim 3.5.2. *With high probability, it holds that*

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

This follows from applying Bernstein's Inequality (Theorem 3.2.12) for fixed y and then taking a union bound over all $y \in \{-1,1\}^{n^{k/2}}$. Using the claim, we see that Theorem 5.2.3 allows us to certify that the value of the first term in (3.29) is at most $O(\sqrt{p}n^{3k/4}\log n)$.

We will use the next claim to bound the second term of (3.29).

Claim 3.5.3. *With high probability, it holds that*

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

Since the $|w_T| \leq 1$ and $\mathbf{Pr}[w_T \neq 0] \leq p$, the claim follows from the Chernoff Bound. The second term of (3.29) is upper bounded by $\sum_{U\in[n]^{k/2}} |w_{U,U}|$ and we can compute this quantity in polynomial time to certify that its value is at most $O(\sqrt{p}n^{3k/4})$.

Random matrix approach. Observe that (3.28) is $y^\top By$ for a matrix B indexed by $U \in [n]^{k/2}$ so that $B_{U_1,U_2} = w(U_1,U_2)$. Then $y^\top By \leq \|B\| \|y\|^2$. To certify that $y^\top By$ is small, we compute $\|B\|$. We need to show that $\|B\|$ is small with high probability. First, note that $\|B\|$ is equal to the norm of the $2n^{k/2} \times 2n^{k/2}$ symmetric matrix

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

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

Proposition 3.5.4. [Tao12, Proposition 2.3.13] *Let M be a random symmetric matrix $n \times n$ whose upper triangular entries M_{ij} with $i \geq j$ are independent random variables with mean 0, variance at most 1, and magnitude at most K. Then, with high probability,*

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

Let $\tilde{B}' = \frac{1}{\sqrt{p}}\tilde{B}$. The upper triangular entries of $\tilde{B}'$ are independent random variables with mean 0, variance at most 1, and magnitude at most $1/\sqrt{p}$. Applying Proposition 3.5.4 to $\tilde{B}'$ shows that $\|B\| = O\left(kn^{k/4}\sqrt{p}\log n \cdot \max\left\{1, \frac{1}{\sqrt{p}n^{k/4}}\right\}\right)$ with high probability. Since $\|y\|_\infty \leq 1$ by assumption, $\|y\|^2 \leq n^{k/2}$ and (3.28) is at most $O(k\sqrt{p}n^{3k/4}\log n)$ with high probability when $p \geq n^{-k/2}$.

3.5.2 The odd arity case

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

$$\sum_{T\in[n]^k} w(T)x^T \leq \sum_{i\in[n]} W_i \leq \sqrt{n}\sqrt{\sum_{i\in[n]} W_i^2}, \tag{3.30}$$

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

$$\begin{aligned}\sum_{i\in[n]} W_i^2 &= \sum_{T,U\in[n]^{k-1}} \sum_{i\in[n]} w(T,i)w(U,i)x^T x^U \\ &= \sum_{T_1',T_2',U_1',U_2'\in[n]^{\frac{k-1}{2}}} \sum_{i\in[n]} w(T_1',U_1',i)w(T_2',U_2',i)x^{(T_1',T_2')}x^{(U_1',U_2')}. \end{aligned} \tag{3.31}$$

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

$$A_{(i_1,i_2),(j_1,j_2)} = \begin{cases} \sum_{\ell\in[n]} w(i_1,j_1,\ell)w(i_2,j_2,\ell) & \text{if } (i_1,j_1)\neq(i_2,j_2) \\ 0 & \text{otherwise,} \end{cases} \tag{3.32}$$

where we have divided the indices of A into 2 blocks of $\frac{k-1}{2}$ coordinates each. Define $x^{\otimes k-1}\in\mathbb{R}^{[n]^{k-1}}$ so that $x^{\otimes k-1}(T) = x^T$. Then (5.2) is equal to

$$(x^{\otimes k-1})^\top A x^{\otimes k-1} + \sum_{T,U\in[n]^{\frac{k-1}{2}}} (x^T)^2(x^U)^2 \sum_{i\in[n]} w(T,U,i)^2. \tag{3.33}$$

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

Lemma 3.5.5. *Let $k\geq 3$ and $p\geq n^{-k/2}$. Let $\{w(T)\}_{T\in[n]^k}$ be indepedent random variables satisfying conditions (3.6), (3.7), and (3.8) above. Let A be defined as in (3.32). With high probability,*

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

We can therefore certify that the first term is $2^{O(k)} p n^{3k/2-1}\log^3 n$. We will prove the lemma in Section 3.5.4.

The second term of (3.33) is at most $\sum_{T\in[n]^k} w(T)^2$. We can easily compute this and the Chernoff Bound implies that its value is at most $pn^{3k/2-1}$ with high probability.

So far, with high probability we can certify that $\sum_{i\in[n]} W_i^2 = 2^{O(k)} p n^{3k/2-1}\log^3 n$. Plugging this bound into (3.30) concludes the proof.

Remark 3.5.6. It would have been more natural to have written $\sum_{i\in[n]} W_i^2 = (x^{\otimes k-1})^\top A' x^{\otimes k-1}$ for A' such that $A'_{T,U} = \sum_{i\in[n]} w(T,i)w(U,i)$. However, $\|A'\|$ could be too large because of the contribution of the second term in (3.33). We use the additional assumption that $\|x\|_\infty \leq 1$ to get around this issue.

Remark 3.5.7. We have defined A so that $A_{(i_1,i_2),(j_1,j_2)} = \sum_{\ell\in[n]} w(i_1,j_1,\ell)w(i_2,j_2,\ell)$, not $A_{i,j} = \sum_{\ell\in[n]} w(i,\ell)w(j,\ell)$. The reduces the correlation among entries $w(b,c)$ and $w(b,c')$ for $c\neq c'$. Intuitively, A looks more like a random matrix with independent entries, so we can bound its norm using the trace method. See the proof of Lemma 3.5.5 in Section 3.5.4.

3.5.3 An SOS version

In this section, we will prove the SOS version of Theorem 3.2.1.

Theorem 3.4.1. For $k \geq 2$ and $p \geq n^{-k/2}$, let $\{w(T)\}_{T\in[n]^k}$ be independent random variables such that for each $T \in [n]^k$,

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

Then, with high probability,

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

Rather than writing out the full proof, we will indicate the small changes required to convert the above proof of Theorem 3.2.1 into SOS form.

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

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

so

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

Odd arity. A couple of additional issues arise in the odd case. First of all, the square root in (3.30) is not easily expressed in SOS, so we instead prove the squared version

$$\left(\sum_{T\in[n]^k} w(T)x^T\right)^2 \leq 2^{O(k)}n^{3k/2}\log^3 n. \tag{3.34}$$

By a simple extension of [OZ13, Fact 3.3], (3.34) implies (3.27) in SOS :

Fact 3.5.8.

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

Proof.

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

Secondly, we do not know how to prove the Cauchy-Schwarz inequality (3.30) in SOS. However, O'Donnell and Zhou show that a very similar inequality can be proved in SOS [OZ13, Fact 3.8]:

Fact 3.5.9.

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

Using this fact instead of Cauchy-Schwarz to prove the squared version of (3.30), we can follow the argument above to show that

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

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

3.5.4 Proof of Lemma 3.5.5

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

Lemma 3.5.5. Let $k \geq 3$ and $p \geq n^{-k/2}$. Let $\{w(T)\}_{T\in[n]^k}$ be indepedent random variables satisfying conditions (3.6), (3.7), and (3.8) above. Let A be the $[n]^{k-1} \times [n]^{k-1}$ indexed by $[n]^{k-1}$ that is defined as follows:

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

Then with high probability,

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

The proof closely follows the arguments of [COGL04, Lemma 17] and [BM16, Section 4]. Both proofs use the trace method: To bound the norm of a symmetric random matrix M, it suffices to bound $\mathbf{E}[\mathrm{tr}(M^r)]$ for large r. For non-symmetric matrices, we can instead work with $MM^\top$. In our particular case, we have the following.

Claim 3.5.10. *If $\mathbf{E}[\mathrm{tr}((AA^\top)^r)] \leq n^{O(k)}2^{O(r)}r^{6r}p^{2r}n^{kr}$, then $\|A\| \leq 2^{O(k)}pn^{k/2}\log^3 n$ with high probability.*

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

Remark 3.5.11. We can get arbitrarily small $1/\mathrm{poly}(n)$ probability of failure: This proof shows that $\|A\| \leq K2^{O(k)}pn^{k/2}\log^3 n$ with probability at most $n^{-\log K}$.

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

Lemma 3.5.12. *Under the conditions of Lemma 3.5.5, $\mathbf{E}[\mathrm{tr}((AA^\top)^r)] \leq n^{O(k)}2^{O(r)}r^{6r}p^{2r}n^{kr}$ with high probability.*

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

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

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

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

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

$$(i_s,j_s) \neq (i_{s+1},j_{s+1}) \qquad \text{for } 1 \leq s \leq 2r-1 \tag{3.35}$$

$$(i_{s+2},j_s) \neq (i_{s+3},j_{s+1}) \qquad \text{for } 1 \leq s \leq 2r-3 \tag{3.36}$$

$$(i_1,j_{2r-1}) \neq (i_2,j_{2r}). \tag{3.37}$$

Let Ω be the set of all $(i_1,\ldots,i_{2r},j_1,\ldots,j_{2r}) \in ([n]^{\frac{k-1}{2}})^{4r}$ satisfying (3.35), (3.36), and (3.37). Then for $J \in \Omega$ and $L = (\ell_1,\ldots,\ell_{2r}) \in [n]^{2r}$, define

$$P_{J,L} = w_{i_1,j_1,\ell_1}w_{i_2,j_2,\ell_1}w_{i_3,j_1,\ell_2}w_{i_4,j_2,\ell_2}\cdots w_{i_{2r-1},j_{2r-1},\ell_{2r-1}}w_{i_{2r},j_{2r},\ell_{2r-1}}w_{i_1,j_{2r-1},\ell_{2r}}w_{i_2,j_{2r},\ell_{2r}}. \tag{3.38}$$

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

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

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

Claim 3.5.13. *The number of distinct $w_{i,j,\ell}$ factors in $P_{J,L}$ is at least $2|L|$.*

Proof. For each $\ell \in L$, (3.38) shows that $P_{J,L}$ contains a pair of the form $w_{i_s,j_s,\ell}w_{i_{s+1},j_{s+1},\ell}$ or $w_{i_{s+2}j_s,\ell}w_{i_{s+3},j_{s+1},\ell}$. Since $J \in \Omega$, we know that $(i_s, j_s) \neq (i_{s+1}, j_{s+1})$ or $(i_{s+2}, j_s) \neq (i_{s+3}, j_{s+1})$, so each of these pairs must have two distinct $w_{i,j,\ell}$ factors. We then have at least $2|L|$ distinct $w_{i,j,\ell}$ factors. □

Claim 3.5.14. *The number of distinct $w_{i,j,\ell}$ factors in $P_{J,L}$ is at least $|J| - 2$.*

Proof. Consider looking over the factors of $P_{J,L}$ from left to right in the order of (3.38) until we have seen all elements of J. The first pair $w_{i_1,j_1,\ell_1}w_{i_2,j_2,\ell_1}$ contains at most four previously-unseen elements of J. Every subsequent pair of factors $w_{i_s,j_s,\ell_s}w_{i_{s+1},j_{s+1}}$ or $w_{i_{s+2}j_s,\ell_{s+1}}w_{i_{s+3},j_{s+1},\ell_{s+1}}$ in $P_{J,L}$ shares two variables of J with its preceding pair. Each such pair can then contain at most two new elements of J. After seeing u $w_{i,j,\ell}$'s, we have therefore seen at most $4 + 2\left(\frac{u-2}{2}\right)$ distinct elements of J. To get all $|J|$ elements of J, we must have seen at least $|J| - 2$ $w_{i,j,\ell}$'s and these must be distinct. □

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

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

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

Claim 3.5.15. *If $|L| > r$, then $\mathbf{E}[P_{J,L}] = 0$.*

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

Assume for a contradiction that $|L| > r$ and every $w_{i,j,\ell}$ factor occurs at least twice. Since there are at least $2|L|$ distinct $w_{i,j,\ell}$'s, there must be at least $4|L| > 4r$ total $w_{i,j,\ell}$'s. However, looking at (3.38), $P_{J,L}$ has at most $4r$ $w_{i,j,\ell}$ factors. □

Claim 3.5.16. *If $|J| > 2r + 2$, then $\mathbf{E}[P_{J,L}] = 0$.*

This can be proved in exactly the same manner.

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

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

We bound each term of the sum.

Claim 3.5.17.

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

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

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

If $2b \leq a-2$,

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

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

To conclude, observe that

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

□

Remark 3.5.18. If we did not have conditions (3.35), (3.36), and (3.37), we would only have been able to show that $|L| \leq 2r$. This would have led to a weaker bound of $O(\sqrt{n})$.

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

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

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

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

Theorem 3.6.1. (Coja-Oghlan–Goerdt–Lanka [COGL07, Theorem 3]). *For $H \sim \mathcal{H}(n,p,3)$, there is a polynomial time algorithm certifying that $\alpha(H) < \epsilon n$ with high probability for any constant $\epsilon > 0$ when $\overline{m} > n^{3/2}\ln^6 n$ and $\overline{m} = o(n^2)$.*

Theorem 3.6.2. (Coja-Oghlan–Goerdt–Lanka [COGL07, implicit in Section 4]). *For $H \sim \mathcal{H}(n,p,4)$, there is a polynomial time algorithm certifying that $\alpha(H) < \epsilon n$ with high probability for any constant $\epsilon > 0$ when $\overline{m} \geq O\left(\frac{n^2}{\epsilon^4}\right)$.*

Theorem 3.6.3. (Coja-Oghlan–Goerdt–Lanka [COGL07, Theorem 4]). *For $H \sim \mathcal{H}(n,p,4)$, there is a polynomial time algorithm certifying that $\chi(H) > \xi$ with high probability for constant ξ when $\overline{m} \geq O(\xi^4 n^2)$.*

We generalize these results to k-uniform hypergraphs:

Theorem 3.6.4. *For $H \sim \mathcal{H}(n,p,k)$, there is a polynomial time algorithm certifying that $\alpha(H) < \beta$ with high probability when $\overline{m} \geq O_k\left(\frac{n^{3k/2}\log^3 n}{\beta^k}\right)$.*

Theorem 3.6.5. *For $H \sim \mathcal{H}(n,p,k)$, there is a polynomial time algorithm certifying that $\chi(H) > \xi$ with high probability when $\overline{m} \geq O_k\left(\xi^k n^{k/2}\log^3 n\right)$.*

The proofs are simple extensions of the $k = 3$ and $k = 4$ cases from [COGL07]. We will first prove Theorem 3.6.4 using Theorem 3.2.1 and this will almost immediately imply Theorem 3.6.5.

Proof of Theorem 3.6.4. Recall that Theorem 3.2.1 deals with k-tuples, not sets of size k. It is easy to express a hypergraph in terms of k-tuples rather than sets of size k. For a set S and $t \in \mathbb{Z}_{\geq 0}$, recall the notation $\binom{S}{t} = \{T \subseteq S \mid |T| = t\}$. For each possible hyperedge $e \in \binom{[n]}{k}$, we associate an arbitrary tuple T_e from among the $k!$ tuples in $[n]^k$ containing the same k elements. To draw from $\mathcal{H}(n,p,k)$, we include each T_e independently with probability p and include all other $T \in [n]^k$ with probability 0.

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

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

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

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

where the second term is 0 because I is an independent set. We proceed in a similar manner to the proof of Theorem 3.2.1, except with a few small changes. First, note that the Cauchy-Schwarz Inequality implies that

$$\sum_{T \in [n]^k} w(T)x^T \leq \sum_{i \in [n]} |x_i| W_i \leq \sqrt{|I|}\sqrt{\sum_{i \in [n]} W_i^2}. \tag{3.39}$$

Continuing as in the proof of Theorem 3.2.1, we bound $\sum_{i \in [n]} W_i^2$ by

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

The first term is upper bounded by $\|A\|\,|I|^{k-1}$. Lemma 3.5.5 implies that this quantity is at most $2^{O(k)}|I|^{k-1}pn^{k/2}\log^3 n$.

To bound the sum, note that it has at most $\binom{|I|}{(k-1)/2}^2 \leq |I|^{k-1}$ nonzero terms. Using (3.7) and (3.8), its expected value is then at most $|I|^{k-1}np$. Each term is independent, and, since $\overline{m} \geq O_k\left(\frac{n^{3k/2}\log^3 n}{\beta^k}\right)$ and $1 \leq |I| \leq n$ imply that $|I|^{k-1}pn^{k/2} \geq 1$, the Chernoff Bound implies that the total value is at most $|I|^{k-1}pn^{k/2}\log n$ with high probability. Therefore, the bound

$$\sum_{i \in [n]} W_i^2 \leq 2^{O(k)}|I|^{k-1}pn^{k/2}\log^3 n$$

holds with high probability. Plugging this into (3.39), we see that we can certify that

$$p\binom{|I|}{k} \leq 2^{O(k)}|I|^{k/2}\sqrt{p}n^{k/4}\log^{3/2} n.$$

Rearranging, we get

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

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

□

Proof of Theorem 3.6.5. For a coloring of a hypergraph H, each color class is an independent set of H. If $\chi(H) \leq \xi$, then there exists a color class of size at least $\frac{n}{\xi}$ and therefore $\alpha(H) \geq \frac{n}{\xi}$. We can then certify that $\alpha(H) < \frac{n}{\xi}$ using Theorem 3.6.4. □

3.7 Extension to larger alphabets

3.7.1 Preliminaries

CSPs over larger domains. We begin by discussing CSPs over domains of size $q > 2$. We prefer to identify such domains with $\mathbb{Z}_q$, so our predicates are $P : \mathbb{Z}_q^k \to \{0,1\}$. The extensions of the definitions and facts from Section 2.1 are straightforward; the only slightly nonobvious notion is that of a literal. We take the fairly standard [Aus08] definition that a literal for variable x_i is any $x_i + c$ for $c \in \mathbb{Z}_q$. Thus there are now q^k possible "negation patterns" c for a P-constraint. We denote by $\mathcal{F}_{q,P}(n,p)$ the distribution over instances of CSP(P) in which each of the $q^k n^k$ constraints is included with probability p; the expected number of constraints is therefore $\overline{m} = q^k n^k p$. We have the following slight variant of Fact 2.1.6.

Fact 3.7.1. *Let $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$. Then the following statements hold with high probability.*

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

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

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

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

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

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

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

Theorem 3.7.2.

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

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

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

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

3.7.2 Conversion to Boolean functions

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

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

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

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

Claim 3.7.3. *The degree of f^b is equal to the degree of f.*

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

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

Now observe that

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

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

3.7.3 Quasirandomness and strong refutation

To prove quasiandomness and strong refutation results for CSPs over larger alphabets, we proceed exactly as in the binary case. We used the $t = k$ case of Lemma 2.1.13 (the Vazirani XOR Lemma [Vaz86, Gol11]) to certify quasirandomness for binary CSPs. A generalization of this case holds for Abelian groups [Rao07, Lemma 4.2].

Lemma 3.7.4. *Let G be an Abelian group and let $\mathcal{U}_G$ be the uniform distribution over G. Also, let $\{\chi_\sigma\}_{\sigma \in G}$ be an orthonormal basis for $L^2(G, \mathcal{U}_G)$ and let $D : G \to \mathbb{R}$ be a distribution over G. If $\widehat{D}(\sigma) \leq \epsilon$ for all $\sigma \in G$, then $d_{\mathrm{TV}}(D, \mathcal{U}_G) \leq \frac{1}{2}|G|^{3/2}\epsilon$.*

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

Lemma 3.7.5. *Let $\sigma \in \mathbb{Z}_q^k$ such that $\sigma \neq \mathbf{0}$ and $|\sigma| = s$. There is an algorithm that with high probability certifies that*

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

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

Proof. The proof is essentially identical to the proof of Lemma 3.2.3. We highlight the differences. First of all, we can write

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

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

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

where $w_{\sigma,\alpha}(T) = \sum_{c \in \mathbb{Z}_q^k} 1_{\{(T,c) \in \mathcal{I}\}} \chi_\sigma(\alpha + c(\sigma))$. Observe that $\mathbf{E}[w_{\sigma,\alpha}(T)] = 0$ and $\mathbf{Pr}[w_{\sigma,\alpha}(T) \neq 0] \leq q^k p$. Since $\|\chi_\sigma\| = 1$, observe that the Cauchy-Schwarz Inequality implies that $|\chi_\sigma| \leq q^{k/2}$ for all σ. Then $|w_{\sigma,\alpha}(T)| \leq q^{3k/2}$ for all α and σ. For every α, we can then apply Lemma 3.2.4 just as in the proof of Lemma 3.2.3. □

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

Theorem 3.7.6. *There is an efficient algorithm that certifies that an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$ is γ-quasirandom with high probability when $m \geq \frac{q^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.*

Theorem 3.7.7. *There is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq \mathbf{E}[P] + \gamma$ with high probability when $m \geq \frac{q^{O(k)} n^{k/2} \log^5 n}{\gamma^2}$.*

3.7.4 Refutation of non-t-wise supporting CSPs

We will show that the dual polynomial characterization of being far from t-wise supporting described in Section 3.1.2 generalizes to larger alphabets. We extend the definitions of t-wise supporting and δ-separating polynomials to the $\mathbb{Z}_q$ case in the natural way.

Lemma 3.7.8. *For $P : \mathbb{Z}_q^k \to \{0,1\}$ and $0 \le \delta < 1$, there exists a polynomial $Q : \mathbb{Z}_q^k \to \mathbb{R}$ of degree at most t that δ-separates P if and only if P is δ-far from supporting a t-wise uniform distribution.*

Proof. The proof uses the following dual linear programs exactly as in the proof of Lemma 3.1.9.

$$\begin{aligned}
&\text{minimize} && \sum_{z \in \mathbb{Z}_q^k} (1 - P(z))\mathcal{D}(z) && && (3.40)\\
&\text{s.t.} && \sum_{z \in \mathbb{Z}_q^k} \mathcal{D}(z)\chi_\sigma(z) = q^k \widehat{\mathcal{D}}(\sigma) = 0 && \forall \sigma \in \mathbb{Z}_q^k \quad 0 < |\sigma| \le t && (3.41)\\
& && \sum_{z \in \mathbb{Z}_q^k} \mathcal{D}(z) = 1 && && \\
& && \mathcal{D}(z) \ge 0 && \forall z \in \mathbb{Z}_q^k &&
\end{aligned}$$

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

To prove Lemma 3.1.9, we needed to show in the binary case that feasible solutions to the primal LP (3.1) were t-wise uniform. We now argue that the constraint (3.41) is a sufficient condition for t-wise uniformity of $\mathcal{D}$ in the q-ary case. For a distribution $\mathcal{D}$ over $\mathbb{Z}_q^k$ and $S \subseteq [k]$, define $\mathcal{D}_S$ to be the marginal distribution of $\mathcal{D}$ on $(\mathbb{Z}_q^k)_S$, i.e., $\mathcal{D}_S(z) = \sum_{z' \in \mathbb{Z}_q^k, z'_S = z} \mathcal{D}(z')$. We need to show that (3.41) implies that $\mathcal{D}_S = \mathcal{U}_q^{|S|}$ for all $S \subseteq [k]$ with $1 \le |S| \le t$.

Fix such an S and let $|S| = s$. Consider the basis $\{\chi_\alpha\}_{\alpha \in \mathbb{Z}_q^s}$. Lemma 3.7.4 implies that it suffices to show that $\mathbf{E}_{\boldsymbol{z} \sim \mathcal{U}_q^s}[D_S(\boldsymbol{z})\chi_\alpha(\boldsymbol{z})] = 0$ for all $\alpha \in \mathbb{Z}_q^s$. Observe that $\mathbf{E}_{\boldsymbol{z} \sim \mathcal{U}_q^s}[D_S(\boldsymbol{z})\chi_\alpha(\boldsymbol{z})] = \mathbf{E}_{\boldsymbol{z}' \sim \mathcal{U}_q^k}[\mathcal{D}(\boldsymbol{z}')\chi_\sigma(\boldsymbol{z}')]$ for $\sigma \in \mathbb{Z}_q^k$ such that $\sigma_i = \alpha_i$ for $i \in S$ and $\sigma_i = 0$ otherwise. Since $|S| \le t$, we know that $|\sigma| \le t$ and (3.41) implies $\mathbf{E}_{\boldsymbol{z}' \sim \mathcal{U}_q^k}[\mathcal{D}(\boldsymbol{z}')\chi_\sigma(\boldsymbol{z}')] = 0$.

The rest of the proof is exactly as in the binary case. □

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

Theorem 3.7.9. *Let P be δ-far from being t-wise supporting. There exists an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \le 1 - \delta + \gamma$ with high probability when $\overline{m} \ge \frac{q^{O(k)} n^{t/2} \log^5 n}{\gamma^2}$ and $t \ge 2$.*

The proof is essentially identical to Proof 2 of Theorem 3.2.9.

Corollary 3.2.11 also extends to larger alphabets.

Corollary 3.7.10. *Let P be a predicate that does not support any t-wise uniform distribution. Then there is an efficient algorithm that, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of $\mathrm{CSP}(P)$, certifies that $\mathrm{Opt}(\mathcal{I}) \le 1 - 2^{-\widetilde{O}(q^t k^t)}$ with high probability when $\overline{m} \ge 2^{\widetilde{O}(q^t k^t)} n^{t/2} \log^5 n$ and $t \ge 2$.*

This follows directly from Theorem 3.7.9 and the following extension of Corollary 3.1.10 to larger alphabets.

Corollary 3.7.11. *Suppose $P : \mathbb{Z}_q^k \to \{0,1\}$ is not t-wise supporting. Then it is in fact δ-far from t-wise supporting for $\delta = 2^{-\widetilde{O}(q^t k^t)}$.*

The proof is essentially identical to the proof of Corollary 3.1.10: Observe that the LP (3.40) has at most $q^t k^t$ variables and proceed exactly as before.

3.7.5 SOS proofs

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

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

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

Lemma 3.7.12. *Let $\mathbf{0} \neq \sigma \in \mathbb{Z}_q^k$ with $|\sigma| = s$. Then*

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

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

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

Proof. In the proof of Lemma 3.7.5, we certify that $|\widehat{D_{\mathcal{I},y}}(\sigma)|$ is small by certifying that $\left|\widehat{D_{\mathcal{I},y}}(\sigma)^{\mathrm{poly}}\right|$ is small. The proof of Lemma 3.7.5 relies only on Lemma 3.2.4; we can replace this with its SOS version Lemma 3.4.2. □

Remark 3.7.13. We stated the lemma with the weaker set of axioms $\{y(i,a)^2 \le 1\}_{i \in [n], a \in \mathbb{Z}_q}$. Since $y(i,a)^2 = y(i,a)$ implies $y(i,a)^2 \le 1$ in degree-2 SOS, the lemma holds with the axioms $\{y(i,a)^2 = y(i,a)\}_{i \in [n], a \in \mathbb{Z}_q}$ as well.

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

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

Theorem 3.7.14. *Given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of* CSP(P)*,*

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

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

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

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

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

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

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

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

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

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

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

This is equal to

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

Since $|P(z)| \le 1$ and $|\chi_\sigma(z)| \le q^{O(k)}$, $\sum_{\sigma\in\mathbb{Z}_q^k} |\widehat{P}(\sigma)| \le q^{O(k)}$. We can then apply Lemma 3.7.12 for each σ to complete the proof. □

SOS refutation of non-t-wise supporting CSPs.

Theorem 3.7.15. *Let P be δ-far from being t-wise supporting. Then, given an instance $\mathcal{I} \sim \mathcal{F}_{q,P}(n,p)$ of* CSP(P)*,*

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

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

To prove this theorem, we need a version of Claim 3.4.6 for larger alphabets.

Claim 3.7.16. *Let $f : \mathbb{Z}_q^k \to \mathbb{R}$ such that $f(z) \geq 0$ for all $z \in \mathbb{Z}_q^k$ and let $f^b(v) = \sum_{\alpha \in \mathbb{Z}_q^k} f(\alpha) \prod_{i \in [k]} v(i, \alpha_i)$. Then*

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

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

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

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

With this claim, the proof of the theorem exactly follows that of Theorem 3.2.9.

Proof of Theorem 3.7.15. Claim 3.7.16 implies that

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

Summing over all constraints, we get that

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

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

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

Just as in the proof of Theorem 3.7.14, we can rewrite this in degree-k SOS as

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

We then rearrange to get

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

Since $\mathbf{E}[Q] = 0$ and $Q \geq -1$, we know that $|Q| \leq q^{O(k)}$ and therefore $|\widehat{Q}(\sigma)| \leq q^{O(k)}$. We can then apply Lemma 3.7.12 for each σ to complete the proof. □

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

The setting of [DLSS14] fixes the number of constraints in a CSP instance, whereas the model described in Chapter 2 includes each possible constraint in the instance with some probability p. Here we show that results from our setting easily extend to that of [DLSS14] by giving an algorithm that simulates the behavior of our model when the number of constraints is fixed.

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

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

Theorem 3.8.1. *Suppose there exists an efficient algorithm R that, on a given CSP instance $\mathcal{I} \sim \mathcal{F}_P(n,p)$, for all $p \geq p_{\min}$, certifies that $\mathrm{Opt}(\mathcal{I}) \leq \eta$ for some $0 \leq \eta < 1$ with high probability. Then there exists an efficient algorithm $\mathcal{A}$ that certifies that a random instance $\mathcal{J}$ of $\mathrm{CSP}(P)$ with μ constraints has $\mathrm{Opt}(\mathcal{J}) \leq \eta + 2\left(\mu^{-1}\ln\mu\right)^{1/2}$ with high probability when $\mu\left(1-\left(\mu^{-1}\ln\mu\right)^{1/2}\right) \geq (2n)^k p_{\min}$.*

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

Algorithm 1

Algorithm $\mathcal{A}$

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

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

Furthermore, the probability of failing to refute an instance with value at most $1 - \eta + 2d$ due to exiting at step 2 is $o_{k,t}(1)$. We treat m as a sum of $(2n)^k$ independent Bernoulli variables with

probability p and denote $\mathbf{E}[m]$ by $\overline{m}$. Applying a Chernoff bound yields the following.

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

Similarly,

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

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

Chapter 4

Sum of squares lower bounds for refuting any CSP

4.1 Overview of results

In this chapter, we prove lower bounds on the SOS degree required to refute random instances of CSP(P). Actually, our results hold in a slightly more general model. Let Ω be a finite alphabet and let $\mathcal{P}$ be a collection of nontrivial predicates $\Omega^k \to \{0,1\}$. An input $\mathcal{I}$ to the problem CSP($\mathcal{P}$) consists of n variables $x_1, \ldots, x_n$, 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. Typical examples involve a binary alphabet $\Omega = \{0,1\}$, a fixed predicate $P : \{0,1\}^k \to \{0,1\}$, 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"). The CSP(P) problem introduced in Chapter 1 is CSP($P^{\pm}$) in the notation of this chapter. For example, if P is the k-bit logical OR function, then CSP($P^{\pm}$) is simply the k-SAT problem.

Using this notation, we restate the results we will prove in this chapter.

Theorem 1.5.4 ([KMOW17])**.** 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 with high probability, degree-$\widetilde{\Omega}\left(\frac{n}{\Delta^{2/(t-1)}}\right)$ SOS fails to $(\delta_P(t) + o(1))$-refute $\mathcal{I}$.

When $\delta_P(t) = 0$, our result does not need the additive $o(1)$ in refutation strength.

Theorem 1.5.5 ([KMOW17])**.** 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 CSP($P^{\pm}$) with $m = \Delta n$ constraints, with high probability degree-$\widetilde{\Omega}\left(\frac{n}{\Delta^{2/(\mathcal{C}(P)-2)}}\right)$ SOS fails to (weakly) refute $\mathcal{I}$.

4.2 Technical framework

Up to this point, 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.

4.2.1 Constraint satisfaction notation

Unlike our upper bounds in Chapter 3, our lower bounds have no dependence on alphabet size. We therefore prove them for CSPs over arbitrary alphabets. Here, we introduce some notation.

Notation 4.2.1. 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. Rather than our full Theorem 1.5.4 concerning δ-refutation, the reader is advised to mainly keep in mind our Theorem 1.5.5, which is concerned with (weak) refutation of CSPs for which the predicates support a $(\tau - 1)$-wise uniform distribution. Given our proof of Theorem 1.5.5, the more general Theorem 1.5.4 will fall out fairly easily.

Notation 4.2.2. 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 4.2.3 (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 4.1, consider our Theorem 1.5.5 in which we have CSP($P^{\pm}$) instances, where $P : \{\pm 1\}^k \to \{0,1\}$ 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" $\ell \in \{\pm\}^k$, the distribution μ_ℓ gotten by negating inputs to μ according to ℓ is also $(\tau - 1)$-wise uniform. In the 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 1.5.4 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.)

4.2.2 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 4.4, but here we will give the briefest possible definition.

Notation 4.2.4. We introduce two parameters: $1 \leq \mathtt{SMALL} \leq n/2$ and $0 < \zeta < 1$. (For the sake of intuition, the reader might think of, e.g., $\mathtt{SMALL} = n^{\Omega(1)}$ and $\zeta = \frac{1}{\log n}$.) The parameters are assumed to satisfy $K \leq \zeta \cdot \mathtt{SMALL}$.

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 $c \leq 2 \cdot \mathtt{SMALL}$. 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 4.2.5. *(Roughly stated; see Theorem 4.4.12 for a precise statement.) A random G with constraint density Δ will satisfy the Plausibility Assumption whp provided* $\mathtt{SMALL} \ll \dfrac{n}{\Delta^{2/(\tau-2-\zeta)}}$.

The Plausibility Assumption is highly similar to the assumption that G has good vertex-expansion, and indeed our proof of Theorem 4.4.12 in Appendix 4.8 is a completely standard variant of the well-known proof that random bipartite graphs have good vertex-expansion.

4.2.3 Main result

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

Theorem 4.2.6 (Roughly stated; cf. Theorem 4.6.1.)**.** *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 $D = \frac{1}{3}\zeta \cdot \mathtt{SMALL}$, there exists a degree-$D$ pseudoexpectation $\widetilde{\mathbf{E}}$ 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* $\mathrm{supp}(\mu_f)$*." (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 4.2.5 and Theorem 4.2.6, we can now point out how the constraint density vs. SOS-degree tradeoff arises in our Theorem 1.5.5. For $\mathrm{CSP}(P^{\pm})$ 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 $\widetilde{\Omega}\left(\frac{n}{\Delta^{2/(\mathcal{C}(P)-2)}}\right)$. More precise details of parameter-setting are given in Section 4.7.

4.3 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 $\widetilde{\mathbf{E}}$ as described in Theorem 4.2.6.

4.3.1 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.

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 $\widetilde{\mathbf{E}}$ is a degree-d pseudoexpectation, then for every collection S of at most $d/2$ variables, $\widetilde{\mathbf{E}}$ 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 $\widetilde{\mathbf{E}}$ 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: $\widetilde{\mathbf{E}}[p] \geq 0$ for p for every truly nonnegative polynomial p depending on at most d variables. However, when trying to show that the more global $\widetilde{\mathbf{E}}[p^2]$ 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 $\widetilde{\mathbf{E}}$ 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 $\widetilde{\mathbf{E}}$ 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 4.5.

4.3.2 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 $\widetilde{\mathbf{E}}$ 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 $\widetilde{\mathbf{E}}[\cdot]$, meaning, the pseudoexpectation of the product of any distinct pair of basis polynomials is 0.

Fact 4.3.1. *Suppose there exists a basis $f_1, f_2, \ldots$ for degree-d polynomials such that the following two properties hold:*

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

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

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

Notice that the standard Fourier monomial basis guarantees us positivity (since $\widetilde{\mathbf{E}}$ 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 $\widetilde{\mathbf{E}}[\cdot]$ is to perform the Gram–Schmidt process on, say, the monomial basis $1, x_1, x_2, \ldots, x_1 x_2, \ldots$ to get a new basis $f_1, f_2, \ldots$. 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 $\widetilde{\mathbf{E}}$ 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 the 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 $\widetilde{\mathbf{E}}$), 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 orthogonal under $\widetilde{\mathbf{E}}[\cdot]$; e.g., consider the following 3-XOR

instance:

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

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 $\widetilde{\mathbf{E}}[\cdot]$ satisfies all constraints then $\widetilde{\mathbf{E}}[x_1] = 1$ and $\widetilde{\mathbf{E}}[y_1] = -1$. So neither x_1 nor y_1 are orthogonal to 1. Since the two sets of equations are disjoint, we also know that $\widetilde{\mathbf{E}}[x_1y_1] = -1$, 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 $\overline{x}_{i1}$ (where we use the notation $\overline{z} := z - \widetilde{\mathbf{E}}[z]$). Now we have that $\overline{x}_{i1}$ is orthogonal to 1 and, by independence of the blocks, $\widetilde{\mathbf{E}}[\overline{x}_{i1} \cdot \overline{x}_{j1}] = 0$ 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 $\{1, \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n, \overline{x_1x_2}, \ldots\}$ 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 $\overline{x_iy_i}$ and $\overline{x_jy_j}$ are not orthogonal under $\widetilde{\mathbf{E}}[\cdot]$:

$$\widetilde{\mathbf{E}}[\overline{x_iy_i} \cdot \overline{x_jy_j}] = \widetilde{\mathbf{E}}[x_iy_i \cdot x_jy_j] - \widetilde{\mathbf{E}}[x_iy_i]\,\widetilde{\mathbf{E}}[x_jy_j] = b_ib_j - 0 = b_ib_j.$$

Instead, consider the basis

$$\widehat{1} = 1,\ \widehat{x_0} = x_0,\ \widehat{x_1} = x_1,\ \ldots,\ \widehat{y_1} = y_1,\ \widehat{y_2} = y_2,\ \ldots,\ \widehat{x_1y_1} = x_1y_1 - b_1x_0,\ \widehat{x_2y_2} = x_2y_2 - b_2x_0,\ \ldots$$

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 4.3.1 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

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

Next, note that

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

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 $\widetilde{\mathbf{E}}[x^S \cdot y_T] \neq 0$. That is, x^S and y_T are correlated under $\widetilde{\mathbf{E}}[\cdot]$. 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 4.6.

4.4 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 4.4.1 (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 4.4.2. For H a subgraph, we write $\mathrm{vbls}(H)$ for the set of variables appearing in H, $\mathrm{cons}(H)$ for the set of constraints appearing in H, and $\mathrm{edges}(H)$ for the set of edges appearing in H.

Notation 4.4.3. Given $f \in \mathrm{cons}(H)$, we write $N_H(f) = \{i \in \mathrm{vbls}(H) : (f, i) \in \mathrm{edges}(H)\}$. Note that this is *not* necessarily the same thing as $N(f) \cap \mathrm{vbls}(H)$.

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

Definition 4.4.4 (Small subgraphs)**.** We say that subgraph H is *small* if $|\mathrm{cons}(H)| \leq \mathtt{SMALL}$.

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 1.5.7. 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 4.1; call it H.

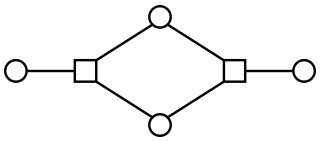

Figure 4.1: An example small subgraph. Constraint-vertices are squares, variable-vertices are circles.

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 $2\cdot(\tau-\zeta)/2+(4-6) \geq 0$; 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 $c = |\text{cons}(H)|$, $v = |\text{vbls}(H)|$, $e = |\text{edges}(H)|$,

$$\mathbf{E}[\#\text{ copies of } H] \approx m^c n^v n^{-e} = n^{c\cdot(\tau-\zeta)/2+v-e} \implies H \text{ "plausibly occurs" iff } c\cdot(\tau-\zeta)/2+(v-e) \geq 0. \tag{4.1}$$

This inequality is precisely the one occurring in the Plausibility Assumption from Section 4.2.2.

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 4.5, we will be concerned almost exclusively with subgraphs of G in which all constraint-vertices have degree at least τ:

Definition 4.4.5 (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 4.4.6. 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 4.4.7 (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 4.4.8 (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 4.4.9 (Plausible τ-subgraphs)**.** Let H be a τ-subgraph. We say that H is *plausible* if $I(H) \geq 0$.

Remark 4.4.10. H being plausible implies (indeed, is equivalent to) $|\mathrm{cons}(H)| \leq \frac{1}{\zeta} \cdot R(H)$. 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 (4.1).

Lemma 4.4.11. *Let H be a τ-subgraph with $c = |\mathrm{cons}(H)|$, $v = |\mathrm{vbls}(H)|$, $e = |\mathrm{edges}(H)|$, 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. □

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 $|\mathrm{cons}(H)| \leq 2 \cdot \mathtt{SMALL}$ are plausible.*

As mentioned earlier, for an appropriate choice of SMALL, the Plausibility Assumption holds for a random instance. More precisely, in Appendix 4.8 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 1.5.5 concerning weak refutation. The second claim ("Moreover...") is a technical variant needed to extend our results to give Theorem 1.5.4 concerning δ-refutation.

Theorem 4.4.12. *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 $\boldsymbol{G}$ is a random instance with $m = \Delta n$ constraints, the Plausibility Assumption holds provided*

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

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

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

4.5 Defining the pseudoexpectation

4.5.1 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 4.5.1 (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 4.5.2. 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 4.5.3 (Closure, $\mathrm{cl}(S)$)**.** Let S be a set of variables. We define the *closure* of S, written $\mathrm{cl}(S)$, to be the union of all *small* S-closed τ-subgraphs H. Note that $\mathrm{cl}(S)$ is itself an S-closed τ-subgraph.

Remark 4.5.4. A key warning to remember: we do not necessarily have $S \subseteq \mathrm{vbls}(\mathrm{cl}(S))$.

Remark 4.5.5. Let $T \subseteq S$. Then if H is T-closed, it is also S-closed. It follows that $\mathrm{cl}(T) \subseteq \mathrm{cl}(S)$.

Fact 4.5.6. *The only plausible $\emptyset$-closed τ-subgraph H is $H = \emptyset$. It follows that* $\mathrm{cl}(\emptyset) = \emptyset$.

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

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

Theorem 4.5.7. *Let S be a set of variables with $|S| \leq \zeta \cdot \mathtt{SMALL}$. Then $\mathrm{cl}(S)$ is small and satisfies $R(\mathrm{cl}(S)) \leq |S|$.*

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

Lemma 4.5.8. *Suppose that H is a τ-subgraph formed as a union, $H = H_1 \cup \cdots \cup H_t$, where each H_j is small and where we have $R(H_1 \cup \cdots \cup H_j) \leq \zeta \cdot \mathtt{SMALL}$ 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 $H' = H_1 \cup \cdots \cup H_{t-1}$ is small. Since H_t is also small we have $|\mathrm{cons}(H')|, |\mathrm{cons}(H_t)| \leq \mathtt{SMALL}$ and hence $|\mathrm{cons}(H' \cup H_t)| \leq 2 \cdot \mathtt{SMALL}$. Thus $H' \cup H_t$ is plausible and so

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

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

In proving Theorem 4.5.7, 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 $\mathrm{cl}(S)$ in this way, let $V = \mathrm{vbls}(\mathrm{cl}(S))$, 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 $|S| \leq \zeta \cdot \mathtt{SMALL}$, and Lemma 4.5.8 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 $\mathrm{cl}(S)$. Thus we have shown:

Theorem 4.5.9. *Let S be a set of variables with $|S| \leq \zeta \cdot \mathtt{SMALL}$. Then $\mathrm{cl}(S \cup \mathrm{vbls}(\mathrm{cl}(S))) = \mathrm{cl}(S)$.*

4.5.2 The planted distribution

Definition 4.5.10 (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 $\boldsymbol{x} \in \Omega^n$ to the variables of G, defined as follows: For each constraint $f \in \mathrm{cons}(H)$ we independently draw an assignment $\boldsymbol{w}_f \in \Omega^{N(f)}$ according to μ_f. We write its component associated to variable $i \in N_H(f)$ as $\boldsymbol{w}_{f,i}$, and think of it as an assignment "suggested" for this variable. (Note that we will ignore the components of $\boldsymbol{w}_f$ correspoding to variables not in $N_H(f)$.) Now each variable $i \in \mathrm{vbls}(H)$ has one or more assignments in Ω suggested by its adjacent constraints. We get a unique assignment $\boldsymbol{x}_i$ for it by *conditioning* on all the suggestions being consistent. (We will show later in (4.8) 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 4.5.11. For each $i \in \mathrm{vbls}(G)$ 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 $\mathrm{cl}(S)$, 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 4.5.12. *Let S be a set of variables and let $H \supseteq \mathrm{cl}(S)$ be a small subgraph. Then the marginal of η_H on S is the same as the marginal of $\eta_{\mathrm{cl}(S)}$ on S.*

Remark 4.5.13. 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 $v_H = |\mathrm{vbls}(H)|$ and $e_H = |\mathrm{edges}(H)|$. 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

$$\begin{aligned}
p_H(x) &\coloneqq \mathop{\mathbf{Pr}}_{\boldsymbol{w}}[\boldsymbol{w} \text{ suggestions consistent, and the assignment to } \mathrm{vbls}(H) \text{ they agree on is } x] \\
&= q^{-(n-v_H)} \cdot \mathop{\mathbf{E}}_{\boldsymbol{w}} \prod_{(f,i)\in\mathrm{edges}(H)} \left(q^{-1}(1 + \overline{1_{\boldsymbol{w}_{f,i}}}(x_i)\right) \\
&= q^{v_H - e_H - n} \cdot \sum_{H' \subseteq H} \mathop{\mathbf{E}}_{\boldsymbol{w}} \prod_{(f,i)\in\mathrm{edges}(H')} \overline{1_{\boldsymbol{w}_{f,i}}}(x_i) \qquad (4.2) \\
&= q^{v_H - e_H - n} \cdot \sum_{H' \subseteq H} \prod_{f\in\mathrm{cons}(H')} \mathop{\mathbf{E}}_{\boldsymbol{w}_f \sim \mu_f} \prod_{i \in N_{H'}(f)} \overline{1_{\boldsymbol{w}_{f,i}}}(x_i), \qquad (4.3)
\end{aligned}$$

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

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

since $\mathbf{E}_{\boldsymbol{c} \sim \Omega}[\overline{1_{\boldsymbol{c}}}(x_i)] = 0$ for any fixed value $x \in \Omega$. Thus in (4.3) it is equivalent to sum over τ-subgraphs H', and so returning to (4.2) we get

$$p_H(x) = q^{v_H - e_H - n} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ a } \tau\text{-subgraph}}} \mathop{\mathbf{E}}_{\boldsymbol{w}} \prod_{(f,i)\in\mathrm{edges}(H')} \overline{1_{\boldsymbol{w}_{f,i}}}(x_i). \qquad (4.4)$$

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 $x_{\overline{T}}$ onto the coordinates not in T. Then

$$\begin{aligned} p_H(x_T) &:= \mathop{\mathbf{Pr}}_{\boldsymbol{w}}[\boldsymbol{w} \text{ suggestions consistent, and the assignment to } T \text{ they agree on is } x_T] \\ &= \sum_{x_{\overline{T}} \in \Omega^{\overline{T}}} p_H(x_T, x_{\overline{T}}) = q^{n-|T|} \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{x}_{\overline{T}} \sim \Omega^{\overline{T}} \\ \text{uniform}}} [p_H(x_T, \boldsymbol{x}_{\overline{T}})] \\ &= q^{v_H - e_H - |T|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ a } \tau\text{-subgraph}}} \mathop{\mathbf{E}}_{\boldsymbol{w}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in T}} \overline{1_{\boldsymbol{w}_{f,i}}}(x_i) \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{x}_{\overline{T}} \sim \Omega^{\overline{T}} \\ \text{uniform}}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in \overline{T}}} \overline{1_{\boldsymbol{w}_{f,i}}}(\boldsymbol{x}_i), \end{aligned} \tag{4.5}$$

where we used (4.4). Now suppose the τ-subgraph H' has a leaf vertex j that is in $\overline{T}$; i.e., it's *not* in T. Then $\boldsymbol{x}_j$ appears exactly once in the above, within the expression

$$\mathop{\mathbf{E}}_{\substack{\boldsymbol{x}_{\overline{T}} \sim \Omega^{\overline{T}} \\ \text{uniform}}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in \overline{T}}} \overline{1_{\boldsymbol{w}_{f,i}}}(\boldsymbol{x}_i). \tag{4.6}$$

As $\boldsymbol{x}_j$ is chosen uniformly and independently of all random variables, the above contains a factor of the form $\mathbf{E}_{\boldsymbol{x}_j \sim \Omega}[\overline{1_{\boldsymbol{w}_{f,j}}}(\boldsymbol{x}_j)]$. But for *any* fixed outcome of $\boldsymbol{w}_{f,j}$, this expectation is 0, meaning (4.6) will vanish. Thus any summand H' in (4.5) will vanish if H' has a leaf variable outside T. Thus we may equivalently sum only over T-*closed* H'. That is,

$$\begin{aligned} p_H(x_T) &= \mathop{\mathbf{Pr}}_{\boldsymbol{w}}[\boldsymbol{w} \text{ suggestions consistent, and the assignment to } T \text{ they agree on is } x_T] \\ &= q^{v_H - e_H - |T|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ is } T\text{-closed}}} \mathop{\mathbf{E}}_{\boldsymbol{w}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in T}} \overline{1_{\boldsymbol{w}_{f,i}}}(x_i) \cdot \mathop{\mathbf{E}}_{\substack{\boldsymbol{x}_{\overline{T}} \sim \Omega^{\overline{T}} \\ \text{uniform}}} \prod_{\substack{(f,i) \in \text{edges}(H') \\ i \in \overline{T}}} \overline{1_{\boldsymbol{w}_{f,i}}}(\boldsymbol{x}_i) \\ &= q^{v_H - e_H - |T|} \cdot \sum_{\substack{H' \subseteq H \\ H' \text{ is } T\text{-closed}}} \mathop{\mathbf{E}}_{\boldsymbol{w}} \mathop{\mathbf{E}}_{\substack{\boldsymbol{x} \sim \Omega^n \text{ unif.}, \\ \text{condit. on } \boldsymbol{x}_T = x_T}} \prod_{(f,i) \in \text{edges}(H')} \overline{1_{\boldsymbol{w}_{f,i}}}(\boldsymbol{x}_i). \end{aligned} \tag{4.7}$$

Suppose we took $T = \emptyset$ above. Since H is small, every subgraph H' is plausible and hence Fact 4.5.6 implies that the above has only one summand, corresponding to $H' = \emptyset$. The summand is trivially 1, and hence

$$\mathop{\mathbf{Pr}}_{\boldsymbol{w}}[\boldsymbol{w} \text{ suggestions consistent}] = q^{v_H - e_H}. \tag{4.8}$$

Observe that this does not depend at all on the μ_f's; in particular, it is easily seen to the be the probability of consistent suggestions under completely uniform μ_f's. In any case, since (4.8) is positive, as promised, we may condition on the associated event; thus from (4.7) we obtain

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

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

4.5.3 Pseudoexpectations

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

Definition 4.5.14. Given a polynomial expression $p(x)$ in the indeterminates $1_c(x_i)$, we write

$$\mathrm{vbls}(p) = \{i : \text{at least one } 1_c(x_i) \text{ appears in } p(x)\},$$
$$\deg_{\mathrm{mlin}}(p) = \max\{|\mathrm{vbls}(M)| : M(x) \text{ is a monomial in } p(x)\}.$$

We call the latter the *multilinear-degree*; note that $\deg_{\mathrm{mlin}}(p) \le \deg(p)$ always.

Recall that a *pseudoexpectation* on polynomials of degree at most D is a linear map $\widetilde{\mathbf{E}}[\cdot]$ satisfying $\widetilde{\mathbf{E}}[1] = 1$. 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 $\widetilde{\mathbf{E}}[\cdot]$ *satisfies the identity* $p(x) = 0$ if $\widetilde{\mathbf{E}}[p(x) \cdot q(x)] = 0$ for all polynomials $q(x)$ with $\deg(p \cdot q) \le D$.

Definition 4.5.15 (Our pseudoexpectation)**.** We'll define our pseudoexpectation $\widetilde{\mathbf{E}}[\cdot]$ on all polynomials of *multilinear-degree* at most $\zeta \cdot \mathrm{SMALL}$; in particular, this defines it for all polynomials of (usual) degree at most $\zeta \cdot \mathrm{SMALL}$. We define it by imposing that $\widetilde{\mathbf{E}}[M(x)] = \mathbf{E}_{\mathrm{cl}(\mathrm{vbls}(M))}[M(\boldsymbol{x})]$ for all monomials $M(x)$ having $\deg_{\mathrm{mlin}}(M) \le \zeta \cdot \mathrm{SMALL}$. (Here we are using the abbreviation $\mathbf{E}_C[M(\boldsymbol{x})]$ for $\mathbf{E}_{\boldsymbol{x} \sim \eta_C}[q(\boldsymbol{x})]$.) By Theorem 4.5.7, this makes sense in that $\mathrm{cl}(\mathrm{vbls}(M))$ will always be small. Note that we have $\widetilde{\mathbf{E}}[1] = \mathbf{E}_{\mathrm{cl}(\emptyset)}[1] = 1$, as required.

Theorem 4.5.16. *Let $p(x)$ be a polynomial expression of multilinear-degree at most $\zeta \cdot \mathrm{SMALL}$. Let H be any small subgraph containing*

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

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

$$\widetilde{\mathbf{E}}[p(x)] = \mathop{\mathbf{E}}_{H}[p(\boldsymbol{x})] = \mathop{\mathbf{E}}_{H'}[p(\boldsymbol{x})].$$

Proof. This is immediate from Theorem 4.5.12 and Remark 4.5.5. □

Theorem 4.5.17. *Let $p(x)$ be a polynomial with $S = \mathrm{vbls}(p)$ satisfying $|S| \le \deg(p)$, $|S| \le \zeta \cdot \mathrm{SMALL}$. Assume that $p(\boldsymbol{x})$ is identically zero for $\boldsymbol{x} \sim \eta_{\mathrm{cl}(S)}$. (Note that $\mathrm{cl}(S)$ is small by Theorem 4.5.7.) Then our $\widetilde{\mathbf{E}}[\cdot]$ satisfies the identity $p(x) = 0$.*

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

$$\widetilde{\mathbf{E}}[p(x) \cdot q(x)] = \sum_j \widetilde{\mathbf{E}}[p(x) \cdot M_j(x)] = \sum_j \mathop{\mathbf{E}}_{\mathrm{cl}(S \cup \mathrm{vbls}(M_j))}[p(\boldsymbol{x}) \cdot M_j(\boldsymbol{x})]. \tag{4.9}$$

Here the last equality used Theorem 4.5.16 and the smallness of $\mathrm{cl}(S \cup \mathrm{vbls}(M_j))$, which follows from Theorem 4.5.7 and the fact that $|S \cup \mathrm{vbls}(M_j)| \le \deg(p) + \deg(q) = \deg(p \cdot q) \le \zeta \cdot \mathrm{SMALL}$. But since $\mathrm{cl}(S \cup \mathrm{vbls}(M_j)) \supseteq \mathrm{cl}(S)$ (Remark 4.5.5), Theorem 4.5.12 tells us that $p(\boldsymbol{x})$ has the same distribution under $\eta_{\mathrm{cl}(S \cup \mathrm{vbls}(M_j))}$ and $\eta_{\mathrm{cl}(S)}$; i.e., it is identically 0. Thus (4.9) vanishes, as needed. □

We have the following immediate corollaries:

Corollary 4.5.18. *Our pseudoexpectation $\widetilde{\mathbf{E}}[\cdot]$ 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 $\widetilde{\mathbf{E}}[p(x)] = \widetilde{\mathbf{E}}[\mathrm{multilin}(p(x))]$*, where* $\mathrm{multilin}(p(x))$ *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 4.2.6):

Corollary 4.5.19. *Our pseudoexpectation* $\widetilde{\mathbf{E}}[\cdot]$ *satisfies the identity*

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

for all $f \in \mathrm{cons}(G)$*; i.e., "*$\widetilde{\mathbf{E}}[\cdot]$*'s distribution on* $N(f)$ *is always in* $\mathrm{supp}(\mu_f)$*".*

Proof. We apply Theorem 4.5.17, with $S = N(f)$, which satisfies $|S| = \deg(s_f)$ and $|S| \leq K \leq \zeta \cdot \mathtt{SMALL}$. 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 $H_f \subseteq \mathrm{cl}(S)$. It then follows from the definition of $\boldsymbol{x} \sim \eta_{\mathrm{cl}(S)}$ that $s_f(\boldsymbol{x}) \equiv 1$, since the restriction of $\boldsymbol{x}$ to $N(f)$ will always be supported on $\mathrm{supp}(\mu_f)$. □

4.6 The proof of positive semidefiniteness

4.6.1 Setup

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

Theorem 4.6.1. *If* $p(x)$ *is a polynomial expression of degree at most* D*, then* $\widetilde{\mathbf{E}}[p(x)^2] \geq 0$*.*

In light of Corollary 4.5.18, 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 $p(x) \in \mathrm{span}(x^S : S \in \mathcal{M}^{\leq D})$, where we introduce the following notation:

Definition 4.6.2. 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 4.6.3. 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 4.6.4. All of the ideas in our proof of Theorem 4.6.1 are present in the $q = 2$ case; only notational complexities arise for $q > 2$. Thus the reader is encouraged to keep the Boolean case $\Omega = \{\mathrm{false}, \mathrm{true}\}$ in mind. In this case, since $\widetilde{\mathbf{E}}[\cdot]$ satisfies the identity $1_{\mathrm{false}}(x_i) = 1 - 1_{\mathrm{true}}(x_i)$, one can also ignore the indeterminate $1_{\mathrm{false}}(x_i)$ (since $1 \in \mathrm{span}(\mathcal{M}^{\leq 0})$ already). Then one can more naturally write the indeterminate $1_{\mathrm{true}}(x_i)$ as x_i and the monomial x^S becomes $\prod_{i \in S} x_i$.

4.6.2 Gram–Schmidt overview

Notation 4.6.5. 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 $\mathrm{pr}(S)$ 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 $\widetilde{\mathbf{E}}[\cdot]$ as the "inner product": $\langle p(x), q(x) \rangle \coloneqq \widetilde{\mathbf{E}}[p(x) \cdot q(x)]$. 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 $\widetilde{\mathbf{E}}[\cdot]$ (meaning $\widetilde{\mathbf{E}}[y_S \cdot y_{S'}] = 0$ 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 $z_S = y_S / \sqrt{\widetilde{\mathbf{E}}[y_S^2]}$, satisfying $\widetilde{\mathbf{E}}[z_S^2] = 1$.

We now address the obviously difficulty that $\widetilde{\mathbf{E}}[\cdot]$ 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 $\widetilde{\mathbf{E}}[y_S^2] < 0$. 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 $\widetilde{\mathbf{E}}[y_S^2] = 0$. 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" $\widetilde{\mathbf{E}}[\cdot]$ to be positive *semi*definite. We therefore need a different coping mechanism. For us, when $\widetilde{\mathbf{E}}[y_S^2] = 0$ 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 $\widetilde{\mathbf{E}}[z_S^2] = 1$. Thus we will have an additional burden: we will have to "manually" show that $\widetilde{\mathbf{E}}[y_S^2] = 0$ implies that y_S is orthogonal under $\widetilde{\mathbf{E}}[\cdot]$ 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 $\widetilde{\mathbf{E}}[z_S^2] = 0$. 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 4.6.6. 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:

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

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

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

Of course, if $\widetilde{\mathbf{E}}[y_S^2] < 0$ then we have encountered a positive definiteness problem. Indeed, to be conservative we will treat it as a problem if $\widetilde{\mathbf{E}}[y_{S,T}^2] < 0$ 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 $\widetilde{\mathbf{E}}[\cdot]$. However this relies on $\widetilde{\mathbf{E}}[z_T^2] = 1$, which fails for us if $T \in \text{PvZ}$. Thus we will need to explicitly prove that $T \in \text{PvZ}$ implies $\widetilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T] = 0$. 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 4.6.7. A *positive definiteness problem* occurs at substage (S,T) of modified Gram–Schmidt if either $\widetilde{\mathbf{E}}[y_{S,T}^2] < 0$, or if $T \in \text{PvZ}$ but $\widetilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T] \neq 0$. (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 4.6.8. *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 $T' \preceq T$;*
- $\widetilde{\mathbf{E}}[y_{S,T} \cdot z_{T'}] = 0$ *for all $T' \preceq T$, and hence $\widetilde{\mathbf{E}}[y_{S,T} \cdot q(x)] = 0$ for all polynomials $q(x)$ supported on monomials $x^{T'}$ with $T' \preceq T$;*
- $\widetilde{\mathbf{E}}[y_{S,T}^2] \geq 0$.

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

Our main Theorem 4.6.1 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

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

using Proposition 4.6.8.

4.6.3 Advanced accounting

Definition 4.6.9. 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 4.6.10. For a τ-subgraph$^+$ H, we extend the definition of revenue by assigning *two* credits for all isolated variable-vertices in H.

Remark 4.6.11. 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 4.6.12. *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 4.5.8 (here it does not matter that H is a τ-subgraph$^+$). □

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

Lemma 4.6.13. *In the setup of Lemma 4.6.12, 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. □

4.6.4 The key lemma

Lemma 4.6.14. *Let $y = y(x)$ be a polynomial expression of degree d. Assume $2d \leq \zeta \cdot \text{SMALL}$ and that $\widetilde{\mathbf{E}}[y \cdot p(x)] = 0$ 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*

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

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 4.6.3.)

Proof. Let us define

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

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

Claim 4.6.15. *H_{new} is small.*

Proof. Write $\mathrm{cl}(\mathrm{vbls}(H) \cup T) = H'_1 \cup \cdots \cup H'_t$ for small $(\mathrm{vbls}(H) \cup T)$-closed τ-subgraphs H'_i. Let $H'_{<j} \coloneqq H'_1 \cup \cdots \cup H'_{j-1}$, 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 4.6.12 to $H \cup H'_1$, $(H \cup H'_1) \cup H'_2$, $((H \cup H'_1) \cup H'_2) \cup H'_3$, ... to prove the claim. □

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

For the first of these, given an $(i, c) \in T$ we write $\overline{x}_i = 1_c(x_i) - \widetilde{\mathbf{E}}[1_c(x_i)]$ and $\overline{x}^T = \prod_{i \in T} \overline{x}_i$. Observe that $\overline{x}^T - x^T$ is a polynomial of degree strictly less than d; thus $\widetilde{\mathbf{E}}[y \cdot (\overline{x}^T - x^T)] = 0$ and so $\widetilde{\mathbf{E}}[y \cdot \overline{x}^T] \neq 0$. Now using Theorem 4.5.16 and $B \supseteq \mathrm{cl}(\mathrm{vbls}(H) \cup T) \supseteq \mathrm{cl}(\mathrm{vbls}(y \cdot \overline{x}^T))$, we conclude

$$\mathop{\mathbf{E}}_{B}[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T] \neq 0. \tag{4.10}$$

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

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

in contradiction to (4.10).

It remains to show that $R(H_{\mathrm{new}}) \leq R(H)$, which we will do using Lemma 4.6.13 (with $H' = B$, and $s = |T_{\mathrm{new}}|$, recalling that all of B's leaves are in $\mathrm{vbls}(H) \cup T$). We must show that the number of "boundary edges" — i.e., edges in B that have exactly one endpoint in H — is at least $|T_{\mathrm{new}}|$. Supposing otherwise, the set

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

would satisfy $|V| < |T_{\mathrm{new}}| + |T_{\mathrm{old}}| = |T| \leq d$. We will show that this contradicts (4.10).

Claim 4.6.16. *The deletion of variable-vertices V from B disconnects all variables in T from all variables in* $\mathrm{vbls}(H)$ *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 $\mathrm{vbls}(H)$ within B, as the vertices of T_{old} are already in V. Suppose $j \in T_{\mathrm{new}}$ is connected to some variable $i \in \mathrm{vbls}(H)$ by a path within B. Since $j \notin \mathrm{vbls}(H)$, 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 $\mathrm{vbls}(H)$ must pass through a variable in V. □

Recall that the proof is complete once we show that $|V| < d$ contradicts (4.10). Now

$$\begin{aligned}
\mathop{\mathbf{E}}_{B}[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T] &= \mathop{\mathbf{E}}_{B}\Big[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T \cdot \sum_{\vec{c} \in \Omega^V} \mathbf{1}[\boldsymbol{x}_i = c_i \ \forall i \in V]\Big] \\
&= \sum_{\vec{c} \in \Omega^V} \mathop{\mathbf{E}}_{B}\big[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T \cdot \mathbf{1}[\boldsymbol{x}_i = c_i \ \forall i \in V]\big].
\end{aligned} \tag{4.11}$$

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

$$\mathop{\mathbf{E}}_{B}\big[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T \cdot \mathbf{1}[\boldsymbol{x}_i = c_i \ \forall i \in V]\big] = \mathop{\mathbf{Pr}}_{B}[\boldsymbol{x}_i = c_i \ \forall i \in V] \cdot \mathop{\mathbf{E}}_{B}[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T \mid \boldsymbol{x}_i = c_i \ \forall i \in V].$$

By Claim 4.6.16 and the definition of the planted distribution η_B (and $\mathrm{vbls}(y) \subseteq \mathrm{vbls}(H)$), we have that $\boldsymbol{y}$ and $\overline{\boldsymbol{x}}^T$ are conditionally independent under η_B, conditioned on all $(\boldsymbol{x}_i : i \in V)$. Therefore

$$\mathop{\mathbf{E}}_{B}[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T \mid \boldsymbol{x}_i = c_i \ \forall i \in V] = \mathop{\mathbf{E}}_{B}[\boldsymbol{y} \mid \boldsymbol{x}_i = c_i \ \forall i \in V] \cdot \mathop{\mathbf{E}}_{B}[\overline{\boldsymbol{x}}^T \mid \boldsymbol{x}_i = c_i \ \forall i \in V].$$

Combining the previous two equations yields

$$\mathop{\mathbf{E}}_{B}\big[\boldsymbol{y} \cdot \overline{\boldsymbol{x}}^T \cdot \mathbf{1}[\boldsymbol{x}_i = c_i \ \forall i \in V]\big] = \mathop{\mathbf{E}}_{B}\big[\boldsymbol{y} \cdot \mathbf{1}[\boldsymbol{x}_i = c_i \ \forall i \in V]\big] \cdot \mathop{\mathbf{E}}_{B}[\overline{\boldsymbol{x}}^T \mid \boldsymbol{x}_i = c_i \ \forall i \in V].$$

Finally, using $|V| < d$ we will show that the first factor above is 0 (thereby establishing the claim that every term in (4.11) is 0, in contradiction to (4.10)). To see this, we have

$$\mathop{\mathbf{E}}_{B}\big[\boldsymbol{y} \cdot \mathbf{1}[\boldsymbol{x}_i = c_i \ \forall i \in V]\big] = \widetilde{\mathbf{E}}\big[y \cdot \prod_{i \in V} 1_{c_i}(x_i)\big]$$

because $\mathrm{cl}(\mathrm{vbls}(y) \cup V) \subseteq \mathrm{cl}(\mathrm{vbls}(H) \cup \mathrm{vbls}(B)) \subseteq B$, where we used Theorem 4.5.9. 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 $|V| < d$. □

4.6.5 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 4.6.17. A *witness* for substage $(S, T) \in \mathcal{M}_2^{\leq D}$ is defined to be a small τ-subgraph$^+$ $H_{S,T}$ with $\mathrm{vbls}(H_{S,T}) \supseteq \mathrm{vbls}(y_{S,T})$ and $R(H_{S,T}) \leq 2D$.

Remark 4.6.18. 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 4.6.19. *The existence of a witness $H_{S,T}$ for substage (S, T) implies $\widetilde{\mathbf{E}}[y_{S,T}^2] \geq 0$.*

Proof. By Lemma 4.6.12, we have that $\overline{H} := H_{S,T} \cup \mathrm{cl}(\mathrm{vbls}(y_{S,T}))$ is small. Thus $\widetilde{\mathbf{E}}[y_{S,T}^2] = \mathbf{E}_{\overline{H}}[\boldsymbol{y}_{S,T}^2] \geq 0$, using Theorem 4.5.16. □

We now come to our main technical theorem:

Theorem 4.6.20. *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 4.6.18 and Proposition 4.6.19 imply that $\widetilde{\mathbf{E}}[y_{S,T}^2] \geq 0$. 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 $H_{S,\mathrm{pr}(T)} \supseteq H_{S,\emptyset}$ for substage $(S,\mathrm{pr}(T))$. We now distinguish two cases.

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

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

First, suppose for contradiction that $T \in \mathrm{PvZ}$. We have that $y_{S,\mathrm{pr}(T)} = x^S - q(x)$ for some $q(x)$ supported on monomials $x^{T'}$ with $T' \preceq \mathrm{pr}(T) \prec T$. By Proposition 4.6.8 and induction, z_T is orthogonal to all such polynomials. Thus we deduce

$$0 \neq \widetilde{\mathbf{E}}[y_{S,\mathrm{pr}(T)} \cdot z_T] = \widetilde{\mathbf{E}}[x^S \cdot z_T] = \widetilde{\mathbf{E}}[x^S \cdot y_{T,\mathrm{pr}(T)}], \tag{4.12}$$

the last equality because $T \in \mathrm{PvZ}$ and hence $z_T = y_T = y_{T,\mathrm{pr}(T)}$. By induction (and using Remark 4.6.18), we have a witness $H_{T,\mathrm{pr}(T)}$ for $y_{T,\mathrm{pr}(T)}$. By Lemma 4.6.12 (using $2D + |S| \leq 3D \leq \zeta \cdot \mathrm{SMALL}$) we have that $\overline{H} \coloneqq H_{T,\mathrm{pr}(T)} \cup \mathrm{cl}(S)$ is small. (In writing $\mathrm{cl}(S)$ we used the abuse from Notation 4.6.3.). Now $\mathrm{vbls}(\overline{H}) \supseteq \mathrm{vbls}(H_{T,\mathrm{pr}(T)}) \cup S \supseteq \mathrm{vbls}(x^S \cdot y_{T,\mathrm{pr}(T)})$, so by Theorem 4.5.16 we have

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

the last equality because we're assuming $T \in \mathrm{PvZ}$. But the second identity above shows that $\boldsymbol{y}_{T,\mathrm{pr}(T)}^2$ is identically 0 under $\eta_{\overline{H}}$, meaning the first expression above must be 0. This contradicts (4.12).

Having ruled out the pseudovariance zero problem, we can complete the induction by finding a witness $H_{S,T} \supseteq H_{S,\mathrm{pr}(T)}$ for substage (S,T). By Proposition 4.6.8 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 $T' \preceq \mathrm{pr}(T)$. Furthermore, $y_{S,\mathrm{pr}(T)}$ is orthogonal to $p(x)$ under $\widetilde{\mathbf{E}}[\cdot]$. Thus, since we are in Case 2, we may deduce that

$$\widetilde{\mathbf{E}}[y_{S,\mathrm{pr}(T)} \cdot x^T] \neq 0. \tag{4.13}$$

We may now apply Lemma 4.6.14 (with $y = y_{S,\mathrm{pr}(T)}$, $H = H_{S,\mathrm{pr}(T)}$, and $r = 2D$) to obtain a small τ-subgraph$^+$ $H_{\mathrm{new}} \supseteq H_{S,\mathrm{pr}(T)}$ with $\mathrm{vbls}(H_{\mathrm{new}}) \supseteq \mathrm{vbls}(y_{S,\mathrm{pr}(T)}) \cup T$ and $R(H_{\mathrm{new}}) \leq 2D$.

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 $y_{S,T} = y_{S,\text{pr}(T)} - \widetilde{\mathbf{E}}[y_{S,\text{pr}(T)} \cdot z_T] z_T$. 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). □

4.7 Wrapping things up by setting parameters

To prove our main result on weak refutation, Theorem 1.5.5, we simply need to combine Theorems 4.4.12 and Theorem 4.6.1. Together these give us a pseudoexpectation defined up to degree

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

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 4.4.12 tells us that with high probability we get a pseudoexpectation satisfying Corollaries 4.5.18, 4.5.19. Thus we have established the following more precise version of Theorem 1.5.5:

Theorem 4.7.1. *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*

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

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 1.5.4, requires just a little work. We now imagine that our instance comes from a random $\text{CSP}(P^{\pm})$ as in Theorem 1.5.4. As discussed at the end of Section 4.2.1, 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 4.4.12 tells us moreover that

$$\#\{\text{nonempty } \tau\text{-subgraphs } H \text{ with } |\mathrm{cons}(H)| \le 2 \cdot \mathtt{SMALL} : I(H) \le \tau - 1\} \le \Delta n^{\frac{1+1/\log \Delta}{2}} = 2^{\frac{\log n}{2 \log \Delta}} \cdot \frac{m}{\sqrt{n}}. \tag{4.14}$$

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

Supposing (4.14) 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 $\mathrm{cl}(S) \supseteq H_f$, but we may ask whether $\mathrm{cl}(S)$ 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 $|\mathrm{cons}(H')| \le 2 \cdot \mathtt{SMALL}$. Furthermore, the number of leaf variables in H' must be at least 1 (else H' is $\emptyset$-closed and hence empty by Fact 4.5.6) and strictly less than K (else $H \setminus H_f$ will be $\emptyset$-closed and hence empty). Finally, we claim $R(H') \le \tau - 1$. 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 4.6.13.) We conclude that whenever $\mathrm{cl}(N(f)) \ne H_f$, there must exist a nonempty τ-subgraph H' with the following properties: (i) $|\mathrm{cons}(H')| \le 2 \cdot \mathtt{SMALL}$; (ii) $I(H') \le R(H') \le \tau - 1$; (iii) H' has at least one leaf variable; (iv) all leaves of H' are adjacent to f.

But (4.14) 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:

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

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

$$\widetilde{\mathbf{E}}\Big[1[x \text{ satisfies } f]\Big] = \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \mu_f}\Big[1[\boldsymbol{x} \text{ satisfies } f]\Big] \ge 1 - \delta.$$

Combining the last two deductions yields

$$\widetilde{\mathbf{E}}\Big[\text{fraction of constraints satisfied}\Big] \ge 1 - \delta - 2^{\frac{\log n}{2 \log \Delta}} \cdot \frac{1}{\sqrt{n}}.$$

In summary, we have proven the following more precise version of Theorem 1.5.4:

Theorem 4.7.2. *Let P be a k-ary Boolean predicate and let $1 < t \le k$. Let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P^{\pm})$ 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*

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

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 4.7.3. We should mention that in our δ-refutation result Theorem 4.7.2, our pseudoexpectation does *not* satisfy "solution value $= 1 - \delta_0$" as a constraint for any $\delta_0 \le \delta$; it merely has $\widetilde{\mathbf{E}}[\text{solution value}] \ge 1 - \delta$. Achieving the (stronger) former condition is a direction for future work. By contrast, for our weak refutation result Theorem 1.5.5, the pseudoexpectation *does* satisfy all the constraints and hence also satisfies $\widetilde{\mathbf{E}}[\text{solution value}] = 1$ as a constraint.

4.8 Proof that random graphs satisfy the Plausibility Assumption

Here we prove Theorem 4.4.12, which we restate for convenience:

Theorem 4.4.12 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 $\boldsymbol{G}$ is a random instance with $m = \Delta n$ constraints, the Plausibility Assumption holds provided*

$$\text{SMALL} \leq \gamma \cdot \frac{n}{\Delta^{2/(\lambda-\zeta)}}, \tag{4.15}$$

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

$$\#\{\text{nonempty } \tau\text{-subgraphs } H \text{ with } \text{cons}(H) \leq 2 \cdot \text{SMALL} : I(H) \leq \tau - 1\} \leq \Delta n^{\frac{1+\zeta}{2}}. \tag{4.16}$$

Proof. A remark before we begin: the expression in (4.15) was chosen precisely so that

$$c \leq 2 \cdot \text{SMALL} \implies 20^K \cdot \Delta \cdot \left(\tfrac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2}} \leq \beta/50^K, \tag{4.17}$$

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

$$\mathbf{E}[\#\{\tau\text{-subgraphs with } c \text{ constraints, } v \text{ vertices, and income at most } I_0\}]. \tag{4.18}$$

There are $\binom{m}{c}$ choices for the constraints and $\binom{n}{v}$ choices for the variables. Then by using Lemma 4.4.11,

$$(4.18) \leq \binom{m}{c}\binom{n}{v} \mathbf{Pr}[\text{fixed set of } c \text{ constraints and } v \text{ variables gets at least } A \text{ edges}], \tag{4.19}$$

where $A \coloneqq \frac{\tau-\zeta}{2} \cdot c + v - \frac{I_0}{2}$. In (4.19), 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

$$\begin{aligned}(4.19) \leq 2^{Kc}\binom{m}{c}\binom{n}{v}\left(\frac{v}{n}\right)^A \leq 2^{Kc}\left(\frac{em}{c}\right)^c\left(\frac{en}{v}\right)^v\left(\frac{v}{n}\right)^A &= \left(e2^K e^{v/c}(v/c)\right)^c \cdot \Delta^c \cdot \left(\frac{v}{n}\right)^{\frac{\lambda-\zeta}{2}\cdot c - I_0/2} \\ &\leq \left(20^K\right)^c \cdot \Delta^c \cdot \left(\tfrac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2}\cdot c - I_0/2}, \end{aligned} \tag{4.20}$$

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

$$(4.18) \leq (4.20) = \left(20^K \cdot \Delta \cdot \left(\tfrac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2}}\right)^c \leq \left(\tfrac{\beta}{50^K}\right)^c,$$

using (4.17). Summing over the at most Kc possibilities for v gives

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

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

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

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

$$(4.18) \leq (4.20) = \left(20^K \cdot \Delta \cdot \left(\tfrac{Kc}{n}\right)^{\frac{\lambda-\zeta}{2}}\right)^{c-1} \cdot 20^K \cdot \Delta \cdot \left(\tfrac{n}{Kc}\right)^{\frac{1+\zeta}{2}} \leq \left(\tfrac{\beta}{50^K}\right)^{c-1} \cdot 20^K \cdot \Delta n^{\frac{1+\zeta}{2}},$$

again using (4.17). 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

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

and again Markov's inequality establishes that (4.16) holds except with probability at most β. □

Chapter 5

Directions for future work

5.1 Upper bounds for more general random CSP models

It would be interesting to show analogous efficient refutation results for models of random CSP(P) in which literals are not used. This would allow for results on, say, refuting q-colorability for random k-uniform hypergraphs. For some predicates (e.g., monotone Boolean predicates), random CSP instances are trivially satisfiable when there are no literals. However for such predicates one could consider a "Goldreich [Gol00]-style" model in which each constraint is randomly either P or $\neg P$ applied to k random variables.

5.2 Upper bounds for refutation of semirandom CSPs

5.2.1 Previous work: Feige's semirandom model for 3-SAT

The study of refutation of semirandom CSPs was initiated by Feige [Fei07]. He studied semirandom 3-SAT instances generated by starting with an arbitrary 3-SAT instance and flipping the polarity of each literal in every clause independently with probability ϵ, called the noise. When $m = \Omega(\epsilon^{-3}n)$, a Chernoff Bound argument shows that such a semirandom instance is unsatisfiable with high probability. Feige showed that unsatisfiability can also be efficiently certified when m is large enough.

Theorem 5.2.1 ([Fei07, Theorems 1.1 and 1.2])**.** *For every $n^{-1.4} < \epsilon \leq 1/2$, there exists a constant c such that a semirandom instance with noise value ϵ and $m \geq c\epsilon^{-2}n^{1.5}\sqrt{\log\log n}$ can be refuted in polynomial time with high probability over the random choice of the instance and the random choices of the algorithm.*

5.2.2 Future work: Generalizing to arbitrary CSPs

We can also study Feige's semirandom model for a CSP with an arbitrary k-ary predicate P. Again, we begin with an instance of CSP(P) and flip every literal independently with probability ϵ. It is then natural to ask whether such instances can be refuted when $m \gg n^{k/2}$.

Based on the framework of Chapter 3, it suffices to strongly refute induced XOR instances, although these induced XOR instances are now *semirandom*. Strong refutation of semirandom k-XOR instances would allow us to extend the results of [AOW15] to the semirandom case, implying $\delta_P(t)$-refutation of semirandom CSP(P) instances when $m \gg n^{t/2}$.

Semirandom XOR refutation in the even arity case. For even arity, we can strongly refute a semirandom k-XOR instance when $m \geq \widetilde{O}(n^{k/2})$ using an SDP.

Proposition 5.2.2. *Let k be even and let $\mathcal{I}$ be a semirandom k-XOR instance. Then there exists a polynomial time algorithm that certifies that* $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + o(1)$ *when* $m \geq \widetilde{O}(n^{k/2})$.

We sketch the proof. When $m = \omega(\epsilon^{-k}n)$, our instance has value $1/2 + o(1)$. As in the proof of Theorem 3.2.1, we can reduce to 2-XOR by grouping the variables in each constraint two blocks of $k/2$ variables each and replacing each block with a new variable. We thereby obtain a 2-XOR instance with $n^{k/2}$ variables. Charikar and Wirth proved that by solving an SDP, we can closely approximate the value of a 2-XOR instance.

Theorem 5.2.3 ([CW04, Lemma 5])**.** *If the optimal value of a* 2-XOR *instance is $\frac{1}{2} + \delta$, then the value of the SDP relaxation for the instance is at least $\frac{1}{2} + \Omega\left(\frac{\delta}{\log(1/\delta)}\right)$.*

Given that our 2-XOR instance has value $1/2+\delta$, we can certify that it has value at most $1/2+\delta \log n$. Choosing $m \geq \widetilde{O}(n^{k/2})$, we can make δ small enough that this is $1/2 + o(1)$.

Semirandom XOR refutation in the odd arity case. To obtain refutation of semirandom instances of CSP(P) with $m \geq \widetilde{O}(n^{k/2})$, it remains to show that we can refute semirandom XOR instances with odd arity. Interestingly, this case seems to be much more challenging than the even arity case.

Question 5.2.4. Is it possible to strongly refute semirandom k-XOR instances with odd arity when $m \geq \widetilde{O}(n^{k/2})$?

We will outline one approach to this question: Taking the an algorithm that works in the random and showing that it works in the semirandom case as well.

An algorithm for random instances. As mentioned in Chapter 1, Barak and Moitra showed that this is possible in the case of random instances [BM16]. Here is a high-level view of their algorithm:

1. **Construct a $(2k-2)$-XOR instance** Choose an index $\ell \in [k]$. Create a $(2k-2)$-XOR instance by adding constraints that have the same variable in position ℓ. For concreteness, we will set $\ell = k$. Given the constraints

$$x_{i_1} \oplus \cdots \oplus x_{i_{k-1}} \oplus x^* = b_i \qquad \text{and} \qquad x_{j_1} \oplus \cdots \oplus x_{j_{k-1}} \oplus x^* = b_j,$$

derive the constraint

$$x_{i_1} \oplus \cdots \oplus x_{i_{k-1}} \oplus x_{j_1} \oplus \cdots \oplus x_{j_{k-1}} = b_i \oplus b_j. \tag{5.1}$$

Let $R(\mathcal{I})$ be the $(2k-2)$-XOR instance constructed in this way from the k-XOR instance $\mathcal{I}$.

2. **Construct a** 2-XOR **instance** Form a 2-XOR instance on n^{k-1} variables (labeled y_S for $S \in [n]^{k-1}$) by splitting the $2k-2$ variables of each constraint into two tuples of size $k-1$ and replacing each of the resulting sums of $k-1$ variables with its corresponding new variable. In particular, given constraint (5.1), we construct the 2-XOR constraint

$$y_{i_1,\ldots,i_{\frac{k-1}{2}},j_1,\ldots,j_{\frac{k-1}{2}}} \oplus y_{i_{\frac{k-1}{2}+1},\ldots,i_{k-1},j_{\frac{k-1}{2}+1},\ldots,j_{k-1}} = b_i \oplus b_j. \tag{5.2}$$

Though this way of splitting the variables may look unnatural, it is necessary, as the instance with equations of the form $y_{i_1,\ldots,i_{k-1}} \oplus y_{j_1,\ldots,j_{k-1}} = b_i \oplus b_j$ has value $1 - o(1)$ with high probability. Call this 2-XOR instance $M(R(\mathcal{I}))$.

3. **Refute the 2-XOR instance** This is usually done by taking the spectral norm of the corresponding quadratic form or solving the corresponding SDP relaxation and using Theorem 5.2.3.

To show that this algorithm works, two steps are required. First, we need to show that strong refutation of $M(R(\mathcal{I}))$ implies strong refutation of $\mathcal{I}$.

Theorem 5.2.5 ([BM16])**.** *Let k be odd and let $\mathcal{I}$ be a random k-XOR instance. Then if* $\mathrm{Opt}(M(R(\mathcal{I})) \leq \frac{1}{2} + \delta$*, then* $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + O(\delta)$*.*

For an instance $\mathcal{J}$, it is clear that if $\mathrm{Opt}(M(\mathcal{J})) \leq \frac{1}{2} + \delta$, then $\mathrm{Opt}(\mathcal{J}) \leq \frac{1}{2} + \delta$; a solution to $\mathcal{J}$ corresponds to a solution to $M(\mathcal{J})$ satsifying the same number of constraints. The interesting part is to show that for random instances, if $\mathrm{Opt}(R(\mathcal{I})) \leq \frac{1}{2} + \delta$, then $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + O(\delta)$.

The second part of the proof is to show that the $M(R(\mathcal{I}))$ instance has small value when the instance is random.

Theorem 5.2.6 ([BM16])**.** *Let k be odd and let $\mathcal{I}$ be a random k-XOR instance. Then with high probability* $\mathrm{Opt}(M(R(\mathcal{I}))) \leq \frac{1}{2} + o(1)$ *when* $m \geq \widetilde{O}(n^{k/2})$*.*

The semirandom model? Feige showed that the $(2k-2)$-XOR instance induced by a semirandom k-XOR instance also has small value [Fei15].

Proposition 5.2.7. *Let $\mathcal{I}$ be a semirandom k-XOR instance with $m = \omega(n)$. Then, with high probability,* $\mathrm{Opt}(R(\mathcal{I})) \leq \frac{1}{2} + o(1)$*.*

However, we don't know how to prove that $M(R(\mathcal{I}))$ has value $\frac{1}{2} + o(1)$ for semirandom $\mathcal{I}$.

Question 5.2.8. When $\mathcal{I}$ is semirandom, does it hold that $\mathrm{Opt}(M(R(\mathcal{I}))) \leq \frac{1}{2} + o(1)$?

The argument we used in the even k case no longer works. Given a fixed assignment to the variables of $M(R(\mathcal{I}))$, each constraint is no longer satisfied independently of the others. The RHS of each 2-XOR constraint depends on the random RHS's of both k-XOR constraints from which it was derived.

We also don't know how to show that refuting $M(R(\mathcal{I}))$ is sufficient to refute $\mathcal{I}$.

Question 5.2.9. If $\mathrm{Opt}(M(R(\mathcal{I}))) \leq \frac{1}{2} + \delta$, does it hold that $\mathrm{Opt}(\mathcal{I}) \leq \frac{1}{2} + f(\delta)$ for some function f?

These two questions could be good steps toward resolving Question 5.2.4 and obtaining refutation algorithms for semirandom instances of $\mathrm{CSP}(P)$.

5.3 Understanding nondeterministic refutation

In this section, we consider the question of determining the number of constraints required for *nondeterministic* refutation of random instances of $\mathrm{CSP}(P)$. First, we look at upper bounds. Second, we would like to prove lower bounds on number of constraints required for nondeterministic refutation in specific proof systems, including Resolution, SOS, and Cutting Planes.

5.3.1 Upper bounds

Nondeterministic refutation with $o(n^{1.4})$ constraints. Feige, Kim, and Ofek showed that nondeterministic refutation of 3-SAT instances with many fewer than $n^{\mathcal{C}(k\text{-SAT})} = n^{1.5}$ constraints is possible [FKO06].

Theorem 5.3.1 ([FKO06])**.** *Given a random* 3-SAT *instance with $O(n^{1.4})$ constraints, there exists a polynomial-size certificate of unsatisfiability with high probability.*

Question 5.3.2. Is nondeterministic refutation of 3-SAT possible for any $m = o(n^{1.4})$?

Completed work: Extending FKO to other binary predicates. With Feige, we showed that we can extend Theorem 5.3.1 to arbitrary k-ary predicates P to get nondeterministic refutation of random instances of $\mathrm{CSP}(P)$ with $m = o(n^{k/2})$.

Theorem 5.3.3 ([FW15])**.** *Let P be any predicate of arity k and let $\mathcal{I}$ be a random instance of $\mathrm{CSP}(P)$ with $m \geq \widetilde{O}\left(n^{\frac{k}{2}-\frac{k-2}{2(k+2)}} \log^5 n\right)$. Then with probability at least 0.9, there exists an $\widetilde{O}\left(n^{\frac{k}{2}-\frac{k-4}{2(k+2)}}\right)$-size witness that $\mathcal{I}$ is unsatisfiable. Moreover, such a witness can be found in time $2^{\widetilde{O}(n^{1/(k+2)})}$.*

In light of Corollary 1.5.2, we observe that this theorem is only interesting for k-ary predicates that support a $(k-1)$-wise uniform distribution over satisfying assignments. The proof is a generalization as Feige, Kim, and Ofek's proof of Theorem 5.3.1 to arity-k predicates; we sketch it here.

Proof sketch. First, we reduce weak refutation of $\mathrm{CSP}(P)$ to δ-refutation of k-XOR. Then we use the FKO strategy to δ-refute k-XOR with $o(n^{k/2})$ constraints. As in Chapter 3, we will work in the "$G(n,p)$-style" model in which each possible constraint is included independently with probability p and we let $\overline{m} = pn^k$ be the expected number of constraints. We will again use the notation $\mathbf{E}[P] = \mathbf{E}_{\boldsymbol{z}\sim\{\pm1\}^k}[P(\boldsymbol{z})]$.

Reduction to δ-refutation of k-XOR. Assume for a contradiction that there exists a satisfying assignment x^*; we consider its induced distribution $\mathcal{D}_{\mathcal{I},x^*}$. First, we use Theorem 3.2.8 to certify that $\mathcal{I}$ is $(\delta, k-1)$-quasirandomness for some $\delta = o(1)$. By Lemma 2.1.13, this implies that there exists a $(k-1)$-wise uniform distribution $\mathcal{D}$ such that $d_{\mathrm{TV}}(\mathcal{D}_{\mathcal{I},x^*}, \mathcal{D}) \leq 2^k\delta$.

The only $(k-1)$-uniform distributions are the uniform distribution on k bits $\mathcal{U}_k$, the uniform distribution on satisfying assignments to k-XOR (uniform distribution on odd-parity k-bit strings), and the uniform distribution on satisfying assignments to the negation of k-XOR (uniform distribution on even-parity k-bit strings). If $\mathcal{D}$ were $\mathcal{U}_k$, then $\mathcal{D}_{\mathcal{I},x^*}$ would be $2^k\delta$-close to the uniform distribution, and x^* would satisfy at most a $\mathbf{E}[P] + 2^k\delta < 1$ fraction of constraints.

Since x^* is a satisfying assignment, it must therefore be $2^k\delta$-close to either the uniform distribution over satisfying assignments to k-XOR or the uniform distribution over satisfying assignments to the negation of k-XOR. If we $2^k\delta$-refute k-XOR and its negation, then we have a contradiction and x^* cannot be a satisfying assignment. More formally, the following stronger claim holds.

Claim 5.3.4. *Let P be any predicate of arity k. If there exists an ϵ-refutation algorithm for k-XOR, then there exists an ϵ'-refutation algorithm for $\mathrm{CSP}(P)$, where*

$$\epsilon' = 2|\widehat{P}([k])|\epsilon - \eta$$

and

$$\eta = O_k(\overline{m}^{-1/2} n^{\frac{k-1}{4}} \log^{5/2} n).$$

Refuting k-XOR with $o(n^{k/2})$ constraints. Claim 5.3.4 implies that to weakly refute CSP(P), it suffices to δ-refute k-XOR for $\delta = \Theta_k(\overline{m}^{-1/2} n^{\frac{k-1}{4}} \log^{5/2} n)$. We will do this exactly as in [FKO06]. Thinking of a k-XOR instance as a k-uniform hypergraph, we wish to show that there exists a collection of t 2-regular subhypergraphs $\{R_1, \ldots, R_t\}$ such that the following hold:

1. Each R_i contains r vertices.
2. Each vertex in each R_i has degree 2, so each R_i contains $\frac{2r}{k}$ hyperedges.
3. Each hyperedge occurs in at most d of the R_i's.
4. If $\frac{2r}{k}$ is even, then for every i the total number of negated literals occurring in constraints of R_i is odd. If $\frac{2r}{k}$ is odd, then for every i the total number of negated literals occurring in constraints of R_i is even.

Because of Condition 4, each R_i must have at least one violated constraint. Then Condition 3 implies that at least $\frac{t}{d}$ constraints must be violated. We can ignore Condition 4: a constant fraction of subhypergraphs satisfying the other three conditions will also satisfy this condition. Let $\mathcal{H}(n, p, k)$ be the distribution over k-uniform hypergraphs in which each hyperedge is included independently with probability p. To prove the theorem, it suffices to show the following:

Lemma 5.3.5. *Let $\overline{m} \geq O\left(n^{\frac{k}{2} - \frac{k-2}{2(k+2)}} \log^5 n\right)$ and $r = O\left(n^{\frac{1}{k+2}}\right)$. With probability at least 0.9, $H \sim \mathcal{H}(n, p, k)$ has a set of $\Theta(\overline{m})$ 2-regular subhypergraphs on r vertices such that every hyperedge is contained in at most $\frac{\overline{m}}{r}$ elements.*

This can be proven in exactly the same way as the $k = 3$ case in [FKO06]. The algorithmic statement in the theorem also follows exactly as in [FKO06]. □

FKO for larger alphabets. We can also ask if we can find FKO-style nondeterministic refutations when the alphabet is larger: instead of considering $P : \{0,1\}^k \to \{0,1\}$, we study predicates $P : [q]^k \to \{0,1\}$.

Question 5.3.6. Let $P : [q]^k \to \{0,1\}$. Do nondeterministic refutations of random instances of CSP(P) exist when $m \ll n^{k/2}$ exist for $P : [q]^k \to \{0,1\}$?

Interactive refutation. We can also study stronger proofs. In particular, we can instead ask if there exists a constant-round interactive proof refuting a random instance of 3-SAT with $m = o(n^{1.4})$.

Question 5.3.7. Does there exist a constant-round interactive proof refuting a random 3-SAT instance with $m = o(n^{1.4})$?

In other words, is refutation of such instances in the class AM? In standard nondeterministic refutation, we wanted to show that refuting a random 3-SAT instance, which is trivially in coNP, is also in NP.

Algebraic methods for refutation. SOS requires about $n^{\frac{k}{2}}$ constraints to refute random instances of k-XOR in polynomial time [Sch08], but it is easy to see that a random instance with only $O(n)$ constraints can be refuted in polynomial time using Gaussian elimination. In addition, if P is a degree-d polynomial over $\mathbb{F}_2$, then we can use Gaussian elimination to refute instances with $O(n^d)$ constraints by introducing new variables for every product of up to d variables and performing Gaussian elimination on the linearized instance.

Applebaum and Lovett [AL16] recently showed that the reach of algebraic techniques extends far beyond these cases. They considered the slightly different "Goldreich's function" model, in which the scope of each constraint is still chosen randomly but there are no random negations on inputs to constraints. Instead, each constraint has a random right-hand side. For example, the constraints $P(x_S) = 0$ and $P(x_S) = 1$ are each included with equal probability. In this model, they study two parameters of predicates.

1. Bit-fixing degree. P has r-bit fixing degree e if there exists an assignment to r inputs in P such that the resulting restriction has degree e over $\mathbb{F}_2$.
2. Rational degree. The rational degree of P is the minimum degree of a non-zero polynomial Q over $\mathbb{F}_2$ such that Q covers the roots of P or P's complement.

They prove that when either of these quantities is small, polynomial-time refutation is possible using algebraic methods.

Theorem 5.3.8 ([AL16])**.**

1. *If P has r-bit fixing degree e, then a random instance of* $\mathrm{CSP}(P)$ *in the Goldreich's function model can be refuted when $m \gg n^{r+e}$.*
2. *If P has rational degree e, then a random instance of* $\mathrm{CSP}(P)$ *in the Goldreich's function model can be refuted when $m \gg n^{e}$.*

Applebaum and Lovett also point out that the $2r$-ary predicate $P(z) = z_1 + \cdots + z_r + \prod_{i=r+1}^{2r} z_i$ has $\mathcal{C}(P) = r - 1$ but has 1-bit fixing degree equal to 1. For this predicate, algebraic techniques should allow for polynomial-time refutation at much lower densities than SOS. We ask the following two questions.

Question 5.3.9.

1. Can Theorem 5.3.8 be proven in the uniform random model?
2. Is there a parameter that generalizes and unifies bit fixing degree and rational degree and for which we can prove refutation results? (asked in [AL16])

5.3.2 Size lower bounds

In some cases, width or degree lower bounds imply lower bounds on the size of proofs. These statements often have the following form.

$$\text{If a refutation has width or degree } r\text{, then it must have size } 2^{\Omega(r^2/n)}. \tag{5.3}$$

Examples of such theorems include Impagliazzo et al.'s connection between degree and size in polynomial calculus [IPS99], Ben-Sasson and Wigderson's connection between Resolution size and width [BSW99], Grigoriev et al.'s lower bound on size of SOS refutations of knapsack contradictions based on degree lower bounds [GHP02], and Kojevnikov and Itsykson's proof that high degree implies large size for SOS refutations of Tseitin contradictions [KI06].

Understanding the relationship between size and width in Resolution refutations of random k-SAT. In the Resolution proof system, introduced by Robinson [Rob65], each line of a proof is a disjunction of literals. We begin with a set of clauses, which we can think of as the clauses of an unsatisfiable random SAT instance, and apply the following rules.

$$\frac{A}{A \vee B} \qquad \frac{A \vee x \quad A \vee \neg x}{A \vee B}$$

We can refute an instance if we can derive the empty clause from the clauses of the instance using these two rules. The size of a Resolution proof is the number of clauses it contains. The width of a clause is the number of literals it contains. The width of a Resolution proof is the maximum width of any clause in the proof.

Resolution is probably the most well-studied proof system. Nevertheless, we still do not know how many constraints are required for nondeterministic refutation of k-SAT in Resolution. Ben-Sasson proved superpolynomial size lower bounds for Resolution when the number of clauses is almost $n^{k/2}$ [BS01] by proving that the rank is $\omega(\sqrt{n})$ in this regime.

Theorem 5.3.10. *Any Resolution refutation of an instance of random k-SAT with $m = n^{k/2-\epsilon}$ requires width $\Omega(n^{\frac{1}{2}+\frac{\epsilon}{4(k-2)}})$.*

Using Ben-Sasson and Wigderson's connection between size and width [BSW99], the size lower bound follows [BS01].

Theorem 5.3.11. *Any Resolution refutation of an instance of random k-SAT with $m = n^{k/2-\epsilon}$ requires size $2^{n^{\Omega(\epsilon)}}$.*

However, Ben-Sasson's rank lower bounds are superconstant for number of constraints almost all the way up to n^{k-1}.

Theorem 5.3.12. *Any Resolution refutation of an instance of random k-SAT with $m = n^{k-1-\epsilon}$ constraints requires width $\Omega(n^{\frac{\epsilon}{k-2}})$.*

Furthermore, we know that Resolution refutations can be found in polynomial time when $m = O(n^{k-1}/\log^{k-2} n)$ [BKPS98, BKPS02].

Theorem 5.3.13 ([BKPS02, Theorem 6.1])**.** *With high probability, a random instance of k-SAT with $m = O(n^{k-1}/\log^{k-2} n)$ has a polynomial-size Resolution proof of unsatisfiability. Furthermore, this proof can be found in polynomial time.* [1]

We can ask what happens for number of constraints between $n^{k/2}$ and n^{k-1}.

Question 5.3.14. Which of the following is true?

1. The rank of any Resolution refutation of a random instance of k-SAT with $n^{k-1-\epsilon}$ constraints is at least $n^{1/2+\Omega(\epsilon)}$.
2. The rank of any Resolution refutation of a random instance of k-SAT with $n^{k-1-\epsilon}$ constraints is smaller than $\sqrt{n}$ but the size of any refutation is still superpolynomial.
3. There exists a polynomial-size Resolution refutation of a random instance of k-SAT with $m \ll n^{k-1-\epsilon}$.

In this case, we know that there exist formulas based on ordering principles requiring $\Omega(n)$ Resolution width but having polynomial-size Resolution proofs [BG01, AD08]. This means that size-width tradeoff (5.3) cannot be improved in general. However, it is possible that in the special case of random k-SAT, stronger size lower bounds based on width can be proven. On the other hand, if (5.3) is optimal for random k-SAT, then the third case must hold.

[1] In fact, this proof can be found using ordered DLL, a restriction of Resolution. See Beame et al.'s introduction for details [BKPS02].

Understanding the relationship between size and degree in SOS. To lower bound the number of constraints required for nondeterministic refutation in a proof system, we need to prove size lower bounds. If we use (5.3), we need degree $\Omega(n^{1/2+\epsilon})$ to get superpolynomial size lower bounds. A more careful analysis of the proof of the degree lower bounds in [Sch08] gives conditions under which the techniques used in these works allow such a strong degree lower bound.

Theorem 5.3.15 ([Sch08])**.** *Let P be a k-ary predicate whose satisfying assignments include all those of k-XOR or all those of its negation. For all $\epsilon > 0$, an SOS refutation of a random instance of $\mathrm{CSP}(P)$ with $n^{t/6+2/3-\epsilon}$ constraints requires degree $n^{1/2+\delta(\epsilon)}$ with high probability.*

As mentioned above, (5.3) holds for semialgebraic proof systems and we can derive the following size lower bounds.

Theorem 5.3.16. *Let P be a k-ary predicate whose satisfying assignments include all those of k-XOR or all those of its negation. For all $\epsilon > 0$, an SOS refutation of a random instance of $\mathrm{CSP}(P)$ with $n^{t/6+2/3-\epsilon}$ constraints requires size $2^{n^{\Omega(\epsilon)}}$ with high probability.*

We get superpolynomial size lower bounds up to much lower densities than those at which we get superconstant rank lower bounds. We can then ask the following question.

Question 5.3.17. Let P be a k-ary predicate whose satisfying assignments include all those of k-XOR or all those of its negation. Which of the following is true?

1. The degree of any SOS refutation of a random instance of $\mathrm{CSP}(P)$ with $n^{k/2-\epsilon}$ constraints is at least $n^{1/2+\Omega(\epsilon)}$.
2. The degree of a SOS refutation of a random instance of $\mathrm{CSP}(P)$ with $n^{k/2-\epsilon}$ constraints is smaller than $\sqrt{n}$ but the size of any refutation is still superpolynomial.
3. There exists a polynomial-size static SOS refutation of a random instance of $\mathrm{CSP}(P)$ with $m \ll n^{k/2-\epsilon}$.

If either of the first two alternatives is true, then for $m \gg n^{k/2}$, low-rank, polynomial-size refutations of $\mathrm{CSP}(P)$ exist and for $m \leq n^{k/2-\epsilon}$, no polynomial-size refutations exist. In the first case, the degree quickly drops from $\sqrt{n}$ to a constant as m increases from $n^{k/2-\epsilon}$ to $n^{k/2} \cdot \mathrm{polylog}(n)$. Proving the second case would require new techniques for relating size and degree in SOS. If the third alternative holds, then as the number of constraints increases, we pass through three phases: For small enough m, not even nondeterministic refutation is possible. For intermediate values of m, nondeterministic refutation is possible, but low-degree, polynomial time refutation is not possible For large values of m, low-degree, polynomial-time refutation is possible.

For comparison, Atserias et al. showed that there exist formulas based on the Pigeonhole Principle contradiction that require SA degree n^{δ} and size $n^{\Omega(n^{\delta})}$ for some constant $\delta > 0$ [ALN14]. Lauria and Nördstrom extended this result to SOS [LN15]. This shows that, in general, size $n^{O(r)}$ is necessary and sufficient for degree r proofs. The above question can be thought of asking whether this tight size-degree relation also holds for random CSPs: We know that degree-r static semialgebraic proofs require size at least $2^{\Omega(r^2/n)}$ and that $n^{O(r)}$ size is sufficient; our goal is to close this gap.

Lower bounds for dynamic semialgebraic proof systems. We know comparatively few lower bounds for dynamic semialgebraic proof systems. Indeed, such proof systems can be very powerful, having the ability to, for example, simuate Gaussian elimination [Ats15].

We propose studying Cutting Planes as a first step toward understand the power of dynamic semialgebraic refutation. In contrast with other dynamic semialgebraic proof systems, superpolynomial size lower bounds are known for both Syntactic [Pud97] and Semantic [FHL] Cutting Planes.

Question 5.3.18. Prove size lower bounds for syntactic or semantic Cutting Planes refutations of random 3-SAT with $m = O(n)$.

We could first try to prove this result for a weaker form of Cutting Planes. One way of doing this is to restrict the values that coefficients in the proof can take. Large coefficients seems to make it harder to prove lower bounds (e.g., [BPR97]) and easier to prove upper bounds (e.g., [FHL]). We therefore want to start by studying Cutting Planes proofs in which all coefficients are in $\{-1, 0, 1\}$.

Question 5.3.19. Prove size lower bounds for Syntactic or Semantic Cutting Planes refutations of random 3-SAT with $m = O(n)$ when all coefficients are in $\{-1, 0, 1\}$.

One possible route toward proving this is to define an appropriate notion of width and try to prove that short proofs must have low width as in [IPS99, BSW99].

Question 5.3.20. Is there a notion of width for Syntactic Cutting Planes for which short proofs must have low width?

One candidate for such a measure is rank. Atserias et al. [ABL03] showed that low rank implies short proofs: If a set of axioms has rank d, then there exists a Syntactic Cutting Planes refutation of size $O(n^d)$. However, short proofs do not necessarily have low rank: They also show that there exist formulas of rank $n^{\Omega(1)}$ with refutations of size $n^{O(1)}$. This result means that size lower bounds in Syntactic Cutting Planes cannot be proven using this notion of rank.

Another idea is apply Haken's bottleneck counting method [Hak85]. Haken introduced this technique to prove the first superpolynomial lower bounds on the size of Resolution proofs. For every assignment, there must be a line of the proof that is violated. In the bottleneck counting method, we find a large set of assignments to the variables such that there are not many possible proof lines that can falsify a single assignment. Many of these falsifying proof lines must appear in any refutation, so any refutation must be large. Perhaps the bottleneck counting can be applied to Cutting Planes as well.

Bibliography

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, pages 496–505, 2007. 2.1.2

[AAM+11] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Inapproximability of densest κ-subgraph from average case hardness. 2011. 1.2.1, 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, 1.4

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986. 3.3.5

[ABL03] Albert Atserias, María Luisa Bonet, and Jordi Levy. On Chvátal Rank and Cutting Planes Proofs. *Electronic Colloquium on Computational Complexity (ECCC)*, (041), 2003. 5.3.2

[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. 1.2.2, 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, 1.2.2

[AD08] Albert Atserias and Víctor Dalmau. A combinatorial characterization of resolution width. *J. Comput. System Sci.*, 74(3):323–334, 2008. 5.3.2

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Inform. Process. Lett.*, 88(3):107–110, 2003. 2.1.13, 3.2.3

[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. 1.2.1

[AH09] Per Austrin and Johan Håstad. Randomly supported independence and resistance. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 483–492. 2009. 3.1.2, 3.1.2

[AH13] Per Austrin and Johan Håstad. On the usefulness of predicates. *Transactions on Computation Theory*, 5(1):1, 2013. 2.1

[AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC^0. *SIAM Journal on Computing*, 36(4):845–888, 2006. 1.2.2

[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. 1.2.2, 1, 5.3.1, 5.3.8, 2

[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. 1.2.2

[Ale05] Michael Alekhnovich. Lower bounds for k-DNF resolution on random 3-CNFs. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 251–256. 2005. 1.4

[ALN14] Albert Atserias, Massimo Lauria, and Jakob Nordström. Narrow proofs may be maximally long. In *Proceedings of the 29th Annual Conference on Computational Complexity*, pages 286–297, 2014. 5.3.2

[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), 1.5.1, 1.5.2, 1.5.1, 1.5.2, 1.5.2, 5.2.2

[App12] Benny Applebaum. Pseudorandom generators with long stretch and low locality from random local one-way functions. In *Proceedings of the 44th ACM Symposium on Theory of Computing*, pages 805–816, 2012. 1.2.2

[App13] Benny Applebaum. Cryptographic hardness of random local functions–survey. In *10th Theory of Cryptography Conference*, 2013. 1.2.2

[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

[AS04] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. John Wiley & Sons, 2nd edition, 2004. 3.3.5

[Ats15] Albert Atserias. A Note on Semi-Algebraic Proofs and Gaussian Elimination over Prime Fields. *CoRR*, abs/1502.03974, 2015. 5.3.2

[Aus08] Per Austrin. *Conditional Inapproximability and Limited Independence*. PhD thesis, KTH - Royal Institute of Technology, 2008. 3.7.1, 3.7.1, 3.7.1

[BB02] Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, (003), 2002. 1.1, 3.1.5

[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. 1.5

[BBHJ13] Adrian Balint, Anton Belov, Marijn J.H. Heule, and Matti Järvisalo. Generating the Uniform Random Benchmarks for SAT Competition 2013. In *Proceedings of SAT Competition 2013: Solver and Benchmark Descriptions*, pages 97–98, 2013. 1.1

[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 Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 97–106, 2015. 1.4, 4.3.1, 4.3.2, 4.3.2

[BCMV12] 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

[BDHJ14] Anton Belov, Daniel Diepold, Marijn J.H. Heule, and Matti Järvisalo. Generating the Uniform Random Benchmarks. In *Proceedings of SAT Competition 2014: Solver and Benchmark Descriptions*, page 80, 2014. 1.1

[BG01] Maria Luisa Bonet and Nicola Galesi. Optimality of size-width tradeoffs for resolution. *Comput. Complexity*, 10(4):261–276, 2001. 5.3.2

[BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8:269–289, 2012. 1.4, 4.3.1, 4.3.1

[BJK05] Andrei Bulatov, Peter Jeavons, and Andrei Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM J. Comput.*, 34(3):720–742, 2005. 1.1

[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. 1.2.1

[BKPS98] Paul Beame, Richard Karp, Toniann Pitassi, and Michael Saks. On the complexity of unsatisfiability proofs for random k-CNF formulas. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 561–571, 1998. 1.4, 5.3.2

[BKPS02] Paul Beame, Richard Karp, Toniann Pitassi, and Michael Saks. The efficiency of resolution and Davis-Putnam procedures. *SIAM J. Comput.*, 31(4):1048–1075, 2002. 5.3.2, 5.3.13, 1

[BKS13] Boaz Barak, Guy Kindler, and David Steurer. On the Optimality of Semidefinite Relaxations for Average-Case and Generalized Constraint Satisfaction. In *Proceedings of the 4th Innovations in Theoretical Computer Science conference*, 2013. 1.2.1, 1, 3.1.1

[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. 1.2.3, 3, 1.4, 1.5, 1.5, 3.1, 3.2.1, 3.4.1, 3.5.4, 5.2.2, 5.2.5, 5.2.6

[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, 1.4

[BP96] Paul Beame and Toniann Pitassi. Simplified and improved resolution lower bounds. In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science*, pages 274–282, 1996. 1.4

[BPR97] Maria Bonet, Toniann Pitassi, and Ran Raz. Lower bounds for cutting planes proofs with small coefficients. *J. Symbolic Logic*, 62(3):708–728, 1997. 5.3.2

[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. 1.2.1

[BS] Boaz Barak and David Steurer. Proofs, beliefs, and algorithms through the lens of sum-of-squares. http://sumofsquares.org/public/index.html. 2.2

[BS01] Eli Ben-Sasson. *Expansion in Proof Complexity*. PhD thesis, Hebrew University, 2001. 1.4, 5.3.2, 5.3.2

[BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *Proceedings of International Congress of Mathematicians*, 2014. 1.5

[BSB02] Eli Ben-Sasson and Yonatan Bilu. A gap in average proof complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 9(3), 2002. 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

[BSW99] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. In *Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999)*, pages 517–526 (electronic). ACM, New York, 1999. 1.4, 1, 5.3.2, 5.3.2, 5.3.2

[CD09] Nadia Creignou and Hervé Daudé. The SAT-UNSAT transition for random constraint satisfaction problems. *Discrete Math.*, 309(8):2085–2099, 2009. 1.1

[CEI96] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 174–183, 1996. 1.4

[Cha13] Siu On Chan. Approximation resistance from pairwise independent subgroups. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 447–456, 2013. 2.1

[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. 1.1

[CM01] Mary Cryan and Peter Bro Miltersen. On pseudorandom generators in NC^0. In Jiri Sgall, Ales Pultr, and Petr Kolman, editors, *Mathematical Foundations of Computer Science*, volume 2136 of *Lecture Notes in Computer Science*, pages 272–284. Springer Berlin Heidelberg, 2001. 1.2.2

[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

[COCF10] Amin Coja-Oghlan, Colin Cooper, and Alan Frieze. An efficient sparse regularity

concept. *SIAM J. Discrete Math.*, 23(4):2000–2034, 2009/10. 3.1, 3.1

[COCF09] Amin Coja-Oghlan, Colin Cooper, and Alan Frieze. An efficient sparse regularity concept. In *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 207–216. SIAM, Philadelphia, PA, 2009. 1.4

[COGL04] Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong Refutation Heuristics for Random k-SAT. In Klaus Jansen, Sanjeev Khanna, José D.P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 3122 of *Lecture Notes in Computer Science*, pages 310–321. Springer Berlin Heidelberg, 2004. 1.4, 3.5.4

[COGL07] Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong refutation heuristics for random k-SAT. *Combin. Probab. Comput.*, 16(1):5–28, 2007. 3.1, 3.1.3, 3.1.3, 3.6, 3.6.1, 3.6.2, 3.6.3, 3.6

[CR79] Stephen A. Cook and Robert A. Reckhow. The relative efficiency of propositional proof systems. *J. Symbolic Logic*, 44(1):36–50, 1979. 1.4

[CS88] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *J. Assoc. Comput. Mach.*, 35(4):759–768, 1988. 1.4, 1

[CW04] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: extending Grothendieck's inequality. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60, 2004. 3.5.1, 3.5.1, 5.2.3

[Dan15] Amit Daniely. Complexity Theoretic Limitations on Learning Halfspaces. *CoRR*, abs/1505.05800, 2015. 1.2.3, 1, 2, 3.1.1, 3.1.1

[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. 1.2.1

[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. 1.2.3, 3.1.1

[DLSS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 441–448. ACM, 2014. 1.1, 1.2.3, 1, 2, 3.1.1, 3.1.1, 3.1.1, 3.1.1, 3.3, 3.3.1, 3.3.1, 3.3.2, 3.3.5, 3.8

[DS14] Amit Daniely and Shai Shalev-Shwartz. Complexity theoretic limitations on learning DNF's. Technical Report 1404.3378, arXiv, 2014. 1.2.3, 1, 2, 3.1.1, 3.1.1

[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, 1.2.1, 1.2.1, 1, 3

[Fei07] Uriel Feige. Refuting smoothed 3CNF formulas. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 407–417, 2007. 1.4, 5.2.1, 5.2.1

[Fei15] Uriel Feige. Personal Communication, 2015. 5.2.2

[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, 3.1

[FHL] Yuval Filmus, Pavel Hrubeš, and Massimo Lauria. Semantic versus syntactic cutting planes. 5.3.2, 5.3.2

[FK81] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981. 3.1

[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, 1.4, 5.3.1, 5.3.1, 5.3.1, 5.3.1

[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

[Fri99] Ehud Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain. 1.1

[Fu96] Xudong Fu. *On the complexity of proof systems*. PhD thesis, University of Toronto, 1996. 1.4

[FW15] Uriel Feige and David Witmer. Nondeterministic refutation of any CSP beyond spectral methods. Unpublished manuscript, 2015. 5.3.3

[Gab16] Oliver Gableske. dimetheus. In *Proceedings of SAT Competition 2016: Solver and Benchmark Descriptions*, pages 37–38, 2016. 1.1

[Gal77a] Zvi Galil. On resolution with clauses of bounded size. *SIAM J. Comput.*, 6(3):444–459, 1977. 1

[Gal77b] Zvi Galil. On the complexity of regular resolution and the Davis-Putnam procedure. *Theoret. Comput. Sci.*, 4(1):23–46, 1977. 1

[GHP02] Dima Grigoriev, Edward A. Hirsch, and Dmitrii V. Pasechnik. Complexity of semi-algebraic proofs. In *Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science*, pages 419–430, 2002. 5.3.2

[GJ02] Andreas Goerdt and Tomasz Jurdziński. Some results on random unsatisfiable k-Sat instances and approximation algorithms applied to random structures. In *Mathematical Foundations of Computer Science 2002*, volume 2420 of *Lecture Notes in Comput. Sci.*, pages 280–291. Springer, Berlin, 2002. 3.3.12

[GJ03] Andreas Goerdt and Tomasz Jurdziński. Some results on random unsatisfiable k-Sat instances and approximation algorithms applied to random structures. *Combin. Probab. Comput.*, 12(3):245–267, 2003. Combinatorics, probability and computing (Oberwolfach, 2001). 3.1.5, 3.3.12

[GK01] Andreas Goerdt and Michael Krivelevich. Efficient recognition of random unsatisfiable k-SAT instances by spectral methods. In *STACS 2001 (Dresden)*, volume 2010 of

Lecture Notes in Comput. Sci., pages 294–304. Springer, Berlin, 2001. 1.4

[GL03] Andreas Goerdt and André Lanka. Recognizing more random unsatisfiable 3-SAT instances efficiently. In *Typical case complexity and phase transitions*, volume 16 of *Electron. Notes Discrete Math.* Elsevier, 2003. 2.1

[GL04] Andreas Goerdt and André Lanka. An approximation hardness result for bipartite Clique. *Electronic Colloquium on Computational Complexity (ECCC)*, (048), 2004. 1.2.1, 1

[Gol00] Oded Goldreich. Candidate One-Way Functions Based on Expander Graphs. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 7, 2000. 1.2.2, 5.1

[Gol11] Oded Goldreich. Three XOR-lemmas—an exposition. In *Studies in complexity and cryptography*, volume 6650 of *Lecture Notes in Comput. Sci.*, pages 248–272. Springer, Heidelberg, 2011. 2.1.13, 3.7.3

[Gri01] Dima Grigoriev. Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1-2):613 – 622, 2001. 1.4, 1

[GV01] Dima Grigoriev and Nicolai Vorobjov. Complexity of Null- and Positivstellensatz proofs. *Annals of Pure and Applied Logic*, 113(1–3):153 – 160, 2001. First St. Petersburg Conference on Days of Logic and Computability. (document), 2.2.1

[Hak85] Armin Haken. The intractability of resolution. *Theoret. Comput. Sci.*, 39(2-3):297–308, 1985. 5.3.2

[Hås01] Johan Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. 1.1

[Hua13] Sangxia Huang. Approximation resistance on satisfiable instances for predicates with few accepting inputs. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 457–466, 2013. 1, 3.1.1, 3.3.2, 3.3.2

[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. 1.2.2, 2

[IPS99] R. Impagliazzo, P. Pudlák, and J. Sgall. Lower bounds for the polynomial calculus and the Gröbner basis algorithm. *computational complexity*, 8(2):127–144, 1999. 1.4, 5.3.2, 5.3.2

[Kho06] Subhash Khot. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006. 2.1

[KI06] Arist Kojevnikov and Dmitry Itsykson. Lower Bounds of Static Lovász-Schrijver Calculus Proofs for Tseitin Tautologies. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, 2006. 1.4, 5.3.2

[KL96] Ilia Krasikov and Simon Litsyn. On integral zeros of Krawtchouk polynomials. *Journal of Combinatorial Theory, Series A*, 74(1):71–99, 1996. 3.3.3

[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

[KMOW17] Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of

squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, 2017. (document), 1.5.4, 1.5.5, 1.5.6, 1.5.7, 1.5.4, 1.5.5

[KMRT+07] Florent Krzakała, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007. 1.1

[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. 1.5

[Kra29] Mikhail Krawtchouk. Sur une généralisation des polynomes d'Hermite. *Comptes Rendus*, 189:620–622, 1929. 3.3.3

[Las00] Jean B. Lasserre. Optimisation globale et théorie des moments. *C. R. Acad. Sci. Paris Sér. I Math.*, 331(11):929–934, 2000. (document), 2.2.1

[Las01] Jean B. Lasserre. Global Optimization with Polynomials and the Problem of Moments. *SIAM J. Optim.*, 11(3):796–817, 2001. (document), 2.2.1

[Lau09] Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 157–270. Springer, New York, 2009. 1.5, 2.2

[LN15] Massimo Lauria and Jakob Nordström. Tight Size-Degree Bounds for Sum-of-Squares Proofs. In *Proceedings of the 30th Conference on Computational Complexity*, pages 448–466, 2015. 5.3.2

[LRS15] James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 567–576, 2015. 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

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I.* North-Holland Publishing Co., Amsterdam, 1977. North-Holland Mathematical Library, Vol. 16. 3.3.5

[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. 1.2.2

[MW16] Ryuhei Mori and David Witmer. Lower bounds for CSP refutation by SDP hierarchies. In *RANDOM '16*, 2016. 1.4

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 2.1.2, 3.7.1

[O'D17] Ryan O'Donnell. SOS is not obviously automatizable, even approximately. In *Proceedings of the 8th Innovations in Theoretical Computer Science conference*, 2017. 2.2.1

[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. 1.2.2, 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. 1.2.1

[OZ13] Ryan O'Donnell and Yuan Zhou. Approximability and proof complexity. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1537–1556. SIAM, Philadelphia, PA, 2013. 1.5, 2.2, 3.5.3, 3.5.3

[Par00] Pablo Parillo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000. (document), 2.2.1

[Pud97] Pavel Pudlák. Lower bounds for resolution and cutting plane proofs and monotone computations. *J. Symbolic Logic*, 62(3):981–998, 1997. 5.3.2

[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. 1.1

[Rao07] Anup Rao. An Exposition of Bourgain's 2-Source Extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, 2007. 3.7.3

[Rob65] J. A. Robinson. A machine-oriented logic based on the resolution principle. *J. Assoc. Comput. Mach.*, 12:23–41, 1965. 5.3.2

[RRS17] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, 2017. (document), 1.3, 1.5, 1.5.3

[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

[RW17] Prasad Raghavendra and Benjamin Weitz. On the Bit Complexity of Sum-of-Squares Proofs. Technical Report 1702.05139, arXiv, 2017. 1.5, 2.2.1

[SAT] http://satcompetition.org/2014/certunsat.shtml. 1.1

[Sch78] Thomas J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the 10th Annual ACM Symposium on Theory of Computing*, pages 216–226, 1978. 1.1

[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, 5.3.1, 5.3.2, 5.3.15

[Sho87] N.Z. Shor. Class of global minimum bounds of polynomial functions. *Cybernetics*, 23(6):731–734, 1987. (document), 2.2.1

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. 3.5.1, 3.5.4

[Tse66] Grigori S Tseitin. On the complexity of derivation in propositional calculus. In *Leningrad Seminar on Mathematical Logic*, 1966. 1

[Tul09] Madhur Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, 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

[Val84] Leslie Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, November 1984. 3.3

[Vaz86] Umesh Vazirani. *Randomness, Adversaries and Computation*. PhD thesis, University of California, Berkeley, 1986. 3.7.3

[Wig55] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955. 3.1