# Algorithmic Thresholds for Refuting Random Polynomial Systems

Jun-Ting Hsieh* Pravesh K. Kothari*

October 19, 2021

**Abstract**

Consider a system of $m$ polynomial equations $\{p_i(x) = b_i\}_{i\leqslant m}$ of degree $D \geqslant 2$ in $n$-dimensional variable $x \in \mathbb{R}^n$ such that each coefficient of every $p_i$ and $b_i$s are chosen at random and independently from some continuous distribution. We study the basic question of determining the smallest $m$ – the *algorithmic threshold* – for which efficient algorithms can find *refutations* (i.e. certificates of unsatisfiability) for such systems. This setting generalizes problems such as refuting random SAT instances, low-rank matrix sensing and certifying pseudorandomness of Goldreich's candidate generators and generalizations.

We show that for every $d \in \mathbb{N}$, the $(n+m)^{O(d)}$-time canonical sum-of-squares (SoS) relaxation refutes such a system with high probability whenever $m \geqslant O(n) \cdot (\frac{n}{d})^{D-1}$. We prove a lower bound in the restricted *low-degree polynomial* model of computation which suggests that this trade-off between SoS degree and the number of equations is nearly tight for all $d$. We also confirm the predictions of this lower bound in a limited setting by showing a lower bound on the canonical degree-4 sum-of-squares relaxation for refuting random quadratic polynomials. Together, our results provide evidence for an algorithmic threshold for the problem at $m \gtrsim \widetilde{O}(n) \cdot n^{(1-\delta)(D-1)}$ for $2^{n^\delta}$-time algorithms for all $\delta$.

Our upper-bound relies on establishing a sharp bound on the smallest integer $d$ such that degree $d-D$ polynomial combinations of the input $p_i$s generate all degree-$d$ polynomials in the ideal generated by the $p_i$s. Our lower bound actually holds for the easier problem of distinguishing random polynomial systems as above from a distribution on polynomial systems with a "planted" solution. Our choice of planted distribution is slightly (and necessarily) subtle: it turns out that the natural and well-studied planted distribution for quadratic systems (studied as the *matrix sensing* problem in machine learning) is easily distinguishable whenever $m \geqslant \widetilde{O}(n)$ – a factor $n$ smaller than the threshold in our upper bound above. Thus, our setting provides an example where refutation is harder than search in the natural planted model.

*Carnegie Mellon University, Supported by NSF CAREER Award #2047933.

# Contents

# 1 Introduction

Suppose you are given a system of polynomial equations $\{p_i(x) = b_i\}_{i \leqslant m}$ where each $p_i$ is a homogeneous polynomial of degree $D$ and each $b_i$ and each coefficient of $p_i$ are independent standard Gaussians. When $m \geqslant n+1$, an elementary argument[1] shows that the system has no real (or complex!) solution with probability 1. In this work, we study the problem of finding *refutations* – certificates of unsatisfiability – for such random polynomial systems with $m \geqslant n+1$.

When $D = 1$, the classical Gauss-Jordan elimination for solving linear systems efficiently produces a refutation whenever $m \geqslant n+1$. When $D \geqslant 2$, the problem is NP-hard in the worst-case (it encodes Max-Cut) and the setting above is a (and perhaps, "the") natural average-case formulation. As $m$ increases, finding refutations gets easier and indeed when $m \geqslant N_D = \sum_{i \leqslant D} \binom{n+i-1}{i} = \Omega(n^D)$ (we call this the *linearization threshold*), one can simply "linearize" the polynomials and apply Gauss-Jordan elimination to linear functions on $N_D$ variables to obtain a refutation algorithm.

Can efficient algorithms refute random polynomial systems below the linearization threshold? More generally, what's the *algorithmic threshold* – i.e. the smallest $m = m(n, D)$ – such that efficient algorithms can (with high probability) come up with efficiently verifiable certificates of unsatisfiability of random polynomial systems with $m$ equations?

Let's cut to the chase: in this paper, we design algorithms and prove lower bounds that suggest that polynomial time algorithms can *non-trivially but not appreciably* beat the "linearization" threshold above. Specifically, for any $d \in \mathbb{N}$, and degree $D \geqslant 2$ polynomials, we give an $n^{O(d)}$ time algorithm that succeeds in refuting random polynomial systems with $m \gtrsim O(n) \cdot \left(\frac{n}{d}\right)^{D-1}$ equations and our lower bounds (in restricted models of computation) suggest that this trade-off is nearly tight. This threshold is smaller (but only by a constant factor) than the linearization threshold for any $d \geqslant 2$. This may come as a surprise since for related problems such as *maximizing* low-degree polynomials, the case of degree $D = 2$ polynomials (in contrast to degree $D \geqslant 3$) is often "easy" and exhibits no information-computation gap. For $2^{n^\delta}$ time algorithms more generally, our results suggest that the algorithmic threshold beats the linearization threshold by a multiplicative factor of $\sim n^{\delta(D-1)}$. Taken together, our results suggest that the algorithmic threshold "smoothly" drops from $\sim n^d$ to the information-theoretic threshold of $\sim n$ as the running time budget for the refutation algorithm grows from $\mathrm{poly}(n)$ to $2^n$.

Before presenting our results, we discuss how the problem above is the refutation counterpart of natural algorithmic questions arising in diverse areas.

## 1.1 Random polynomial systems generalize well-studied problems

*In algebraic geometry,* the study of random polynomials and their zeros began with the 1932 paper of Bloch and Pólya [BP31] leading to the seminal work of Kac [Kac49] on average number of real zeros of random univariate polynomials. More recently, beginning with work of Shub and Smale as part of their "Bezout series" [SS93b, SS93a, SS93c, SS96, SS94, Shu09, BS09], an influential sequence of works has focused on estimating the distribution of the number of common zeros of $n$

[1] The classical Bezout's theorem says that the number of common complex zeros of $n$ "generic" polynomials – this condition holds with probability 1 whenever the coefficients are independently drawn from a continuous distribution – of degree $\leqslant d$ in $n$ variables is at most $d^n$. Apply this to the first $n$ equations and observe that the chance that the $n+1$-th polynomial has a zero at any of the $\leqslant d^n$ common zeros of the rest is 0. Via more sophisticated arguments (e.g. [KZ13]), such a result can be extended to random polynomial systems with coefficients chosen from discrete distributions.

Gaussian random $n$-variate polynomials of degree $D$. For example, it is known that the expected number of complex common zeros grows as $\sim D^{m/2}$ – a quadratic improvement over the the "worst-case" bound obtained via Bezout's theorem. Extending this to counting real common zeros requires constraining the combinatorial structure of the monomials with non-zero coefficients and more sophisticated ideas (see Kostlan's work [Kos02] for an overview). As his 17th problem for the 21st century, Smale [Wik] asked if there is a deterministic polynomial time algorithm for finding one such common zero. A sequence of breakthroughs due to Beltrán and Pardo [BP08], Burgisser and Cucker [BC11] and Lairez [Lai17] resolved Smale's question and found a deterministic polynomial time algorithm based on *numerical homotopy methods*.

The problem we study in this work is a natural extension: when the number of equations is $m \leqslant n$, the pertinent algorithmic problem is that of counting and finding common zeros. When $m > n$, the relevant question is of finding refutations (i.e. certificates of unsatisfiability).

*In combinatorial optimization,* refuting random polynomial systems generalizes foundational problems such as certifying bounds on combinatorial quantities like the clique number and chromatic number of random graphs. One well-studied special case is that of refuting random constraint satisfaction problems (CSPs), which is equivalent to refuting *sparse* polynomial equations (one for each "clause") with random coefficients over the hypercube. A long line of work [Fei02, COGL07, Fei07, AOW15, BM16, RRS17, BCK15, KMOW17, AGK21] have led to a complete understanding of algorithmic thresholds for refuting random constraint satisfaction problems in terms of a basic combinatorial property of the underlying predicate. The problem we study in this work is a natural *dense* (i.e. all monomials appear with non-zero coefficients) counterpart to the sparse random polynomial systems arising in the study of random CSPs.

*In statistical learning theory,* random polynomial systems arise naturally and in fact are closely related to the well-studied matrix and tensor sensing problems. For example, in the matrix sensing problem with "random Gaussian measurements", there is an unknown rank-$r$ matrix $X$ such that one is required to reconstruct $X$ from equations of the form $\langle G_i, X \rangle = b_i$ where $G_i$ are random matrices with Gaussian entries. When the rank $r = 1$ (and $X = xx^\top$ is symmetric) this corresponds to the problem of solving random quadratic equations (i.e. $D = 2$) where the right hand sides correspond to evaluation of the polynomials at some unknown vector $x$. The tensor sensing problem is a generalization where instead of matrices, $G_i$s are random Gaussian tensors of order $D$. A long line of work beginning with [Can10, Rec11, Gro11] has led to essentially optimal algorithms based on semidefinite programming for solving the matrix sensing (and variants such as matrix completion) problem and more recently, much progress [BM16, PS17, dKNS20] has been made even on the tensor variants. This work can be seen as studying the refutation variant of the matrix/tensor sensing problems for rank 1 matrices/tensors.

*In cryptography,* random polynomial systems over the reals arise naturally in a recent sequence of works that use conjectures about the *hardness* of solving random polynomial systems. The work of [Lin16] led to a program [Lin17, AS17, LT17] for building *indistinguishability obfuscation* (iO) based, among other components, on a certain variant of Goldreich's [Gol00] candidate pseudorandom generator. An offshoot of this program recently culminated [JLS21] in the discovery of the first construction of indistinguishability obfuscation based on standard assumptions.

At a high-level such works consider maps $f : \mathbb{Z}^n \to \mathbb{Z}^m$ where each of the $m$ outputs is computed as a low-degree $D$ polynomial $p_i$ of the $n$ inputs. The interest is in finding maps $f$ such that $D$ is small (say 2) but for $m \gg n$ (say, $m \sim n^{1.1}$ for concreteness), the $m$-dimensional output is com-

putationally indistinguishable from some distribution where each output is independent. Several candidate constructions of such pseudo-random generators were shown to be insecure by describing efficient algorithms (based on the sum-of-squares hierarchy of semidefinite programs studied in this work) that *invert* the map $f$ – i.e. compute a solution to the system of polynomial equations defined by the map [LV17, BBKK18, BHJ+19]. One candidate construction (see [BHJ+19]) was in fact based on choosing each of the $m$ polynomials to be random quadratic polynomials as in the model studied in this paper. This work provides strong evidence for the algorithmic threshold for the refutation version of this problem.

### 1.2 Our results

**Algorithms.** Our main algorithmic result uses the sum-of-squares hierarchy to non-trivially improve on the linearization trick for refuting random polynomial systems. We note that for the various special cases (such as random constraint satisfaction problems, the matrix/tensor sensing problems and generalizations of Goldreich's pseudo-random generator), semidefinite programs from the sum-of-squares hierarchy provide the state-of-the-art algorithms for solving/refuting polynomial systems.

**Theorem 1.1** (Refutation Algorithm, Informal, See Theorem 3.2 for a formal version)**.** *For every $D \in \mathbb{N}$ and every $d \geqslant D$, there is a $n^{O(d)}$ time algorithm – namely the canonical degree-$d$ sum-of-squares relaxation – that takes input a system of $m$ polynomial equations and either correctly outputs "infeasible" or returns "don't know". When each coefficient of each input polynomial is drawn from an independent nice distribution and $m \geqslant O(n^D/d^{D-1})$, the algorithm outputs "infeasible" with probability $\geqslant 1 - 1/n$.*

We note a few important comments about some implications to settings studied in average-case complexity, cryptography and proof complexity.
*Computational Model:* The algorithm works in the standard word RAM model of computation. We assume that the coefficients of all our polynomials are rational numbers. The bit-complexity of our algorithm (see Theorem 3.2 for details) is polynomial in the input size (including the size of the coefficients of the input polynomials).

*Nice Distributions:* Our algorithm works for any system of random polynomial equations as long as the coefficients are chosen from independent (possibly different for each coefficient) distributions as long as they satisfy two niceness conditions (see Definition 3.1). The first asks that the distributions be supported on rational numbers with some upper bound $B$ on the bit complexity. The running time of our algorithm grows polynomially in $B$. Such a condition is essentially[2] necessary for any algorithmic result. The second condition forces a certain weak *anti-concentration* property and posits that no rational number should have a probability larger than $n^{-O(d)}$. We note that $n^{O(d)}$-bit rational truncations of any natural continuous distribution such as uniform distribution on $[-1, 1]$ or the standard Gaussian distribution $\mathcal{N}(0, 1)$ satisfy such properties.

*Time vs Signal Strength Trade-off:* The algorithm provides a certificate of unsatisfiability of the input polynomial system since whenever the algorithm outputs "infeasible" it is correct. The guarantees of the algorithm provide a trade-off between running time budget (parameterized by $d$) and the

[2]In principle, there could be specialized algorithms that work with non-standard representations of real numbers. We do not know of any such algorithm.

number of equations (a measure of "signal strength" in this setting) required for the algorithm to succeed in refuting with high probability. For any $d$, the smallest $m$ for which the algorithm succeeds improves on the requirement of the linearization trick by a factor of roughly $d^{\Omega(D)}$. On the other hand, to refute at the information-theoretically minimum required $m$, the algorithm runs in time exponential in $n$. In general, by setting $d = n^{\delta}$, we obtain a $2^{O(n^{\delta})}$ time algorithm that succeeds in refuting random degree-$D$ polynomial systems with $O(n) \cdot n^{(1-\delta)(D-1)}$ equations.

One can view this result as a generalization of the work of Bhattiprolu, Guruswami and Lee [BGL17], and Raghavendra, Rao and Schramm [RRS17] who proved a sum-of-squares degree vs signal-strength trade-off for certifying maxima of random tensors and refuting random CSPs, respectively. In particular, the result in [RRS17] can be seen as a degree vs number of equations trade-off that is qualitatively similar to ours above for random 1-*sparse* polynomials (i.e. monomials) over the hypercube.

*The Importance of Solutions with Typical vs Atypical Norm:* There appears to be a key and perhaps surprising difference in the setting of random polynomial system refutation when compared to random CSP refutation (and more generally, related problems such as certifying maxima of random low-degree polynomials) that we wish to highlight. For random polynomial systems arising in the context of refuting CSPs (as in [AOW15, RRS17]), the case of degree $D = 2$ polynomial systems is "easy" and appears to exhibit essentially no information-computation gap. In contrast, in our setting, our upper bound above requires $m = \Omega(n^2)$ for polynomial time algorithms to succeed in refutation. Further, our lower bounds suggest that our algorithm above is in fact essentially optimal in the running time vs number of equations trade-off.

This apparently paradoxical difference is related to the issue of having a good upper bound on the $\ell_2$ norm in the solution space. In the context of CSP refutation, the goal is to find certificates of unsatisfiability over the $n$-dimensional hypercube – in particular, the solution vectors have a fixed $\ell_2$ norm of $\sqrt{n}$. Indeed, the spectral (and thus, SoS-based) refutation algorithms developed in that context continue to work even for refuting random (sparse) polynomial systems over the space of all vectors with "typical" (with respect to the scale of the coefficients of the input polynomials $p_i$s and the right-hand sides $b_i$s) $\ell_2$ norm.

On the other hand, in our setting, the goal is to refute the given random polynomial system over a solution space of vectors of arbitrarily large norms. This crucial difference appears to make our setting harder even for the usually tame case of quadratic polynomials. Indeed, this becomes even more apparent when we construct our low-degree polynomial hardness described below where it's crucial to make a subtle choice of planted distribution where the solution vector needs to be of $\ell_2$-norm about $n^{1/2}$-factor larger than "typical".

Further, this uncertainty in $\ell_2$-norm of the solution space in fact occurs in random polynomial systems that arise in applications. For e.g., in the cryptographic applications discussed above, the "planted" solution vectors have integer coordinates with variance $\mathrm{poly}(n)$ and thus, the $\ell_2$-norm is known only up to some (large) $\mathrm{poly}(n)$-factor. That is precisely the setting where our results apply and appear to suggest a difference from the refutation settings studied in prior works. In particular, it suggests that speculating the hardness of solving/refuting random polynomial systems based on CSP hardness results may lead to incorrect conclusions. We believe that it's an interesting goal to chalk out a full trade-off between sparsity, length of the solution vector and algorithmic thresholds. Such an endeavor is likely to yield interesting insights into the phase tran-

sitions between the qualitatively different behaviors exhibited by random polynomial systems.

*Nullstellensatz vs Sum-of-Squares Refutation:* Our proof of the theorem above works by showing that there is a degree-$d$ sum-of-squares "refutation" (i.e. a proof of unsatisfiability of the polynomial system that can be written in the restricted degree-$d$ *sum-of-squares proof system*) of the input polynomial equations (see Section 2.3). Thus, from a proof complexity perspective, our result shows that there are degree-$d$ sum-of-squares refutations for systems of random polynomials over the reals whenever $m \geqslant O(n^D/d^{D-1})$. Our proof in fact establishes a stronger result: we show that our certificate of unsatisfiability can be written in the (formally) weaker Nullstellensatz proof system. As we discuss next, this shows that for the problem of refuting random polynomial equations, the degree required for Nullstellensatz and sum-of-squares proof systems can only be different by some fixed constant factor.

**Sharp thresholds at degree 2.** For the special case of degree $d = 2$ and $D = 2$ (i.e. quadratic polynomials and degree-2 sum-of-squares algorithm) and standard Gaussian coefficients, we can obtain sharp constants in the threshold $m$.

**Theorem 1.2** (Sharp Thresholds for Degree-2 SoS)**.** *Let $\mathcal{G} = \{g_i(x) = b_i\}_{i\in[m]}$ be a system of $m$ polynomial equations where each coefficient of each $g_i$ is chosen from the standard Gaussian distribution $\mathcal{N}(0,1)$. Then,*

- *if $m \geqslant \frac{n^2}{4} + \widetilde{O}(n)$, there is a degree-2 sum-of-squares refutation of the system $\mathcal{G}$ with probability 0.49.*
- *if $m \leqslant \frac{n^2}{4} - \widetilde{O}(n)$, there is no degree-2 sum-of-squares refutation of the system $\mathcal{G}$ with probability $1 - 1/n$.*

Our proof of Theorem 1.2 is short and is based on direct application of results from [ALMT14] that build methods based on conic integral geometry to analyze the feasibility of convex programs with random inputs. Our present analysis does not give a bit-complexity bound on the degree-2 sum-of-squares proofs obtained via this technology. As a result, they do not immediately imply algorithmic results. However, they do strongly suggest that the threshold value of $m$ at $d = 2$ should be $\sim n^2/4$.

It is not hard to prove that the threshold $m$ for $d = 2$-degree Nullstellensatz refutation is $\sim n^2/2$. Thus, this result implies a factor 2 multiplicative gap between the thresholds for Nullstellensatz and SoS refutations to succeed at degree $d = 2$.

**Lower bounds in the low-degree polynomial model.** Our algorithms beat the linearization trick non-trivially at all degrees $d$. However, polynomial-time methods from our schema still require $\Omega(n^D)$ equations for the refutation to succeed. This is a factor $n^{D-1}$ larger than the information-theoretic threshold of $n + 1$. Thus, it is natural to ask if this *information vs computation* gap is "real" and in particular, if our algorithmic results are suboptimal. We provide lower bounds that suggest that our algorithmic results are tight up to absolute constant factors.

Our lower bounds actually hold for the formally *easier*[3] algorithmic task of distinguishing random systems of polynomial equations from an appropriately designed *planted* distribution on polynomial systems that always admit a solution.

In this work, we prove the following lower bound for the distinguishing variant above in the *low-degree polynomial* model of computation.

**Theorem 1.3** (Low-Degree Hardness, Informal, See Theorem 4.1 for a formal version)**.** *Fix $D \in \mathbb{N}$. For every $d \leqslant \frac{2n}{D}$, whenever $m \leqslant O\left(\frac{n^D}{d^{D-1}}\right)$, there exists a probability distribution $\nu_P$ on systems of $m$ polynomial equations that admit a solution with probability 1 such that degree-$d$ polynomials fail to distinguish between $\nu_P$ and the distribution of random polynomial systems with $m$ equations.*

The trade-off between $d$ and $m$ achieved by Theorem 1.3 matches that of our algorithm in Theorem 1.1 up to absolute constant factors. This suggests, in particular, that the algorithmic threshold of polynomial-time algorithms might be $\Omega(n^D)$.

The *low-degree polynomial method* (see [KWB19] for a great exposition) allows distinguishers that compute thresholds of bounded-degree polynomials of input data. While low-degree polynomials might appear restricted, they capture several algorithms including power iteration, approximate message passing, and local algorithms on graphs (cf. [DMM09, GJW20]). Moreover, it turns out that they are enough to capture the best known spectral algorithms for several canonical problems such as planted clique, community detection, and sparse/tensor principal component analysis [BHK+19, HS17, DKWB19, HKP+17]. This model arose naturally from work on constructing sum-of-squares lower bound for the planted clique problem [BHK+19]. It was formalized in [HKP+17] with a concrete quantitative conjecture (called the *pseudo-calibration conjecture*) which informally says that for average-case refutation problems satisfying some mild niceness conditions, degree-$(d \log n)$ lower bounds for the low-degree polynomial method for distinguishing a random draw from a random draw of some planted distribution imply lower bounds on the canonical sum-of-squares relaxation of degree-$d$ for the refutation problem. Subsequently, starting with [HS17, Hop18], researchers have used the low-degree polynomial method as a technique to demarcate average-case algorithmic thresholds for a number of average-case algorithmic problems including densest $k$-subgraph, sparse/tensor principal component analysis, finding independent sets in random graphs among others [HKP+17, GJW20, SW20, Wei20].

**Sum-of-Squares lower bound at degree 4.** We provide further evidence in favor of the thresholds suggested by both our algorithms and hardness results by proving a lower bound on the degree-4 sum-of-squares relaxation for refuting random *quadratic* polynomial systems. Our proof is based on constructing a dual witness via pseudo-calibration – this has become a standard technique for constructing dual witnesses for sum-of-squares lower bounds [BHK+19, HKP+17, MRX20, GJJ+20]. We believe that it is possible to extend our lower bounds (with the same construction of the dual witness) to both higher-degree random polynomials and higher-degree SoS relaxations. But this will likely require challenging technical work in the analysis.

**Theorem 1.4** (Sum-of-Squares Lower Bound at Degree 4, see Theorem 6.1 for a formal version)**.** *Let $g_1, g_2, \ldots g_m$ be homogeneous degree-2 polynomials in $x_1, x_2, \ldots, x_n$ with independent Gaussian coef-*

[3] Any refutation algorithm provides a distinguishing algorithm that succeeds with high probability in distinguishing between an instance of a random polynomial system from an instance chosen at random from any planted distribution. This algorithm runs the refutation algorithm and simply returns "not planted" if the algorithm outputs "infeasible".

*ficients. Then, whenever* $m \leqslant n^2/\operatorname{poly}(\log n)$*, the canonical degree-*4 *sum-of-squares relaxation fails to refute* $\{g_i(x) = 0\}_{i\leqslant m}$ *with probability at least* $1 - o(1)$ *over the draw of* $g_i$*s.*

**Remark 1.5** (Hardness of Refutation vs Hardness of Natural Planted Variants)**.** Random polynomial systems arising in applications are often studied in two closely related variants: the refutation version for random polynomial system (the *null* model) studied in this work and a related *planted* variant. In the planted setting, the resulting polynomial system has a solution with probability 1. There are three natural problems that are studied in this context: 1) efficiently distinguish between a polynomial system chosen from either random or the planted distribution, 2) efficiently find the planted solution, and 3) efficiently refute the existence of a solution.

For average-case variants of several well-studied problems, the complexities of the three problems for natural planted and null distributions are often conjectured to be related. Indeed, researchers often prove lower bounds for the refutation problem (this turns out to be especially natural in the context of hardness for convex programs) and interpret it as a lower bound for the associated planted variant [4].

The natural, well-studied planted variant in the context of random polynomial systems happens to be the following model: a) choose polynomials $p_1, p_2, \ldots, p_m$ randomly, say, with independent Gaussian coefficients, b) choose a $z \sim \mathcal{N}(0,1)^n$, and c) output $\{p_i(x) = p_i(z)\}_{i\leqslant m}$ with the planted solution $z$. This planted variant captures both the rank-1 case of the matrix/tensor sensing problems in machine learning and the low-degree pseudo-random generators arising in recent works [LV17, BBKK18, BHJ+19] on constructing indistinguishability obfuscation in cryptography.

The planted distribution that we use in proving both low-degree hardness and our sum-of-squares lower bound is actually different from this natural variant and it turns out that this is necessary! Indeed, for $D = 2$, for e.g, the natural planted distribution on quadratic systems above turns out to be solvable at nearly the information-theoretic threshold of $m = \widetilde{O}(n)$ via the nuclear norm minimization semidefinite program ($\sim$ degree-2 SoS). In contrast, our results suggest that refuting random degree-2 polynomial systems in polynomial time likely requires $\Omega(n^2)$ equations. On the other hand, our slightly subtle variant (see Definition 4.3) of the planted distribution that appears hard even with $\Omega(n^2)$ equations for all the three problems.

Beyond the application to polynomial systems, this suggests that care must be taken in speculating hardness of natural planted variants of average-case problems based on the hardness of the refutation variants of the problem.

### 1.3 Overview of our techniques

In this section, we give a brief overview of our techniques.

**Algorithm via completeness of generated ideals.** Let $\{p_i(x) - b_i = 0\}_{i\leqslant m}$ be the input polynomial equations of degree $D$ given to the algorithm. If $x$ satisfies this system, observe that it must hold that $p_i(x)x^\alpha - b_i x^\alpha = 0$ for any monomial $x^\alpha$. Further, if $|\alpha| = d - D$ is the degree of the monomial $x^\alpha$, then this reasoning is "captured" by the degree-$d$ sum-of-squares proof system (in

[4] For e.g., the works [DM15, MPW15, HKP+16, BHK+19] proving sum-of-squares lower bounds for refuting clique number of random graphs has *planted or hidden clique* appearing in the title.

fact, by simply the degree-$d$ Nullstellensatz proof system). Thus, starting from the original polynomial system, we can "derive" a collection of degree-$d$ polynomial equations that must all be true if the original system is: $\{p_i(x)x^\alpha - b_i x^\alpha = 0\}_{i \leqslant m}$ – we call this the *generated ideal at degree d*.

Here's the main idea in our algorithm: suppose that the generated ideal at degree $d$ happens to be *complete* – that is, for every *homogeneous* polynomial $f$ of degree $d$, there are polynomials $a_1, a_2, \ldots, a_m$ of degree $d - D$ such that $\sum_{i \leqslant m} a_i(x)(p_i(x) - b_i) = f(x)$. We claim that it is easy to find a refutation in this case. To see why, suppose that for some $i \leqslant m$, $b_i \neq 0$ (such an $i$ exists whp). Then, note that we can derive $p_i(x)^{d/D} = b_i^{d/D}$ from the input equations in degree $d$ (assuming $d$ is a multiple of $D$). On the other hand, since $p_i(x)^{d/D}$ is a homogeneous polynomial of degree $d$ and the generated ideal at degree $d$ is complete, we must also have that $p_i(x)^{d/D} = \sum_i a_i(x)(p_i(x) - b_i)$ for some polynomials $a_i$ of degree $d - D$. Thus, together, we can infer that $b_i^{d/D} = p_i(x)^{d/D} = \sum_j a_j(x)(p_j(x) - b_j)$ or:

$$1 = b_i^{-d/D} \sum_j a_j(x)(p_j(x) - b_j)\,.$$

This is a (degree-$d$ Nullstellensatz and thus, sum-of-squares) refutation since the LHS is the constant 1 while at any $x$ that satisfies the input system, the RHS must be 0. Finally, we can argue (see Lemma 3.7) that whenever such a polynomial identity as above exists, the $a_i$s can be guaranteed to exist with coefficients of bit-complexity polynomial in $n^d$ and the bit-complexity of the coefficients of the inputs $p_i$s. This immediately implies (via Fact 2.12) that the $n^{O(d)}$ time algorithm (see Algorithm 3.3) for approximately solving the canonical degree-$d$ SoS relaxation of the polynomial system above succeeds in refuting the input polynomial system.

Thus, our task reduces to establishing that when $m \gg n^D/d^{D-1}$, the generated ideals at degree $d$ of random $(p_i - b_i)$s are complete. Such a condition naturally yields a system of linear equations so our task reduces to proving that this system admits a solution – that happens if and only if the coefficient matrix of the equations has full row rank. One might be tempted to prove such a claim by showing that when the $p_i$s are random, then for each $i$ and each $\alpha$, $p_i(x)x^\alpha - b_i x^\alpha$ are linearly independent when viewed as their coefficient vectors. This is false – there are several linear dependencies between such vectors.

Indeed, in general, such an argument requires some care as the entries of the matrix defining the linear equations are heavily correlated – this is not surprising since the there are roughly $mn^D \leqslant n^{2D}$ independent random variables in the input while the matrix is of dimension roughly $n^d$ (and $d \gg D$). We analyze this matrix by a careful decomposition (see Definition 3.12) that exploits the structure of the matrix to argue that whenever $m \gg n^D/d^{D-1}$, the resulting matrix is indeed full row rank with probability $1 - n^{-O(d)}$ over the draw of the coefficients of $p_i$s.

**Low-degree hardness and the hard-to-distinguish planted distribution.** In order to construct our lower bound, we need to come up with a planted distribution on polynomial systems with $m \geqslant n + 1$ equations such that 1) every system in the support always admits a solution but at the same time, 2) a draw from the planted distribution is indistinguishable from random polynomial systems that do not have a solution with probability 1 by any low-degree polynomial in the coefficients of the input polynomials. Notice that for e.g., this must mean that low-degree polynomials in the coefficients of the input polynomials cannot approximate the fraction of polynomial constraints satisfied by any $x$. Operationally, this means that we must pick a distribution on poly-

nomial equations that is "as close as possible" to random polynomial systems (the null model) while being satisfiable.

As we discussed in Remark 1.5, the design of the planted distribution is slightly subtle. One might be tempted to use the natural (and well-studied) variant where we pick each $p_i$ randomly just as in the null model and then choose $b_i$s to equal $p_i(x^*)$ for each $i$ for some random $x^*$. Notice that this must introduce correlations in $b_i$s and it in fact turns out that these correlations are strong enough that the resulting planted model can be easily distinguished from the null model by just a degree-4 polynomial[5] in the coefficients of the input polynomials whenever $m \gg n$ – the information theoretic threshold. This is not surprising as there is in fact an algorithm (the so called *nuclear norm minimization* semidefinite program) that recovers the planted $x^*$ when given input a random polynomial system chosen from the planted model above.

Instead, our construction of the planted distribution encodes subtle correlations in the polynomials $p_i$s themselves. Specifically, our planted distribution first picks $b_i$s to be independent draws from the standard Gaussian distribution, chooses a random $x^*$ *of sufficiently large length that* $\to \infty$ *as* $m \to \infty$, and then chooses $p_i$s to be polynomials with standard Gaussian coefficients conditioned on $p_i(x^*) = b_i$. In this version, notice that $b_i$s are clearly independent but unlike the natural planted variant, the coefficients of $p_i$s are mildly correlated. We show that such correlations are subtle enough that no low-degree polynomial can "notice" them. The argument crucially requires that the planted solution $x^*$ have sufficiently large norm – in Remark 4.5 we show that there's a simple distinguisher if the planted solution has bounded or slowly growing norm. It turns out that when the planted solution $x^*$ has sufficiently fast-growing norm, there's a *sharp phase transition* for distinguishability by degree-$d$ polynomials at a threshold $m = O_D(n^D/d^{D-1})$ from the planted model – a threshold that precisely matches the bound at which our algorithm works! This gives a pleasingly tight algorithmic threshold of $\Theta_D(n^D/d^{D-1})$ for distinguishing random polynomial systems from the above planted ones and thus also for refuting them.

Our analysis of the performance of low-degree polynomial distinguishers for the above pair relies on expressing the coefficients of the "likelihood ratio" (ratio of the probability density functions of the planted and the null distributions) in the Hermite basis – this is a standard strategy [KWB19, SW20] employed in proving such results. The performance of the low-degree polynomial distinguishers is related (again via standard ideas from prior works) to the truncated low-degree likelihood ratio – a natural quantity that depends on the density functions of the planted and the null models above. Our analysis then proceeds by combinatorial characterization and estimates for the Hermite coefficients of the planted density function.

**Sum-of-Squares lower bounds.** Our sum-of-squares lower bound shows that for a system of $m$ random homogeneous quadratic equations with RHS all set to 0, there is no degree-4 sum-of-squares refutation as long as $m \leqslant n^2/\operatorname{poly}(\log n)$. As is standard, we show such a statement by exhibiting a dual witness –a *pseudo-distribution* of degree 4 (see Definition 2.9) – that is consistent with the input system of polynomial equations.

More specifically, we view the equations as $\{x^\top G_i x = 0\}_{i \leqslant m}$ where each $G_i$ is a matrix with independent standard Gaussian entries. A pseudo-distribution of degree 4 satisfying such constraints is a linear map $\widetilde{\mathbb{E}}_\mu$ that assigns a real number to every degree $\leqslant 4$ polynomial and satisfies

[5] The degree-4 polynomial $(\sum_i b_i^2)^2$ is a distinguisher whp between the null and planted models.

1) **Normalization:** $\widetilde{\mathbb{E}}_\mu[1] = 1$, 2) **Positivity:** $\widetilde{\mathbb{E}}_\mu[q^2] \geqslant 0$ for every degree $\leqslant 2$ polynomial $q$, and 3) **Constraints:** $\widetilde{\mathbb{E}}_\mu[(x^\top G_i x)q] = 0$ for every degree-2 polynomial $q$ and every $i$.

Our construction of such a map uses *pseudo-calibration* – a general technique for constructing candidate dual witnesses discovered in [BHK+19]. Informally speaking, this technique gives a "mechanical" method of constructing a candidate pseudo-distribution for an average-case refutation problem on some null distribution given a *planted* distribution that is indistinguishable from the null by low-degree polynomials. Our construction (see Definition 6.2) is based on the planted distribution described above in the context of our proof of lower bounds for the low-degree method.

While pseudo-calibration makes the job of coming up with candidate pseudo-distributions easy, the analysis of the resulting construction still essentially needs to be done via techniques specific to a given setting (we note that the pseudo-calibration conjecture of [HKP+17] hypothesizes the existence of a more mechanical translation). Thus, the bulk of our technical work goes into analyzing the construction above.

Our idea (as is standard in such settings) is to use the Hermite polynomial basis to explicitly write down expressions for the pseudo-distribution. Analyzing the pseudo-distribution requires analyzing the spectrum of the *moment matrix* associated with the pseudo-distribution. The moment matrix $\mathcal{M}$ is indexed by indices of monomials $I, J$ of degree $\leqslant 2$ on rows and columns and has its $(I, J)$ entry given by $\widetilde{\mathbb{E}}_\mu[x^I x^J]$. In our case, note that this is a random matrix with heavily correlated entries. The positivity property of $\widetilde{\mathbb{E}}_\mu$ is equivalent to the positive semidefiniteness of the matrix $\mathcal{M}$.

Our analysis works by decomposing the $\mathcal{M}$ into a linear combination of *graph matrices* (analogs of the bases used in prior works e.g. [HKP+17, BHK+19, GJJ+20]) that form a good basis for analyzing the spectra of such correlated random matrices. Thankfully, some of the technology for understanding the spectra of such graph matrices – whose spectra can be directly related to combinatorial properties of associated graphs called *shapes* – was developed in the context of proving $n^{O(1)}$-degree sum-of-squares lower bounds for the Sherrington-Kirkpatrick Hamiltonian by [GJJ+20, AMP20].

The high-level outline of our analysis resembles the strategy adopted by [GJJ+20] though the details differ because of the difference in the structure of the pseudo-distribution. First, the construction above does not quite exactly satisfy the constraints but we show that an appropriately small perturbation of it does. To analyze this construction, we study the decomposition of the $\mathcal{M}$ and identify the shapes that are *negligible* (i.e. contribute sufficiently small singular values), *trivial* (these contribute a large positive semidefinite mass) and *spiders* – these can be "killed" – that is, one can show that the contributions of the corresponding terms adds up to 0.

While our analysis follows a similar high-level plan to [GJJ+20] so far, the combinatorial characterization of shapes that fall into each of the three types above differs from that [GJJ+20] and requires an analysis specialized to our setting. This is because [GJJ+20] work with a special form of "rank 1" polynomial constraints relevant to their setting $\{\langle x, g_i\rangle^2 = 1\}_{i \leqslant m}$ where the $g_i$s are random vectors (the "affine planes" problem). As a result, the resulting construction of pseudo-distribution leads to a moment matrix with a different set of shapes playing a prominent role – 2 uniform graphs as opposed to 3-uniform hypergraphs in our case.

With the above techniques, it is possible to obtain a lower bound that works whenever $m \ll n^{1.5}$ as in the work of [GJJ+20]. But just as in their setting, this bound is off from the optimal bound

of $\sim n^2$. This is entirely due to the difficulties in the analysis of the construction above.

In our setting, we are able to obtain (only at degree 4) an analysis of the construction that does work all the way to the threshold of $n^2$ up to polylogarithmic factors in $n$. One crucial ingredient is a further sub-classification of non-negligible shapes appearing in the decomposition into *disconnected* and the rest. We give a different "charging scheme" for the disconnected shapes in order to show that they cannot contribute negative eigenvalues to the spectrum of $\mathcal{M}$.

# 2 Preliminaries

## 2.1 Notations

We use the standard conventions $\mathbb{N} = \{0, 1, 2, \dots\}$ and $[N] = \{1, 2, \dots, N\}$. Consider vectors $x \in \mathbb{R}^N$ and $\alpha \in \mathbb{N}^N$. We use the notation $|\alpha| = \sum_{i=1}^N \alpha_i$ and $\alpha! = \prod_{i=1}^N (\alpha_i!)$, and further denote $x^\alpha := \prod_{i=1}^N x_i^{\alpha_i}$. Moreover, we say $\alpha$ is *simple* if $\alpha \in \{0,1\}^N$, i.e. the monomial $x^\alpha$ is multilinear. With slight abuse of notation, we will often treat $\alpha \in \mathbb{N}^N$ as a multiset of $[N]$.

In this paper, we will encounter the case where $N = m \times n \times n$. The same notations apply: for $\alpha := (\alpha^1, \dots, \alpha^m) \in \mathbb{N}^{m \times n \times n}$, $|\alpha| = \sum_{s \in [m]} \sum_{i,j \in [n]} \alpha_{ij}^s$ and $\alpha! = \prod_{s \in [m]} \prod_{i,j \in [n]} (\alpha_{ij}^s)!$. In this case, we may view $\alpha$ as a *labeled directed multigraph* (with self-loops allowed) on vertex set $[n]$, where each edge has a label in $[m]$. Thus, $|\alpha|$ is the total number of edges, and $|\alpha^s|$ is the number of edges labeled $s$.

In this work, we will deal with algorithms that operate on numerical inputs. In all such cases, we will rely on the standard word RAM model of computation and assume that all the numbers are rational represented as a pair of integers describing the numerator and the denominator. In order to measure the running time of our algorithms, we will need to account for the length of the numbers that arise during the run of the algorithm. The following definition captures the size of the representations of rational numbers:

**Definition 2.1** (Bit Complexity)**.** *The bit complexity of an integer $p \in \mathbb{Z}$ is $1 + \lceil \log_2 p \rceil$. The bit complexity of a rational number $p/q$ where $p, q \in \mathbb{Z}$ is the sum of the bit complexities of $p$ and $q$.*

## 2.2 Hermite polynomials

In this section, we introduce the *Hermite polynomials*, which are orthogonal polynomials with respect to the Gaussian measure (see [Sze39] for a standard reference). The *univariate Hermite polynomials* $\{h_k\}_{k \in \mathbb{N}}$ are defined by the following recurrence:

$$h_0(x) = 1, \quad h_1(x) = x, \quad h_{k+1}(x) = x h_k(x) - k h_{k-1}(x)\,.$$

Next, we define the *multivariate* Hermite polynomials. For an index $\alpha \in \mathbb{N}^N$ and vector $x \in \mathbb{R}^N$, $h_\alpha(x) := \prod_{i=1}^N h_{\alpha_i}(x_i)$. The Hermite polynomials form an orthogonal basis with respect to the Gaussian measure: for $\alpha_1, \alpha_2 \in \mathbb{N}^N$,

$$\mathbb{E}_{x \sim \mathcal{N}(0,\mathbb{I})} \left[ h_{\alpha_1}(x) h_{\alpha_2}(x) \right] = \begin{cases} \alpha_1! & \text{if } \alpha_1 = \alpha_2, \\ 0 & \text{otherwise.} \end{cases}$$

We will need the following facts about Hermite polynomials:

**Fact 2.2.** *For an even integer $k \geqslant 2$, $h_k(0) = (-1)^{k/2}(k-1)!!$. For an odd $k$, $h_k(0) = 0$.*

**Fact 2.3.** *For an even integer $k \geqslant 2$, $(k-1)!! \leqslant (\frac{k}{2})^{k/2}$.*

**Fact 2.4.** *For any $x \in \mathbb{R}$, the generating function of Hermite polynomials is the following,*

$$e^{xt-\frac{t^2}{2}} = \sum_{k=0}^{\infty} h_k(x)\frac{t^k}{k!}\,.$$

## 2.3 Sum-of-Squares and Nullstellensatz proofs vs algorithms

Sum-of-squares proof system is a restricted reasoning system for certifying unsatisfiability of a system of polynomial equality and inequality constraints over the reals. We refer the reader to the monograph [FKP19] for a detailed exposition.

**Definition 2.5** (Sum-of-Squares Refutations)**.** *Let $p_1, p_2, \ldots, p_k$ be polynomials in variables $x_1, \ldots, x_n$ with coefficients over the reals. Given a system of constraints $\{p_i \geqslant 0\}_{i\leqslant k}$, a sum-of-squares refutation of the system is a polynomial identity of the following form:*

$$-1 = \sum_{T\subseteq[k]} S_T \prod_{i\in T} p_i\,, \tag{1}$$

*where $S_0, S_1, \ldots, S_T$ are sum-of-squares polynomials. The* degree *of the sum-of-squares proof is the minimum positive integer $\ell$ such that for every $T \subseteq [k]$ such that $S_T \neq 0$, $\sum_{i\in T} \deg(p_i) + \deg(S_T) \leqslant \ell$.*

Observe that if an identity of the form (1) exists, then it immediately proves that the associated constraint system is unsatisfiable. This is because any $x$ that satisfies the constraint system must make the right hand side evaluate to a non-negative real number while the left hand side is the negative real $-1$. Under mild conditions on the polynomials $p_1, p_2, \ldots, p_k$, a converse holds. Such results are called *positivstellensatz*. We state a general one due to Krivine and Stengle [Kri64, Ste74].

**Fact 2.6** (Krivine/Stengle's Positivstellensatz [Kri64, Ste74], see Theorem 3.73 in [FKP19])**.** *Let $p_1, p_2, \ldots, p_k$ be $n$-variate real-coefficient polynomials in $x_1, x_2, \ldots, x_n$. If there does not exist $x \in \mathbb{R}^n$ such that $p_i(x) \geqslant 0$ for every $i \leqslant k$, then, there are sum-of-squares polynomials $\{S_T\}_{T\subseteq[k]}$ such that the following polynomial identity holds:*

$$-1 = \sum_{T\subseteq[k]} S_T \prod_{i\in T} p_i\,.$$

While positivstellensatz implies that there's always a refutation for all polynomial systems, it provides no bound on the degree of the resulting proof.

It is instructive to compare it with the strictly weaker Nullstellensatz proof system that we will also encounter in this work.

**Definition 2.7** (Nullstellensatz Refutation)**.** *Let $p_1, p_2, \ldots, p_k$ be polynomials in variables $x_1, x_2, \ldots, x_n$ with coefficients over the reals. Given a system of constraints $\{p_i = 0\}_{i\leqslant k}$, a Nullstellensatz refutation of the system is a polynomial identity of the following form:*

$$1 = \sum_{i\leqslant k} a_i p_i\,, \tag{2}$$

*where $a_1, \ldots, a_k$ are arbitrary polynomials. The* degree *of the Nullstellensatz proof is the minimum positive integer $\ell$ such that for every $i$, $\deg(p_i) + \deg(a_i) \leqslant \ell$.*

Unlike the sum-of-squares proof system, the Nullstellensatz proof systems only deals with polynomial *equality* constraints. Analogously to positivstellensatz, the completeness of the Nullstellensatz proof systems is implied by Hilbert's Nullstellensatz.

**Fact 2.8** (Corollary of Hilbert's Nullstellensatz, see for e.g., [Pit97])**.** *Suppose $p_1, p_2, \ldots, p_k$ are real-coefficient polynomials in $x_1, x_2, \ldots, x_n$ such that there is no $x$ satisfying $p_i(x) = 0$ for every $i \leqslant k$. Then, there are polynomials $a_1, a_2, \ldots, a_k$ with real coefficients such that the following polynomial identity holds:*

$$1 = \sum_{i \leqslant k} a_i p_i \,.$$

Informally speaking, the key difference between the sum-of-squares and the Nullstellensatz proof system is the ability to reason about the non-negativity of square polynomials. This seemingly minor change results in a huge difference in the power of the proof systems. For example, the pigeonhole principle requires $\Omega(n)$ degree for Nullstellensatz to refute but has a degree-4 SoS refutation (see Claim 3.59 on Page 125 of [FKP19] for a short proof).

## 2.4 Pseudo-distributions

Pseudo-distributions are generalizations of probability distributions and form dual objects to sum-of-squares proofs in a precise sense that we will describe below.

**Definition 2.9** (Pseudo-distribution, Pseudo-expectations, Pseudo-moments)**.** *A* degree-$\ell$ pseudo-distribution *is a finitely-supported function $\mu : \mathbb{R}^n \to \mathbb{R}$ such that $\sum_x \mu(x) = 1$ and $\sum_x \mu(x) f(x)^2 \geqslant 0$ for every polynomial $f$ of degree at most $\ell/2$. (Here, the summations are over the support of $\mu$.)*

*The* pseudo-expectation *of a function $f$ on $\mathbb{R}^d$ with respect to a pseudo-distribution $\mu$, denoted $\widetilde{\mathbb{E}}_{\mu(x)} f(x)$, as*

$$\widetilde{\mathbb{E}}_{\mu(x)} f(x) = \sum_x \mu(x) f(x) \,. \tag{3}$$

*The degree-$\ell$ moment tensor of a pseudo-distribution $\mu$ is the tensor $\mathbb{E}_{\mu(x)}(1, x_1, x_2, \ldots, x_n)^{\otimes \ell}$. In particular, the moment tensor has an entry corresponding to the pseudo-expectation of every monomial of degree at most $\ell$ in $x$.*

Observe that if a pseudo-distribution $\mu$ satisfies, in addition, that $\mu(x) \geqslant 0$ for every $x$, then it is a mass function of some probability distribution. Further, a straightforward polynomial-interpolation argument shows that every degree-$\infty$ pseudo-distribution satisfies $\mu \geqslant 0$ and is thus an actual probability distribution. The set of all degree-$\ell$ moment tensors of probability distribution is a convex set. Similarly, the set of all degree-$\ell$ moment tensors of degree-$d$ pseudo-distributions is also convex.

**Definition 2.10** (Constrained pseudo-distributions)**.** *Let $\mu$ be a degree-$\ell$ pseudo-distribution over $\mathbb{R}^n$. Let $\mathcal{A} = \{p_1 \geqslant 0, p_2 \geqslant 0, \ldots, p_m \geqslant 0\}$ be a system of $m$ real-coefficient polynomial inequality constraints. We say that $\mu$* satisfies the system of constraints $\mathcal{A}$ at degree $\ell$ *if for every sum-of-squares polynomial $h$ and any $T \subseteq [m]$ such that $\deg(h) + \sum_{i \in T} \deg(p_i) \leqslant \ell$, $\widetilde{\mathbb{E}}_\mu [h \cdot \prod_{i \in T} p_i] \geqslant 0$.*

The following fact describes the precise sense in which pseudo-distributions are duals to sum-of-squares proofs.

**Fact 2.11** (Strong Duality, [JH16], see Theorem 3.70 in [FKP19] for an exposition)**.** *Let $p_1, p_2, \ldots, p_k$ be real-coefficient polynomials in $x_1, x_2, \ldots, x_n$. Suppose there is a degree-$d$ sum-of-squares refutation of the system $\{p_i(x) \geqslant 0\}_{i \leqslant k}$. Then, there is no pseudo-distribution $\mu$ of degree $\geqslant d$ satisfying $\{p_i(x) \geqslant 0\}_{i \leqslant k}$. On the other hand, suppose that there is a pseudo-distribution $\mu$ of degree $d$ consistent with $\{p_i(x) \geqslant 0\}_{i \leqslant k}$. Suppose further that the set $\{p_1, p_2, \ldots, p_k\}$ contains the quadratic polynomial $R - \sum_i x_i^2$ for some $R > 0$. Then, there is no degree-$d$ sum-of-squares refutation of the system $\{p_i(x) \geqslant 0\}_{i \leqslant k}$.*

## 2.5 Algorithms and numerical accuracy

The sum-of-squares proof system is automatizable via semidefinite programming in an appropriate sense that we describe next. Informally, this means that degree-bounded sum-of-squares proofs and low-degree pseudo-distributions satisfying a system of constraints can be found via efficient algorithms. Such algorithms deal with numerical inputs and thus, in the context of algorithms, we only allow our input polynomial systems to have rational coefficients.

The following fact follows by using the ellipsoid algorithm for semidefinite programming. The resulting algorithm to compute pseudo-distributions approximately satisfying a given set of polynomial constraints is called the *sum-of-squares algorithm*.

**Fact 2.12** (Computing pseudo-distributions consistent with a set of constraints [Sho87, Par00, Nes00, Las01])**.** *There is an algorithm with the following properties: The algorithm takes input $B \in \mathbb{N}$, $\tau > 0$, and polynomials $p_1, p_2, \ldots, p_k$ of degree $\ell$ with rational coefficients of bit complexity $B$. If there is a pseudo-distribution of degree $d$ consistent with the constraints $\{p_i(x) = 0\}_{i \leqslant k}$, the algorithm in time $\mathrm{poly}(B, \frac{1}{\tau}) \cdot n^{O(d)}$ outputs a pseudo-distribution $\mu$ of degree $d$ satisfying $|\widetilde{\mathbb{E}}_\mu p_i(x) x^\alpha| \leqslant \tau$ if it exists and otherwise outputs "infeasible".*

## 2.6 Background on the low-degree polynomial method

The low-degree polynomial method is a restricted class of computationally bounded algorithms for hypothesis testing problems arising in statistics.

In order to describe this method, let $\nu_N$ (for "null") and $\nu_P$ (for "planted" distribution; often called the "alternative" distribution in statistics) be a pair of probability distributions on $\mathbb{R}^K$. Informally, we will set $\nu_N$ to be a distribution on instances of some optimization problem that admit no solutions with high probability (such as random polynomial systems in our case) while $\nu_P$ will be the distribution on random polynomial systems that always admit a solution.

In the hypothesis testing problem, the algorithm is given a sample $z$ with the promise that it is generated by the mixture $0.5\nu_N + 0.5\nu_P$. The goal is to determine correctly with high probability if $z$ is generated from $\nu_N$ or $\nu_P$. Often $\nu_N$ and $\nu_P$ are parameterized family of distributions (for e.g, the degree $D$, the number of variables $n$ or equations $m$ in our setting).

The key question is to determine the parameter regimes under which the hypothesis testing problem is solvable with high (say $1 - o_K(1)$) probability. Any such "testing" algorithm can be seen as computing some function $T_K : \mathbb{R}^K \to \mathbb{R}$ on the input sample $z$ and outputting "null" if $T_K(z)$ exceeds some threshold $\tau$. Observe that a family of tests $\{T_K\}_K$ succeeds with probability $1 - o_K(1)$ as $K \to \infty$ if $\mathbb{E}_{\nu_P} T_K - \mathbb{E}_{\nu_N} T_K \to \infty$ as $K \to \infty$.

Information-theoretically speaking, the classical Neyman-Pearson lemma identifies an optimal (in the sense of achieving optimal trade-off between false positives and false negatives) statistical test – the *likelihood ratio* – that distinguishes the given pair of distributions.

**Restricting to low-degree polynomial tests.** While the likelihood ratio test is statistically optimal, it is often hard to compute and thus does not yield an efficiently computable distinguisher. The *low-degree polynomial method* restricts the algorithm to a smaller class of statistical tests so as to gain computational efficiency.

Specifically, such tests $T$ are restricted to 1) evaluating some degree-$d$ polynomial $f$ on the input sample $z$ and 2) "accepting" if $f(z)$ exceeds some chosen threshold $\tau$. Such a test is clearly computable in $K^{O(d)}$ time by explicitly evaluating each monomial of $f$.

While such tests may appear restricted, recent works showed that $O(\log n)$-degree polynomial tests in fact can *simulate* algorithms such as power iteration (and thus computing spectral norms), approximate message passing, and local algorithms applied to $z$ and more generally matrices/tensors with entries set to constant-degree polynomials of $z$. This allows the method to capture the strongest known algorithms for fundamental distinguishing tasks including planted clique and spiked Wigner models, and more generally, random optimization problems such as clique/independent set and densest $k$-subgraph in random graphs. In what can be construed to be an even more evidence of the power of the method, recent work [BBH+20] shows that under appropriate restrictions, algorithms in the $O(\log n)$-degree polynomial model are as powerful as polynomial time algorithms in the statistical query model studied arising in learning theory and recently applied [FGR+17] to prove lower bounds for average-case variants of several foundational combinatorial and statistical learning problems. The low-degree likelihood ratio and the low-degree polynomial tests were introduced in the context of establishing sum-of-squares lower bounds implicitly in [BHK+19] and formalized explicitly in [HKP+17]. In particular, for average-case distinguishing problems satisfying some mild "niceness" conditions, [HKP+17] conjecture (this is called the *pseudo-calibration conjecture*) that indistinguishability by degree-$d$ polynomials implies lower bounds for a canonical $\widetilde{O}(d)$-degree SoS relaxation for the associated refutation problem.

Subsequent works (starting with [HS17], see Conjecture 2.2.4 in [Hop18] and 1.16 in [KWB19]) have proposed the stronger conjecture that concludes a lower bound against all $n^{\widetilde{O}(d)}$ time distinguishing algorithms.

The following definition presents a formal, quantitative version of what it means to use low-degree polynomials to distinguish between a pair of distributions as above.

**Definition 2.13** (Distinguishing by Low-Degree Polynomials)**.** *Let $\nu_N, \nu_P$ be a pair of null and planted distributions on $\mathbb{R}^K$. We say that degree-$d$ polynomials succeed in $(1-\delta)$-distinguishing between $\nu_N$ and $\nu_P$ from a single sample if there is a degree $\leqslant d$ polynomial $f : \mathbb{R}^K \to \mathbb{R}$ such that:*

1. $\mathbb{E}_{\nu_N}[f^2] = 1$.
2. $\mathbb{E}_{\nu_P}[f] \geqslant \frac{1}{\delta}$.

It turns out that it is possible to precisely characterize the *best* low-degree polynomial distinguisher $f$ in terms of the density functions of the associated pair of distributions.

**Proposition 2.14** (Truncated Low-Degree Likelihood Ratio; see [HKP+17] and Proposition 1.15 of [KWB19])**.** *Let $\nu_N, \nu_P$ be a pair of probability distributions on $\mathbb{R}^K$. The truncated low-degree likelihood ratio $L^{\leqslant d}$ at degree $d$ is defined as the unique solution to* $\arg\min_f \mathbb{E}_{z\sim\nu_N}[(L(z) - f(z))^2]$ *where the minimization is over all degree $\leqslant d$ polynomials $f$. The* normalized *truncated likelihood ratio* $L^{\leqslant d} / \mathbb{E}_{\nu_N}\left[(L^{\leqslant d})^2\right]^{1/2}$

*is then the optimal solution to the following optimization problem:*

$$\max \mathbb{E}_{\nu_P} f \quad \text{s.t. } \mathbb{E}_{\nu_N} f^2 = 1 \text{ and } f \text{ is a degree-}d \text{ polynomial.}$$

*Moreover, the value of the optimization problem is* $\mathbb{E}_{\nu_N}\left[(L^{\leqslant d})^2\right]^{1/2}$. *In particular,* $\nu_N$ *and* $\nu_P$ *are* $(1-\delta)$*-indistinguishable by degree* $\leqslant d$ *polynomials if* $\mathbb{E}_{\nu_N}\left[(L^{\leqslant d})^2\right]^{1/2} \leqslant \frac{1}{\delta}$.

# 3 Algorithmic Thresholds: Upper Bound

In this section, we describe and analyze our algorithm for refuting random polynomial systems. Our algorithmic results apply to all random polynomial systems where all coefficients are independent from some distribution on rational numbers that satisfies some niceness properties. Such properties are satisfied by the uniform distribution on a large enough subset of rational numbers, a polynomial bit truncation of the standard Gaussian distribution among others.

**Definition 3.1** (Nice Rational Distributions)**.** *For* $B \in \mathbb{N}$*, we say that a probability distribution* $\nu$ *on* $\mathbb{Q}$ *is B-nice if the following hold:*

1. $\nu$ ***is supported on low-bit complexity rationals****: The support of* $\nu$ *are rational numbers with numerator and denominator in* $[-2^B, 2^B]$.
2. $\nu$ ***is spread-out****: for any* $q \in \mathbb{Q}$, $\mathbf{Pr}_{x\sim\nu}[x = q] \leqslant \frac{1}{B^{100}}$.

The main result of this section is the following theorem:

**Theorem 3.2** (Refutation Algorithm for Random Polynomial Systems)**.** *Fix* $D \in \mathbb{N}$*. There is an algorithm with the following properties: the algorithm takes input m polynomial equations* $\{g_i(x) = b_i\}_{i\in[m]}$ *where each* $g_i$ *is a polynomial of degree D with rational coefficients of bit-complexity B, and in* $(Bn)^{O(d)}$ *time, either correctly outputs "infeasible" or returns "don't know". Further, if* $m \geqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$ *and* $g_i, b_i$ *are obtained by sampling each coefficient of each* $g_i$ *and each* $b_i$ *from (possibly different) independent* $n^{2d}$*-nice rational distributions, then, with probability* $1 - n^{-d}$ *over the choice of the input equations, the algorithm outputs "infeasible".*

Our algorithm is quite simple. It approximately solves the degree-$d$ SoS relaxation for the constraint system $\{p_i(x) = 0\}_{i\leqslant m}$ where $p_i(x) = g_i(x) - b_i$, and returns "infeasible" if the SDP outputs infeasible and "don't know" otherwise. More precisely:

**Algorithm 3.3** (Refute Random Polynomials)**.**

**Given:** A rational accuracy parameter $\tau = \exp\left(-n^{O(d)}B\right)$ and degree-$D$ polynomials $p_1, \ldots, p_m$ with rational coefficients of bit complexity at most $B$ for $B \in \mathbb{N}$.

**Output:** "Infeasible" or "Don't Know".

**Operation:**

1. Find a degree-$d$ pseudo-distribution $\mu$ such that $|\widetilde{\mathbb{E}}_\mu p_i(x)x^\alpha| \leqslant \tau$ for every $i \leqslant m$ and monomial index $\alpha$ of degree at most $d - \deg(p_i)$.

2. If no such pseudo-distribution exists, return "Infeasible".
3. Otherwise output "don't know".

**Analysis of algorithm.** The key to the proof of the theorem is the following lemma that guarantees the existence of a sum-of-squares refutation for the input random polynomial system.

**Lemma 3.4** (Sum-of-squares refutation for random polynomial systems)**.** *Let $D \in \mathbb{N}$ and $d \geqslant D$ be a multiple of $D$. Let $\mathcal{F} = \{g_1, \ldots, g_m\}$ be a set of homogeneous degree-$D$ polynomials with each coefficient of each $g_i$ chosen from an independent $B$-nice rational distribution. Let $b_1, b_2, \ldots, b_m$ be independent samples from a $B$-nice rational distribution. Then, whenever $m \geqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$, with probability at least $1 - n^{-d}$ over the choice of the $g_i$s and $b_i$s, there exist polynomials $a_1, a_2, \ldots, a_m$ of degree $d - D$ such that the following polynomial identity holds:*

$$-1 = \sum_{i \leqslant m} a_i(g_i - b_i)\,. \tag{4}$$

*Further, the coefficients of $a_i$s are rational numbers with bit complexity at most $O(n^{5d} d \log n + n^{5d} B)$.*

**Remark 3.5** (Nullstellensatz vs Sum-of-Squares)**.** Observe that in the refutation identity, there is no additive sum-of-squares term. As a result, our refutation is in fact a Nullstellensatz refutation (Definition 2.7). As we show, there's a strong indication (see the next section on lower bounds) that the trade-off achieved by Lemma 3.4 between $m$ and $d$ is tight up to absolute constant factors for the sum-of-squares proof system. Thus, in this case, we expect that the $m$ vs $d$ trade-off for Nullstellensatz and SoS proof systems to be essentially the same. Interestingly, the constant factor gap allowed by our upper and lower bounds might be "real". At degree $d = 2$, it is not hard to argue that $m \geqslant \frac{n^2}{2}$ is necessary for a Nullstellensatz refutation to exist. However, Theorem 5.1 shows that $m \gtrsim \frac{n^2}{4}$ is sufficient for degree-2 SoS.

It is easy to complete the analysis of the algorithm using this lemma.

*Proof of Theorem 3.2.* The running time of the algorithm follows immediately by applying Fact 2.12. In order to prove correctness of the algorithm, let's assume that for the given set of $g_i$s, a sum-of-squares proof of the form promised by Lemma 3.4 holds. Let $B' = O(n^{5d} d \log n + n^{5d} B)$ be an upper-bound on the bit complexity of the coefficients of $a_i$. By Lemma 3.4, such an event happens with probability $1 - n^{-d}$ over the choice of $g_i$s and $b_i$s. We will prove that conditioned on this event, the algorithm outputs "infeasible" with probability 1.

By Fact 2.12, if there is a pseudo-distribution of degree $d$ consistent with $\{g_i(x) = b_i\}_{i \leqslant m}$ then the sum-of-squares algorithm finds a pseudo-distribution $\mu$ such that $|\widetilde{\mathbb{E}}_\mu[x^\alpha(g_i - b_i)]| \leqslant \tau$ for each monomial index $\alpha$ of degree $\leqslant d - D$. We will show that there does not exist a pseudo-distribution satisfying the latter condition. Thus, the SDP solver must output "infeasible" as desired.

Assume for the sake of contradiction that for $\tau = 0.5 \cdot 2^{-B'}(n+1)^{-d} m^{-1}$, there is a pseudo-distribution $\mu$ satisfying $|\widetilde{\mathbb{E}}_\mu[(g_i - b_i)x^\alpha]| \leqslant \tau$ for every $i \leqslant m$ and every monomial index $\alpha$ of degree $\leqslant d - D$. Then, since all $\leqslant (n+1)^d$ coefficients of each of the $a_i$ are of bit complexity at most $B'$, the pseudo-expectation under $\mu$ of the RHS of (4) can be upper-bounded by:

$$\left|\sum_{i=1}^{m} \widetilde{\mathbb{E}}_\mu[a_i(g_i - b_i)]\right| \leqslant \sum_{i \leqslant m, \alpha} 2^{B'}\tau \leqslant m(n+1)^d 2^{B'}\tau \leqslant 0.5\,.$$

On the other hand, the pseudo-expectation under $\mu$ of the LHS satisfies: $|\widetilde{\mathbb{E}}[-1]| = 1$. This is a contradiction. Thus, there is no such pseudo-distribution $\mu$. □

### 3.1 Proof of Lemma 3.4

**Generated ideals.** Our analysis relies on the key idea of *generated ideals* and their completeness that we define and discuss below. Intuitively speaking, given a set of constraints $\mathcal{F} = \{f_1 = 0, f_2 = 0, \ldots, f_m = 0\}$, the generated ideal of $\mathcal{F}$ at degree $d$ is the set of all degree-$d$ polynomials that the sum-of-squares proof system (and in fact, the Nullstellensatz proof system) can infer to be 0 at any simultaneous solutions of $\mathcal{F}$. The following definition captures this idea.

**Definition 3.6** (Generated Ideal at Degree $d$)**.** *Let $D, d \in \mathbb{N}$ and $D \leqslant d$. Let $\mathcal{F} = \{f_1, \ldots, f_m\}$ be a set of degree-$D$ polynomials. The* generated ideal *of $\mathcal{F}$ at degree $d$ is defined as the following set of degree-$d$ polynomials:*

$$\mathrm{Gen}_d(\mathcal{F}) := \{a_1 f_1 + \cdots + a_m f_m : \forall i \ \deg(a_i) \leqslant d - D\}.$$

*We say that the generated ideal is* complete *at degree $d$ if $\mathcal{P}_d \subseteq \mathrm{Gen}_d(\mathcal{F})$ where $\mathcal{P}_d$ is the set of all homogeneous degree-$d$ polynomials.*

One important consequence of completeness of generated ideals at degree $d$ is the following important lemma that shows that every homogeneous degree-$d$ polynomial can be written as a polynomial combination of the $f_i$s such that the coefficients of all the polynomials appearing in the representation are of polynomial bit complexity.

**Lemma 3.7** (Low-Bit Complexity Representations in Complete Generated Ideals)**.** *Let $D, d \in \mathbb{N}$ and $D \leqslant d$. Let $\mathcal{F} = \{g_1, \ldots, g_m\}$ be a set of degree-$D$ polynomials with rational coefficients of bit complexity $B$ such that the generated ideal $\mathrm{Gen}_d(\mathcal{F})$ is complete. Let $N_{d-D} \leqslant n^{d-D}$ be the number of all monomials in $x_1, x_2, \ldots, x_n$ of total degree exactly $d - D$. Let $f$ be an arbitrary homogeneous polynomial of degree $d$ with rational coefficients of bit-complexity $B$.*

*Then, there is a vector $v \in \mathbb{Q}^{m \cdot N_{d-D}}$ with entries of bit complexity at most $O(n^{5d} d \log n + n^{5d} B)$ such that $\sum_{i \leqslant m, \alpha} v_{i,\alpha} g_i(x) x^\alpha = f(x)$.*

We will use the following fact that appears in a classical work of Kannan [Kan85].

**Fact 3.8** (Bit-Complexity of Solutions to Integer Systems, see Proposition 2.1 in [Kan85])**.** *Let $Ax = u$ for $A \in \mathbb{Z}^{m \times n}$ and $u \in \mathbb{Z}^m$ be a system of $m$ linear equations in $n$ variables $x$ such that each entry of $A$ and $u$ is an integer of magnitude $\leqslant B$.*

*Suppose that the system is soluble over $\mathbb{Q}$ – i.e., there is an $x \in \mathbb{Q}^n$ such that $Ax = u$. Then, there is in fact an $x \in \mathbb{Q}^n$ such that $Ax = u$ where the entries of $x$ have bit complexity $O(n(B + \log n))$.*

*Proof of Lemma 3.7.* Since $\mathrm{Gen}_d(\mathcal{F})$ is complete at degree $d$ and $f$ is a homogeneous polynomial of degree $d$, $f$ must belong to $\mathrm{Gen}_d(\mathcal{F})$. Thus, there are polynomials $a_1, a_2, \ldots, a_m$ of degree $\leqslant d - k$ such that $\sum_i a_i g_i = f$.

For each $i$, write $a_i(x) = \sum_\alpha a_i^\alpha x^\alpha$ where the sum ranges over monomial indices $\alpha$ of total degree $\leqslant d - k$. Then, we know that $f = \sum_{i,\alpha} a_i^\alpha x^\alpha g_i$. By matching the $\leqslant (n+1)^d$ coefficients of $f$ on both sides, we obtain a system of linear equations with rational coefficients. We are guaranteed that this system has a solution over the reals. In fact, since all the coefficients are rational numbers, we can infer that there must be a solution over the rationals.

Each coefficient in this linear system is a sum of at most $m$ different coefficients of one from each $g_i$. Since each coefficient of each $g_i$ has bit complexity at most $B$, the coefficients of the resulting linear system have bit complexity at most $B + O(d \log n)$.

The lowest common multiple of all the denominators appearing in the $\leqslant (n+1)^{2d}$ entries of the equation is at most their product that has bit complexity at most $O(n^{3d} d \log n + n^{3d} B)$. By multiplying all the equations by this integer, we obtain a system of linear equations over the integers. By Fact 3.8, such a system has a solution of bit complexity at most $O(n^{5d} d \log n + n^{5d} B)$. Thus the original system has a solution over the rationals with bit complexity at most $O(n^{5d} d \log n + n^{5d} B)$. This completes the proof. □

Our task thus reduces to showing that the generated ideal of the input polynomials is complete at degree $d$ when $m \geqslant O(n) \cdot \left(\frac{n}{d}\right)^{D-1}$.

**Completeness of generated ideal at degree $d$.** The key to the proof of Lemma 3.4 is the following lemma that identifies a non-trivial $d$ such that the generated ideal at degree $d$ of a collection of $m$ random polynomials is complete.

**Lemma 3.9** (Completeness of Generated Ideals)**.** *Let $D \in \mathbb{N}$ be a constant, let $d, n \in \mathbb{N}$ such that $2 \leqslant D \leqslant d \leqslant n$, and let $m \geqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$. Suppose $\mathcal{G} = \{g_1(x) - b_1, \ldots, g_m(x) - b_m\}$ is a set of $m$ degree-$D$ polynomials obtained by choosing each coefficient of each $g_i$ and each $b_i$ from independent $n^{2d}$-nice rational distributions. Then, the generated ideal of $\mathcal{G}$ at degree $d$ is complete with probability $1 - n^{-d}$.*

*Proof of Lemma 3.4 by Lemma 3.9.* Consider the first polynomial equation $g_1(x) = b_1$. Then, $b_1 \neq 0$ with probability $1 - n^{-200d}$ since it is sampled from a $n^{2d}$-nice distribution. Let's condition on $b_1 \neq 0$ in the following. Let $p(x) := \frac{1}{b_1} g_1(x)$, and let $q(x) := \frac{1}{b_1} \sum_{k=0}^{d/D-1} p(x)^k$ (since $d$ is a multiple of $D$). Then, we have

$$(g_1(x) - b_1) q(x) = p(x)^{d/D} - 1\,. \tag{5}$$

Thus, the polynomial $p^{d/D} - 1 \in \mathrm{Gen}_d(\mathcal{G})$ and moreover $p^{d/D}$ is a homogeneous polynomial of degree $d$. Thus, by Lemma 3.9, the following polynomial identity holds for some polynomials $a_1, a_2, \ldots, a_m$ of degree $\leqslant d - D$ such that each coefficient has bit-complexity $O(n^{5d}(B + \log n))$:

$$\sum_{i=1}^{m} (g_i(x) - b_i) a_i(x) = -p(x)^{d/D}\,, \tag{6}$$

Adding the identities from (5) and (6), we obtain:

$$\sum_{i=1}^{m} (g_i(x) - b_i) a_i(x) + (g_1(x) - b_1) q(x) = -1\,.$$

This completes the proof. □

We now focus on proving Lemma 3.9.

**Reduction to rank lower bounds.** Let $f$ be an arbitrary polynomial such that there exist polynomials $a_1, a_2, \ldots, a_m$ of degree $d - D$ such that $f = \sum_{i=1}^{m} (g_i - b_i)a_i \in \mathrm{Gen}_d(\mathcal{G})$. This polynomial identity holds if and only if the coefficients of $a_i$s satisfy a system of linear equations as we describe next. To prove that $\mathrm{Gen}_d(\mathcal{G})$ is complete, it suffices to restrict the polynomials $a_i$ to be homogeneous degree $d - D$.

For every $i \in [m]$, let $g_i(x) = \sum_{\gamma:|\gamma|=D} \widehat{g}_i(\gamma)x^\gamma$ where $\gamma \in \mathbb{N}^n$ ranges over indices of monomials in $x_1, x_2, \ldots, x_n$ of total degree $D$. Let $f(x) = \sum_{|\alpha|=d,d-D} \widehat{f}(\alpha)x^\alpha$ and $a_i(x) = \sum_{|\beta|=d-D} \widehat{a}_i(\beta)x^\beta$, where $\alpha, \beta \in \mathbb{N}^n$ are multisets indexing monomials in $x_1, x_2, \ldots, x_n$. Then, we have

$$f(x) = \sum_{|\alpha|=d,d-D} \widehat{f}(\alpha)x^\alpha = \sum_{i=1}^{m} \sum_{|\gamma|=D} \sum_{|\beta|=d-D} \widehat{g}_i(\gamma) \cdot \widehat{a}_i(\beta)x^{\beta+\gamma} - \sum_{|\beta|=d-D} \sum_{i=1}^{m} b_i \cdot \widehat{a}_i(\beta)x^\beta .$$

Comparing coefficients on both sides, we get $\widehat{f} = M_{g,b} \cdot \widehat{a}$, where $\widehat{f}$ has dimension $\binom{n+d-1}{d} + \binom{n+d-D-1}{d-D}$ (the number of degree $d$ and $d - D$ monomials) and $\widehat{a}$ has dimension $\binom{n+d-D-1}{d-D}$.

Let's write such equations as $f$ varies over all monomials of total degree exactly $d$. If all the resulting equations admit a solution, then clearly, every homogeneous polynomial of degree $d$ is in $\mathrm{Gen}_d(\mathcal{G})$. The coefficient matrix $M_{g,b}$ of the resulting linear system has the following structure:

$$M_{g,b} = \begin{bmatrix} M_g \\ M_b \end{bmatrix} .$$

Here, the rows of $M_g$ and $M_b$ are indexed by multisets $\alpha$ with $|\alpha| = d$ and $d - D$, respectively. The columns of $M_{g,b}$ are indexed by $(\beta, i)$ with $|\beta| = d - D$ and $i \in [m]$. Writing out the entries of $M_{g,b}$ explicitly:

$$M_g(\alpha, (\beta, i)) = \begin{cases} \widehat{g}_i(\gamma) & \alpha = \beta + \gamma \text{ where } |\gamma| = D \\ 0 & \text{otherwise} \end{cases}, \quad M_b(\alpha', (\beta, i)) = \begin{cases} -b_i & \alpha' = \beta \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

To prove Lemma 3.9, it suffices to show that $M_{g,b}$ has full row rank. We will prove this by showing that that the rows of $M_{g,b}$ are linearly independent.

**Lemma 3.10.** *Let $D \in \mathbb{N}$ be a constant, let $d, n \in \mathbb{N}$ such that $2 \leqslant D \leqslant d \leqslant n$, and let $B \geqslant n^{2d}$. Consider the matrix $M_{g,b}$ defined in (7), where each nonzero entry is sampled from a $B$-nice rational distribution. If $m \geqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$, then the rows of $M_{g,b}$ are linearly independent with probability $1 - n^{-d}$.*

**Remark 3.11.** Observe that $m$ must be at least $\binom{n+d-1}{d}/\binom{n+d-D-1}{d-D} + 1$ for $M_{g,b}$ to have more columns than rows. Thus, for small $d$ (e.g. $d = o(n)$), $m \geqslant \Omega(\frac{n^D}{d^D})$ is necessary for the generated ideal of $\mathcal{G}$ at degree $d$ to be complete.

Lemma 3.9 is an immediate corollary of Lemma 3.10. We proceed to prove Lemma 3.10 in the next section.

### 3.2 Rank lower bound by row-decomposition of $M_{g,b}$

To prove that $M_{g,b}$ is full row rank, it's enough to work with an appropriate permutation of rows/columns and delete any column from $M_{g,b}$. If the modified matrix is full row rank, then the original matrix is full row rank as well.

The main insight in the proof is that although $M_{g,b}$ is difficult to analyze, we can extract square submatrices $M_1, \ldots, M_N$ of $M_{g,b}$ that are full rank, and more importantly, can be "stitched together" to show that $M_{g,b}$ is full row rank. To do so, we define the following,

**Definition 3.12** (row-rank decomposition)**.** *We say that the collection $(M_1, \ldots, M_N)$ of square submatrices of $M_{g,b}$ is a row-rank decomposition of $M_{g,b}$ if*

1. *they cover all the rows of $M_{g,b}$ (i.e. each row of $M_{g,b}$ appears in at least one $M_i$),*
2. *they have disjoint columns of $M_{g,b}$ (i.e. no column of $M_{g,b}$ appears in more than one $M_i$),*
3. *the entries in the diagonal of each $M_i$ are independent of the off-diagonal entries of $M_i$ and the entries of $M_j$ for every $j \neq i$.*

The following lemma illustrates why the existence of a row-rank decomposition suffices to prove that $M_{g,b}$ is full row rank.

**Lemma 3.13.** *Let $A, B$ be submatrices of a matrix $M$ such that $A$ is full row rank and $A, B$ have disjoint columns. Let $M'$ be the submatrix of $M$ with rows (columns, respectively) equal to the union of rows (columns, respectively) of $A, B$. Suppose further that $B$ is $K \times K$ for some $K \leqslant n^{2d}$ and $B = B' + g\mathbb{I}$, where $g$ is a scalar sampled from a $n^{2d}$-nice rational distribution independent of $B'$ and the other entries in $M'$. Then, $M'$ is full row rank with probability $1 - n^{-100d}$.*

*Proof.* First, we write $M'$ (up to permutations of rows and columns) as

$$M' = \begin{bmatrix} A' & C_2 \\ C_1 & B \end{bmatrix} = \begin{bmatrix} A' & C_2 \\ C_1 & B' + g\mathbb{I} \end{bmatrix},$$

where $A'$ is the matrix $A$ with the rows that overlap with $B$ removed (those rows are now in $C_1$). $A'$ may not be square, but since $A'$ is still full row rank (the rows are linearly independent), we may delete some columns from $A'$ (and $C_1$) such that $A'$ is square and full rank. Hence, we may assume that $A'$ and $M'$ are square matrices without loss of generality.

$A'$ being full rank implies that $(A')^{-1}$ exists. Then, $M'$ is full rank if and only if the Schur complement

$$B - C_1(A')^{-1}C_2 = g\mathbb{I} + B' - C_1(A')^{-1}C_2$$

is full rank. Suppose not, then the matrix $g\mathbb{I} + B' - C_1(A')^{-1}C_2$ is rank-deficient, which implies that $g$ is an eigenvalue of $C_1(A')^{-1}C_2 - B'$. However, since $g$ is sampled from a $n^{2d}$-nice distribution and is independent of $C_1, C_2, A', B'$, the probability that $g$ is exactly one of the $K$ eigenvalues is $\leqslant Kn^{-200d} \leqslant n^{-100d}$. □

As an immediate corollary,

**Corollary 3.14.** *Let $d \in \mathbb{N}$ and $B \geqslant n^{2d}$. If there exists a row-rank decomposition $(M_1, \ldots, M_N)$ of $M_{g,b}$ for $N \leqslant n^{2d}$, then $M_{g,b}$ is full row rank with probability $1 - n^{-d}$.*

*Proof.* We apply Lemma 3.13 inductively to $M_1, M_2, \ldots, M_N$. Each submatrix $M_i$ has dimension at most $n^{2d}$, thus by the union bound, $M_{g,b}$ is full row rank with probability $1 - Nn^{-100d} \geqslant 1 - n^{-d}$. □

Thus, to prove Lemma 3.10, it suffices to construct a row-rank decomposition. For clarity of exposition, we will first prove Lemma 3.10 for the special case of $D = 2$ in the subsequent sections, and then show how the ideas extend to the case of $D > 2$ in Section 3.4.

### 3.3 Proof of Lemma 3.10, $D = 2$ case

Recall that the rows and columns of $M_{g,b}$ are indexed by $\alpha$ and $(\beta, i)$ respectively, where $\alpha, \beta$ are multisets with $|\alpha| = d$ or $d-2$, $|\beta| = d-2$, and $i \in [m]$. To ensure that the decomposition have disjoint columns, the submatrices will be constructed using different $i$, i.e. selected from disjoint subsets of $[m]$. Thus, we need $m$ to be sufficiently large so that we have enough "fresh random equations" to select from. Using different $i$ also ensures that each (random) submatrix is independent of each other, especially the diagonal entries. All other columns not present in the decomposition are ignored since we can delete columns arbitrarily.

**Covering rows of $M_g$.** To extract a submatrix for the decomposition, we first select a pair $\gamma = \{j_1, j_2\} \subseteq [n]$ ($j_1 = j_2$ is allowed) and consider all multisets $\alpha$ such that $\alpha = \beta \cup \gamma$ where $|\beta| = d-2$, and pick one "fresh" $i \in [m]$. This gives a square submatrix $A_\gamma$ where the columns are indexed by $\beta$ and the rows are indexed by $\alpha = \beta \cup \gamma$ and the entries are defined to be:

$$A_\gamma(\alpha, \beta) = \begin{cases} \widehat{g}_i(\alpha \setminus \beta) & \text{if } \beta \subset \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

For example, say $\gamma = \{1,2\}$ and $d = 4$; the first three columns are indexed by $\{1,1\}, \{1,2\}, \{1,3\}$, and the first three rows are indexed by $\{1,1,1,2\}, \{1,1,2,2\}, \{1,1,2,3\}$.

$$A_{\{1,2\}} = \begin{bmatrix} \widehat{g}_i(\{1,2\}) & \widehat{g}_i(\{1,1\}) & 0 & \cdots & 0 \\ \widehat{g}_i(\{2,2\}) & \widehat{g}_i(\{1,2\}) & 0 & \cdots & 0 \\ \widehat{g}_i(\{2,3\}) & \widehat{g}_i(\{1,3\}) & \widehat{g}_i(\{1,2\}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \widehat{g}_i(\{1,2\}) \end{bmatrix}.$$

Note that there are non-zero off-diagonal entries, but they are all independent of the diagonal entries $\widehat{g}_i(\gamma)$. By Lemma 3.13, $A_\gamma$ is full rank with high probability and satisfies the conditions of the row-rank decomposition (Definition 3.12). Crucially, all multisets $\alpha$ containing $\gamma$ are covered by $A_\gamma$.

Now, to construct the row-rank decomposition, we will select pairs $\gamma_1, \ldots, \gamma_N$ such that $|\gamma_k| = 2$ and that $A_{\gamma_1}, \ldots, A_{\gamma_N}$ cover all $\alpha$s with $|\alpha| = d$.

**Lemma 3.15.** *Let $2 \leqslant d \leqslant n$. There exist pairs $\gamma_1, \ldots, \gamma_N$ for $N \leqslant \frac{n^2}{2(d-1)} + O(n)$ such that the rows of $A_{\gamma_1}, \ldots, A_{\gamma_N}$ cover all multisets $\alpha$ of size $d$.*

*Proof.* First, we split $[n]$ into $d-1$ buckets, each bucket contains at most $\lceil \frac{n}{d-1} \rceil$ items. Within each bucket, we choose all pairs in the bucket, giving us $\binom{\lceil \frac{n}{d-1} \rceil}{2} + \lceil \frac{n}{d-1} \rceil$ pairs. The total number of pairs

$$(d-1)\binom{\lceil \frac{n}{d-1} \rceil}{2} \leqslant \frac{n^2}{2(d-1)} + O(n),$$

using the fact that $\lceil \frac{n}{d-1} \rceil \leqslant \frac{n}{d-1} + 1$.

Now, it suffices to prove that all $\alpha$s are covered. Observe that any multiset $\alpha$ that intersects a bucket in more than 1 element must be covered: if the intersection contains $\{j_1, j_2\}$, then $A_{\{j_1,j_2\}}$ covers $\alpha$. Thus, any uncovered $\alpha$ can only have 1 element in each bucket. However, there are only $d-1$ buckets whereas $|\alpha| = d$, hence every $\alpha$ must be covered. □

**Covering rows of $M_b$.** We use a single submatrix to cover all rows of $M_b$ (indexed by $\alpha$ with $|\alpha| = d-2$). Since $M_b(\alpha, (\beta, i)) = -b_i$ when $\alpha = \beta$ and 0 otherwise, we simply take the submatrix of a single $i$:

$$B = -b_i \cdot \mathbb{I}.$$

$B$ is full rank and cover all rows in $M_b$.

**Putting things together.** Recall that $M_{g,b}$ being full row rank (Lemma 3.10) implies that the generated ideal of $\mathcal{G} = \{g_1(x) - b_1, \ldots, g_m(x) - b_m\}$ at degree $d$ is complete (Lemma 3.9), which then implies our refutation result (Lemma 3.4).

*Proof of Lemma 3.10, $D = 2$ case.* By Lemma 3.15, the submatrices $A_{\gamma_1}, \ldots, A_{\gamma_N}$ cover the rows of $M_g$, and $B$ covers the rows in $M_b$. Together they form a valid row-rank decomposition of $M_{g,b}$.

The total number of equations required is

$$N + 1 \leqslant \frac{n^2}{2(d-1)} + O(n).$$

Thus, by Corollary 3.14, as long as $m \geqslant \frac{n^2}{2(d-1)} + O(n)$, $M_{g,b}$ is full row rank with probability $1 - n^{-d}$. □

### 3.4 Proof of Lemma 3.10, $D > 2$ case

The proof strategy is very similar to the case of $D = 2$: we construct a row-rank decomposition of $M_{g,b}$ by considering the rows of $M_g$ and $M_b$ separately. We first prove the following analog of Lemma 3.15,

**Lemma 3.16.** *Let $D \in \mathbb{N}$ be a constant and $d \in \mathbb{N}$ such that $3 \leqslant D \leqslant d \leqslant n$. There exist multisets $\gamma_1, \ldots, \gamma_N$ of size $D$ for $N \leqslant O\left(\frac{n^D}{d^{D-1}}\right)$ such that the rows of $A_{\gamma_1}, \ldots, A_{\gamma_N}$ cover all multisets $\alpha$ of size $d$.*

*Proof.* We split $[n]$ into $t := \left\lfloor \frac{d-1}{D-1} \right\rfloor$ buckets of size at most $\left\lceil \frac{n}{t} \right\rceil$. Within each bucket, we choose all size-$D$ multisets, which gives $\binom{\left\lceil \frac{n}{t} \right\rceil + D - 1}{D}$. The total number is

$$t \cdot \binom{\left\lceil \frac{n}{t} \right\rceil + D - 1}{D} \leqslant \frac{t}{D!} \left(\frac{n}{t} + D\right)\left(\frac{n}{t} + D - 1\right)\left(\frac{n}{t} + D - 2\right) \cdots \leqslant \frac{t}{D!} \left(\frac{n}{t} + D - 1\right)^D,$$

using the fact that $\left\lceil \frac{n}{t} \right\rceil \leqslant \frac{n}{t} + 1$. Next, we have $\frac{d}{D-1} \leqslant 2 \cdot \left\lfloor \frac{d-1}{D-1} \right\rfloor = 2t$ since $d \geqslant D \geqslant 3$. Furthermore, by $(D-1)t \leqslant d - 1 \leqslant n$ and $(D-1)! \geqslant \left(\frac{D-1}{e}\right)^{D-1}$,

$$\frac{t}{D!} \left(\frac{n}{t} + D - 1\right)^D \leqslant \frac{(4e)^D}{D} \cdot \frac{n^D}{d^{D-1}} = O_D\left(\frac{n^D}{d^{D-1}}\right).$$

By construction, any uncovered $\alpha$ can only intersect each bucket in $D-1$ elements, hence $|\alpha|$ is at most $t(D-1) < d$, contradicting that $|\alpha| = d$. Therefore, all $\alpha$s are covered. □

Finally, the same matrix $B$ covers all rows of $M_b$. Thus, we have a row-rank decomposition and are ready to prove Lemma 3.10.

*Proof of Lemma 3.10.* The submatrices $A_{\gamma_1}, \ldots, A_{\gamma_N}$ and $B$ together form a valid row-rank decomposition of $M_{g,b}$. By Lemma 3.16, the total number of equations required is

$$N + 1 \leqslant O_D\left(\frac{n^D}{d^{D-1}}\right).$$

Thus, by Corollary 3.14, as long as $m \geqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$, $M_{g,b}$ is full row rank with probability $1 - n^{-d}$. □

# 4 Algorithmic Thresholds: Lower Bounds

In this section, we prove a lower bound for the problem of distinguishing random polynomial systems from a carefully constructed "planted" distribution on random polynomial systems that admit a solution with probability 1. This algorithmic task is formally easier than refutation: observe that any refutation algorithm for random polynomial systems also serves as a distinguishing algorithm. Our lower bounds hold for algorithms in the restricted computation model called the *low-degree polynomial method* and match (up to constant factors) the trade-offs achieved by our refutation algorithm from the previous section.

Specifically, we will prove the following theorem in this section.

**Theorem 4.1** (Low-Degree Hardness of Distinguishing Planted vs Null Polynomial Systems)**.** *Let $D \geqslant 2$ be a constant and $d, n, m \in \mathbb{N}$. Let $\nu_N$ be the probability distribution of the system of degree-$D$ $n$-variate polynomial equations $\{g_i(x) = b_i\}_{i\in[m]}$ such that $g_i(x) = \langle G_i, x^{\otimes D}\rangle$ for a $D$-th order coefficient tensor $G_i$ such that each entry of $G_i$ is chosen to be an independent standard Gaussian. Then, for every $d \leqslant \frac{2n}{D}$, whenever $m \leqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$, there exists a probability distribution $\nu_P$ supported on solvable systems of $m$ polynomial equations such that degree-$d$ polynomials fail to $(1/2)$-distinguish between $\nu_N$ and $\nu_P$.*

**Remark 4.2.** The random polynomials appearing in the theorem above are obtained by choosing a random-entry tensor instead of choosing the coefficients of the polynomial directly. This leads to the coefficients of different monomials to have variances that differ by constant factors. This choice is convenient for our analysis but not necessary for the result to hold though we do not formally prove this.

We will prove Theorem 4.1 by exhibiting an explicit planted distribution defined below.

**Definition 4.3** (Planted distribution $\nu_P$)**.** *Fix a parameter $c = o\left(\frac{1}{d\sqrt{m}}\right)$, the planted distribution $\nu_P$ is sampled as follows,*

1. *Sample $z$ uniformly from $\{\pm\frac{1}{\sqrt{n}}\}^n$.*
2. *For each $i \in [m]$, sample $b_i \sim \mathcal{N}(0,1)$ independently.*
3. *For each $i \in [m]$, sample tensor $G_i \in (\mathbb{R}^n)^{\otimes D}$ with i.i.d. standard Gaussian entries conditioned on $\langle G_i, z^{\otimes D}\rangle = cb_i$.*

From Proposition 2.14, the task of proving indistinguishability of $\nu_N$ and $\nu_P$ by low-degree polynomials reduces to analyzing the truncated low-degree likelihood $L^{\leqslant d}$ of the pair $\nu_N$ and $\nu_P$. We will analyze $L^{\leqslant d}$ by computing a Hermite expansion for it:

$$L^{\leqslant d}(G,b) = \sum_{\alpha,\beta:|\alpha|+|\beta|\leqslant d} \widehat{L}_{\alpha,\beta} \cdot h_\alpha(G) h_\beta(b)$$

To analyze $L^{\leqslant d}$, we will show the following key technical claim:

**Lemma 4.4.** *Let $D \geqslant 2$ be a constant, let $d, n, m \in \mathbb{N}$ such that $0 < d \leqslant \frac{2n}{D}$, and let $c = o\left(\frac{1}{d\sqrt{m}}\right)$. Let $\{h_\beta\}_{|\beta|\leqslant d}$ be the multivariate Hermite polynomials. If $m \leqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$, then*

$$\sum_{\substack{\alpha,\beta:\\ 1\leqslant|\alpha|+|\beta|\leqslant d}} \mathbb{E}_{\nu_P}\left[h_\alpha(G)h_\beta(b)\right]^2 \leqslant 1\,.$$

We finish the proof of Theorem 4.1 modulo this claim:

*Proof of Theorem 4.1 by Lemma 4.4.* From Definition 2.13, it's enough to prove that $\mathbb{E}_{\nu_N}[(L^{\leqslant d})^2] \leqslant 2$. We first write the Hermite expansion of $L^{\leqslant d}$, as a function of $G$ and $b$, in the (unnormalized) Hermite basis,

$$L^{\leqslant d}(G,b) = \sum_{\alpha,\beta:|\alpha|+|\beta|\leqslant d} \widehat{L}_{\alpha,\beta} \cdot h_\alpha(G) h_\beta(b)$$

where $\alpha \in \mathbb{N}^{m\times n\times n}$ and $\beta \in \mathbb{N}^m$ are the Hermite indices. Since $\{h_\alpha(G)h_\beta(b)\}_{\alpha,\beta}$ are orthogonal with respect to $\nu_N$, the degree $\leqslant d$ Hermite coefficients of $L^{\leqslant d}$ equal that of $L$.

Thus, the Hermite coefficients $\widehat{L}_{\alpha,\beta}$ can be computed as:

$$\begin{aligned}\widehat{L}_{\alpha,\beta} &= \mathbb{E}_{(G,b)\sim\nu_N}\left[L(G,b)\cdot h_\alpha(G)h_\beta(b)\right]\cdot\frac{1}{\alpha!\beta!} = \mathbb{E}_{(G,b)\sim\nu_N}\left[(\nu_P/\nu_N)\cdot h_\alpha(G)h_\beta(b)\right]\cdot\frac{1}{\alpha!\beta!}\\ &= \mathbb{E}_{(G,b)\sim\nu_P}\left[h_\alpha(G)h_\beta(b)\right]\cdot\frac{1}{\alpha!\beta!}\,.\end{aligned}$$

Note that for $\alpha,\beta = \vec{0}$ (the first coefficient), $\widehat{L}_{0,0} = 1$. Then, by Lemma 4.4,

$$\mathbb{E}_{\nu_N}\left[(L^{\leqslant d})^2\right] = \sum_{|\alpha|+|\beta|\leqslant d}\widehat{L}^2_{\alpha,\beta}\cdot\alpha!\beta! = 1 + \sum_{1\leqslant|\alpha|+|\beta|\leqslant d}\mathbb{E}_{(G,b)\sim\nu_P}\left[h_\alpha(G)h_\beta(b)\right]^2\cdot\frac{1}{\alpha!\beta!} \leqslant 2\,.$$

Along with Proposition 2.14, this shows that the value $\mathbb{E}_{\nu_P}[f]$ is at most $\mathbb{E}_{\nu_N}\left[(L^{\leqslant d})^2\right]^{1/2} \leqslant \sqrt{2}$ for any degree $\leqslant d$ polynomial $f$ such that $\mathbb{E}_{\nu_N}[f^2] = 1$. □

**Remark 4.5** (The Importance of Scaling $c$)**.** The planted distribution outputs a feasible system of polynomial equations $\{\langle G_i, x^{\otimes D}\rangle = b_i\}_{i\in[m]}$, where the satisfying assignment is $x = \frac{z}{c^{1/D}}$. Note that $x$ has large norm: $\|x\|_2 = \frac{1}{c^{1/D}}$. We note that our proof of indistinguishability requires that the scaling $c$ be appropriately small. This is necessary. In particular, there is an efficient distinguisher if $c \gg \sqrt{n/m}$. Given input $(G,b)$, calculate the tensor

$$Q := \sum_{i=1}^m G_i \cdot \mathrm{sgn}(b_i)\,.$$

For the null distribution $\nu_N$, $Q$ is distributed as $\sqrt{m}H$ where $H \in (\mathbb{R}^n)^{\otimes D}$ is a tensor with i.i.d. standard Gaussian entries. On the other hand, for the planted distribution, $Q = c\left(\sum_{i=1}^m |b_i|\right) z^{\otimes D} + \mathcal{L} \cdot \sqrt{m}H$, where $\mathcal{L}$ is a linear operator of norm 1 operating on the flattened vector of $H$ for $\|z\|_2 = 1$.

In the case of $D = 2$, $\|\sqrt{m}H\| = O(\sqrt{mn})$, whereas $\|c\left(\sum_{i=1}^m |b_i|\right) zz^\top\| = \Omega(cm)$ with high probability. Thus, if $c \gg \sqrt{n/m}$, then the algorithm that computes the spectral norm of $Q$ is a distinguisher for $\nu_N$ and $\nu_P$. We note that there's an analogous distinguisher based on spectral relaxations of tensor norm for $D > 2$.

For clarity of exposition, we will first prove Lemma 4.4 for the special case of $D = 2$ then show that the ideas generalize to the case of $D > 2$.

## 4.1 Computing Hermite coefficients of $L^{\leqslant d}$ for $D = 2$

The Hermite coefficients of the truncated likelihood $L^{\leqslant d}$ are naturally characterized if we attach a certain combinatorial interpretation to each Hermite index. Towards this goal, let's associate every index $\alpha \in \mathbb{N}^{m\times n\times n}$ with a *labeled directed multigraph* (with self-loops allowed) with $n$ vertices and $|\alpha|$ edges with labels from $[m]$.

**Notations.** From here on, we will use $s$ to denote an index in $[m]$, and $i, j$ to denote indices in $[n]$. For each $s \in [m]$, $\alpha^s \in \mathbb{N}^{n\times n}$ corresponds to the adjacency matrix of the subgraph whose edges have label $s$, hence $|\alpha^s|$ is the number of edges labeled $s$. Furthermore, define $\Delta \in \mathbb{N}^n$ such that $\Delta_i := \sum_{s=1}^m \sum_{j=1}^n \alpha^s_{ij} + \alpha^s_{ji}$ for $i \in [n]$, interpreted as the total degree of vertex $i$. Note that $\alpha$ can have self-loops and each self-loop contributes an additive 2 to the definition of $\Delta$.

**Lemma 4.6** (Hermite Coefficients of $L$)**.** *Let $\alpha \in \mathbb{N}^{m\times n\times n}$, $\beta \in \mathbb{N}^m$. Let $\Delta = \Delta(\alpha) \in \mathbb{N}^n$ such that $\Delta_i$ is the total degree of vertex $i$ in the labeled directed graph associated with $\alpha$. Then, if 1) $\Delta_i$ is even for all $i \in [n]$, 2) $\beta_s \leqslant |\alpha^s|$, and 3) $|\alpha^s| + \beta_s \equiv 0 \pmod 2$ for all $s \in [m]$, then*

$$\mathbb{E}_{(G,b)\sim\nu_P}\left[h_\alpha(G)h_\beta(b)\right] = n^{-|\alpha|}\prod_{s=1}^m \xi_{|\alpha^s|,\beta_s}(c),$$

$$\text{where } \xi_{k,\ell}(c) := \mathbb{E}_{g\sim\mathcal{N}(0,1)}\left[h_k(cg)h_\ell(g)\right] = c^\ell \cdot \frac{k!}{(\frac{k-\ell}{2})!}\left(-\frac{1-c^2}{2}\right)^{\frac{k-\ell}{2}}.$$

*Otherwise, $\mathbb{E}_{(G,b)\sim\nu_P}\left[h_\alpha(G)h_\beta(b)\right] = 0$.*

To prove Lemma 4.6, we first look at the term $\xi_{k,\ell}(c)$:

**Lemma 4.7.** *For any $k, \ell \in \mathbb{N}$ and $c \in [0, 1]$,*

$$\xi_{k,\ell}(c) := \mathbb{E}_{g\sim\mathcal{N}(0,1)}\left[h_k(cg)h_\ell(g)\right] = c^\ell \cdot \frac{k!}{(\frac{k-\ell}{2})!}\left(-\frac{1-c^2}{2}\right)^{\frac{k-\ell}{2}}$$

*if $\ell \leqslant k$ and $k + \ell \equiv 0 \pmod 2$. Otherwise, $\xi_{k,\ell}(c) = 0$.*

*Proof.* First, let us write the function $h_k(cx)$ in the Hermite basis,

$$h_k(cx) = \sum_{\ell=0}^{\infty} \xi_{k,\ell}(c) \frac{h_\ell(x)}{\ell!},$$

such that the coefficients $\xi_{k,\ell}(c)$ exactly equals $\mathbb{E}_{g\sim\mathcal{N}(0,1)}[h_k(cg)h_\ell(g)]$.

Using the generating function of Hermite polynomials (Fact 2.4), for any $x, t \in \mathbb{R}$,

$$e^{cxt-\frac{t^2}{2}} = \sum_{k=0}^{\infty} h_k(cx)\frac{t^k}{k!} = \sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty} \xi_{k,\ell}(c)\frac{t^k}{k!}\cdot\frac{h_\ell(x)}{\ell!}.$$

On the other hand, we can rewrite the left-hand side:

$$\begin{aligned} e^{cxt-\frac{t^2}{2}} = e^{x\cdot ct-\frac{c^2t^2}{2}}\cdot e^{-\frac{t^2}{2}(1-c^2)} &= \sum_{\ell=0}^{\infty} h_\ell(x)\frac{(ct)^\ell}{\ell!}\cdot\sum_{i=0}^{\infty}\frac{1}{i!}\left(-\frac{t^2(1-c^2)}{2}\right)^i \\ &= \sum_{\ell=0}^{\infty}\frac{h_\ell(x)}{\ell!}\sum_{i=0}^{\infty}\frac{c^\ell}{i!}\left(-\frac{1-c^2}{2}\right)^i t^{\ell+2i}. \end{aligned}$$

Matching coefficients, we see that $\xi_{k,\ell}(c)$ is nonzero only if $k = \ell + 2i$ for some $i \geqslant 0$, i.e. $\ell \leqslant k$ and $k + \ell \equiv 0 \pmod 2$. In this case,

$$\xi_{k,\ell}(c) = c^\ell \cdot \frac{k!}{i!}\left(-\frac{1-c^2}{2}\right)^i,$$

where $i = \frac{k-\ell}{2}$. This completes the proof. ∎

We will rely on the following technical computation from [GJJ+20]:

**Lemma 4.8** ([GJJ+20, Lemma 4.5]). *Let $\alpha \in \mathbb{N}^N$, and fix $v \in \mathbb{R}^N$ and $b \in \mathbb{R}$ such that $\|v\|_2 = 1$. Suppose $g \in \mathbb{R}^N$ is sampled from $\mathcal{N}(0, \mathbb{I})$ conditioned on $\langle g, v\rangle = b$, then*

$$\mathbb{E}_g[h_\alpha(g)] = v^\alpha \cdot h_{|\alpha|}(b).$$

We are now ready to prove Lemma 4.6.

*Proof of Lemma 4.6.* In our planted distribution $\nu_P$, each $G_s$ is sampled conditioned on $z^\top G_s z = \langle G_s, zz^\top\rangle = cb_s$. Thus, applying Lemma 4.8 with $v = zz^\top$ (a vector in $\mathbb{R}^{n^2}$),

$$\begin{aligned} \mathbb{E}_{(G,b)\sim\nu_P}\left[h_\alpha(G)h_{\beta_s}(b)\right] &= \mathbb{E}_{z,b}\left[\prod_{s=1}^{m}(zz^\top)^{\alpha^s}h_{|\alpha^s|}(cb_s)h_{\beta_s}(b_s)\right] \\ &= \mathbb{E}_{z\sim\{\pm\frac{1}{\sqrt{n}}\}^n}\left[\prod_{i=1}^{n} z_i^{\Delta_i}\right]\cdot\prod_{s=1}^{m}\mathbb{E}_{b_s\sim\mathcal{N}(0,1)}\left[h_{|\alpha^s|}(cb_s)h_{\beta_s}(b_s)\right], \end{aligned} \tag{8}$$

since $\prod_{s=1}^m (zz^\top)^{\alpha^s} = \prod_{i=1}^n z_i^{\sum_{s,j}\alpha_{ij}^s+\alpha_{ji}^s} = \prod_{i=1}^n z_i^{\Delta_i}$.

Note that $\sum_i \Delta_i = 2|\alpha|$, thus $\mathbb{E}_z\left[\prod_i z_i^{\Delta_i}\right] = n^{-|\alpha|}$ if every $\Delta_i$ is even and 0 otherwise. Moreover, by Lemma 4.7, $\mathbb{E}_{b_s\sim\mathcal{N}(0,1)}\left[h_{|\alpha^s|}(cb_s)h_\beta(b_s)\right] = \xi_{|\alpha^s|,\beta_s}(c)$ if $\beta_s \leqslant |\alpha^s|$ and have the same parity, and 0 otherwise. ∎

## 4.2 Bounding Hermite coefficients of $L^{\leqslant d}$ for $D = 2$

In this section, we prove Lemma 4.4 for the special case of $D = 2$. We first give a sketch.

**Proof sketch.** To begin, we divide the $\alpha$s based on $|\alpha|$ (number of edges $e$) and write the summation as

$$\sum_{e=1}^{d} \sum_{\alpha:|\alpha|=e} \sum_{\beta:|\beta|\leqslant d-e} \mathbb{E}_{\nu_p} \left[h_\alpha(G) h_\beta(b)\right]^2.$$

We upper bound the above in the following steps. First, we show that for a fixed $\alpha$, the innermost sum is dominated by the $\beta$s where $\beta_s = 0$ or 1 if $|\alpha^s|$ is even of odd, respectively (Lemma 4.9). Moreover, any odd $|\alpha^s|$ introduces an extra factor of $c^2$ (Corollary 4.10). Thus, in the end the terms where $|\alpha^s|$ are all even dominate if $c$ is appropriately small.

Next, recall that $\Delta_i$ must be even by the condition in Lemma 4.6. We show that for all $|\alpha| = e$, the dominating terms are the $\alpha$s with $\Delta_i = 2$ and $|\alpha^s| = 2$ for all nonzero $\Delta_i$ and $|\alpha^s|$ (Lemma 4.11, Lemma 4.12). Viewing $\alpha$ as a graph, the dominating terms are the graphs with $e$ edges and $e$ vertices such that each vertex has degree 2 and each edge label appears exactly twice.

For the sake of a clean sketch, let's ignore all other terms. The number of 2-regular graphs with $e$ edges is $\leqslant 2^{2e}$, and there are $n^e$ ways to label the vertices. For edge labels, we choose $e/2$ labels from $[m]$ and assign to the $e$ edges, thus there are $m^{e/2} e^{e/2}$ ways to do so. Finally, we multiply by $n^{-2e}$ (the coefficient in Lemma 4.6) and summing from $e = 2$ to $d$, we get

$$\sum_{e\geqslant 2,\text{ even}} O\left(\frac{me}{n^2}\right)^{e/2} \leqslant \sum_{e\geqslant 2,\text{ even}} \left(\frac{e}{2d}\right)^{e/2} \leqslant 1,$$

when $m = O(\frac{n^2}{d})$. This completes the sketch.

**Contributions from $\beta$ for fixed $\alpha$.** Suppose we fix an $\alpha$ with $|\alpha| = e$. Note that we must have $\beta_s \leqslant |\alpha^s|$ due to the condition in Lemma 4.6. Thus,

$$\sum_{\beta:|\beta|\leqslant d-e} \mathbb{E}_{\nu_p} \left[h_\alpha(G) h_\beta(\boldsymbol{b})\right]^2 = n^{-2e} \sum_{\beta:|\beta|\leqslant d-e} \prod_{s=1}^{m} \xi_{|\alpha^s|,\beta_s}(c)^2 \leqslant n^{-2e} \prod_{s=1}^{m} \sum_{\beta_s\leqslant|\alpha^s|} \xi_{|\alpha^s|,\beta_s}(c)^2. \tag{9}$$

Next, we show that the dominating term is when $\beta_s = 0$ or 1 for all $s \in [m]$ (depending on the parity of $|\alpha^s|$).

**Lemma 4.9.** *For any $k > 0$ and $c = o(\frac{1}{\sqrt{k}})$,*

$$\sum_{\ell\leqslant k} \xi_{k,\ell}(c)^2 \leqslant (1 + o(1)) \cdot \begin{cases} ((k-1)!!)^2 & \text{if } k \text{ is even,} \\ c^2 (k!!)^2 & \text{if } k \text{ is odd.} \end{cases}$$

*Proof.* Using Lemma 4.7,

$$\sum_{\ell\leqslant k} \xi_{k,\ell}(c)^2 \leqslant \sum_{\substack{\ell\leqslant k \\ \ell+k\equiv 0 \bmod 2}} c^{2\ell} \left(\frac{k!}{(\frac{k-\ell}{2})!}\right)^2 \left(\frac{1-c^2}{2}\right)^{k-\ell}.$$

Let $a_\ell$ be the summand. We have that $\frac{a_{\ell+2}}{a_\ell} = c^4(\frac{k-\ell}{2})^2(\frac{2}{1-c^2})^2$, which is $o(1)$ if $c = o(\frac{1}{\sqrt{k}})$. Thus, the term with the smallest $\ell$ in the summation dominates. If $k$ is even, then $\ell = 0$ dominates,

$$\sum_{\ell \leqslant k} \xi_{k,\ell}(c)^2 \leqslant (1+o(1)) \cdot \left(\frac{k!}{(\frac{k}{2})!}\right)^2 2^{-k} = (1+o(1)) \cdot ((k-1)!!)^2.$$

If $k$ is odd, then $\ell = 1$ dominates,

$$\sum_{\ell \leqslant k} \xi_{k,\ell}(c)^2 \leqslant (1+o(1)) \cdot c^2 \left(\frac{k!}{(\frac{k-1}{2})!}\right)^2 2^{-(k-1)} = (1+o(1)) \cdot c^2 (k!!)^2. \qquad \square$$

As an immediate corollary, we can upper bound (9) based on the parity of $|\alpha^s|$:

**Corollary 4.10.** *Fix an $\alpha \in \mathbb{N}^{m\times n\times n}$ with $|\alpha| = e \leqslant d$. Let $\mathrm{odd}(\alpha) = \{s \in [m] : |\alpha^s| \text{ odd}\}$, and $\mathrm{even}(\alpha) = \{s \in [m] : |\alpha^s| > 0, \text{ even}\}$. Then,*

$$\sum_{\beta:|\beta|\leqslant d-e} \mathbb{E}_{\nu_p}\left[h_\alpha(G)h_\beta(b)\right]^2 \leqslant n^{-2e} \cdot \prod_{s\in \mathrm{odd}(\alpha)} c^2(|\alpha^s|!!)^2 \prod_{s\in\mathrm{even}(\alpha)} ((|\alpha^s|-1)!!)^2 .$$

**Contributions from $|\alpha| = e$.** Fix the number of edges $e \leqslant d$, we upper bound the contribution of all $\alpha$ with $|\alpha| = e$. The nonzero condition of Lemma 4.6 means that the only nonzero terms are the $\alpha$s (viewed as graphs) where each vertex has even degree (counting each self-loop twice). To upper bound the total contribution of all such graphs, we

1. upper bound the number of graphs with even degrees where the vertices have labels in $[n]$,
2. upper bound the contributions from assigning labels in $[m]$ to the edges.

Note that the contribution of each $\alpha$ can vary based on how we label the edges.

**Lemma 4.11.** *Let $e \in \mathbb{N}$ such that $0 < e \leqslant d \leqslant n$. Consider directed graphs with $e$ unlabeled edges (parallel edges and self-loops allowed) such that the vertices have even degrees and have distinct labels in $[n]$. The number of such graphs is upper bounded by $(8n)^e$.*

*Proof.* We first count the number of unlabeled graphs. Let $\mathcal{G}(e,v)$ be the set of unlabeled undirected graphs with $e$ edges, $v$ vertices, and has even degrees. We will prove an upper bound on undirected graphs. For directed graphs, we can simply multiply our upper bound by $2^e$, since each edge can be in either direction.

First, we look at the case when $v = e$. In this case, all vertices must have degree 2, hence the graphs must consist of disjoint cycles and isolated vertices with self-loops. This is easily upper bounded by the number of ways to partition $e$ identical elements. The number of ways to partition $e$ identical elements into $j$ non-empty groups is $\binom{e-1}{j-1}$. Thus,

$$|\mathcal{G}(e,e)| \leqslant \sum_{j=1}^{e} \binom{e-1}{j-1} = 2^{e-1} \leqslant 2^e.$$

For $v < e$, observe that every graph in $\mathcal{G}(e,v)$ can be obtained by contracting $e-v$ vertices from a graph in $\mathcal{G}(e,e)$ without deleting any edge (possibly forming self-loops or parallel edges).

The number of ways to do so can be upper bounded the number of ways to partition $e$ distinct items into $v$ non-empty identical buckets, which we can bound by

$$\binom{e-1}{e-v} \cdot \frac{e!}{v!} \leqslant e^{e-v}\binom{e}{v}.$$

Thus,

$$|\mathcal{G}(e,v)| \leqslant e^{e-v}\binom{e}{v} \cdot |\mathcal{G}(e,e)| \leqslant 2^e e^{e-v}\binom{e}{v}.$$

For directed graphs, we multiply the upper bound by an additional $2^e$.

Next, we assign labels to the vertices. For graphs with $v$ vertices, there are $n^v$ ways to assign labels.

$$\sum_{v=1}^{e} n^v \cdot 2^{2e} e^{e-v}\binom{e}{v} = (4e)^e \sum_{v=1}^{e}\binom{e}{v}\left(\frac{n}{e}\right)^v \leqslant (4e)^e\left(1+\frac{n}{e}\right)^e \leqslant (8n)^e.$$

Here we use the fact $e \leqslant d \leqslant n$. □

Next, fix a graph $H$, we assign labels to the edges.

**Lemma 4.12.** *Let $e \in \mathbb{N}$ such that $0 < e \leqslant d \leqslant n$ and let $c = o\left(\frac{1}{d\sqrt{m}}\right)$. Let $H$ be any directed graph with $e$ (unlabeled) edges and $n$ vertices, and let $A_H$ be the set of $\alpha$s that have $H$ as its graph, i.e. $\sum_s \alpha^s$ is the adjacency matrix of $H$. Then,*

$$\sum_{\alpha\in A_H}\sum_{\beta}\mathbb{E}_{\nu_p}\left[h_\alpha(G)h_\beta(b)\right]^2 \leqslant \begin{cases} n^{-2e}(2me)^{e/2} & e \text{ is even} \\ (c^2 e)n^{-2e}(2me)^{\frac{e+1}{2}} & e \text{ is odd.} \end{cases} \tag{10}$$

*Proof.* We need to handle the odd and even $|\alpha^s|$ differently. Recall that $\text{odd}(\alpha) = \{s \in [m] : |\alpha^s| \text{ odd}\}$ and $\text{even}(\alpha) = \{s \in [m] : |\alpha^s| > 0, \text{ even}\}$. Suppose we assign labels from $[m]$ to edges such that $|\text{even}(\alpha)| = i$ and $|\text{odd}(\alpha)| = j$ where $2i + j \leqslant e$. Note that $e$ and $j$ must have the same parity since $e - j$ must be even.

We first fix $i \leqslant \lfloor e/2 \rfloor$. We will see that $j = 0$ or $1$ is the dominating term, depending on the parity of $e$. We upper bound the contribution as follows:

1. Choose $i$ different labels for $\text{even}(\alpha)$ and $j$ labels for $\text{odd}(\alpha)$. The number of ways to choose is $\binom{m}{i}\binom{m-i}{j}$.
2. Choose the $|\alpha^s|$. First, set the default values: $|\alpha^s| = 2$ for $s \in \text{even}(\alpha)$ and $|\alpha^s| = 1$ for $s \in \text{odd}(\alpha)$. Next, for the other $e - 2i - j$, we can add any even number to any $|\alpha^s|$. This is the same as the number of ways $i + j$ nonnegative integers add up to $\frac{e-2i-j}{2}$, which is
$$\binom{\frac{e-2i-j}{2} + (i+j) - 1}{i+j-1} = \binom{\frac{e+j}{2} - 1}{i+j-1}.$$
3. Assign all $e$ edges: $e!$. Note that each $s$ is double counted $|\alpha^s|!$ times.
4. For each $s \in [m]$, the contribution is scaled by a factor given by Corollary 4.10. In this step, we also adjust the contribution due to the double counting in the previous step.
   - $|\alpha^s|$ even: $\frac{((|\alpha^s|-1)!!)^2}{|\alpha^s|!} \leqslant 1$.

- $|\alpha^s|$ odd: $c^2\frac{(|\alpha^s|!!)^2}{|\alpha^s|!} \leqslant c^2|\alpha^s| \leqslant c^2 e$.

Thus, the contribution is scaled by $n^{-2e}(c^2e)^j$.

For a fixed $i$, the total contribution is

$$\sum_{\substack{j\leqslant e-2i \\ e-j \text{ even}}} n^{-2e}c^{2j}e^j \cdot \binom{m}{i}\binom{m-i}{j} \cdot \binom{\frac{e+j}{2}-1}{i+j-1} \cdot e!\,. \tag{11}$$

Let $a_j$ be the summand. $\frac{a_{j+2}}{a_j} \leqslant m^2 \cdot c^4e^2 \cdot (\frac{e+j}{2})(\frac{e-2i-j}{2}) \leqslant (c^2e^2m)^2 = o(1)$ since $c = o\left(\frac{1}{d\sqrt{m}}\right)$. Thus, the summation is dominated by $j = 0$ and 1 for even and odd $e$ respectively.

If $e$ is even, then $j = 0$ dominates: (11) equals $(1+o(1))n^{-2e}e! \cdot \binom{m}{i}\binom{\frac{e}{2}-1}{\frac{e}{2}-i}$. Summing $i$ from 1 to $e/2$, we get

$$(1+o(1))n^{-2e}e! \cdot \sum_{i=1}^{e/2}\binom{m}{i}\binom{\frac{e}{2}-1}{\frac{e}{2}-i} = (1+o(1))n^{-2e}e! \cdot \binom{m+\frac{e}{2}-1}{\frac{e}{2}}.$$

Since $m \geqslant n$ and $\frac{e}{2} \leqslant \frac{n}{2} \leqslant \frac{m}{2}$, we can upper bound the above by $n^{-2e}e!\frac{(2m)^{e/2}}{(e/2)!}$. Thus,

$$\sum_{\alpha\in A_H}\sum_{\beta}\mathbb{E}_{\nu_P}\left[h_\alpha(G)h_\beta(b)\right]^2 \leqslant n^{-2e}(2m)^{e/2}\frac{e!}{(e/2)!} \leqslant n^{-2e}(2me)^{e/2}.$$

If $e$ is odd, then $j = 1$ dominates: (11) equals $(1+o(1))n^{-2e}e!(c^2e)(m-i) \cdot \binom{m}{i}\binom{e-1}{i}$. In this case, we sum $i$ from 0 to $\frac{e-1}{2}$,

$$(1+o(1))n^{-2e}e!(c^2e) \cdot \sum_{i=0}^{\frac{e-1}{2}}(m-i)\binom{m}{i}\binom{\frac{e-1}{2}}{\frac{e-1}{2}-i} \leqslant (1+o(1))n^{-2e}e!(c^2em) \cdot \binom{m+\frac{e-1}{2}}{\frac{e-1}{2}}.$$

Similar analysis shows that

$$\sum_{\alpha\in A_H}\sum_{\beta}\mathbb{E}_{\nu_P}\left[h_\alpha(G)h_\beta(b)\right]^2 \leqslant (c^2e)n^{-2e}(2me)^{\frac{e+1}{2}}.$$

This completes the proof. □

**Proof of Lemma 4.4.** Now, it suffices to sum up the contributions of $e$ from 1 to $d$.

*Proof of Lemma 4.4.* Combining Lemma 4.11 and Lemma 4.12, in total we have,

$$\begin{aligned}&\sum_{e>0,\text{ even}}(8n)^e \cdot n^{-2e}(2me)^{e/2} + \sum_{e\text{ odd}}(8n)^e \cdot (c^2e)n^{-2e}(2me)^{\frac{e+1}{2}} \\ &= \sum_{e>0,\text{ even}}\left(\frac{128me}{n^2}\right)^{e/2} + O(c^2en)\sum_{e\text{ odd}}\left(\frac{128me}{n^2}\right)^{\frac{e+1}{2}}.\end{aligned}$$

Then, setting $c = o\left(\frac{1}{d\sqrt{m}}\right)$, we can ignore the odd terms. Moreover, take $m = \frac{n^2}{256d}$, we have

$$\sum_{1\leqslant|\alpha|+|\beta|\leqslant d}\mathbb{E}_{\nu_P}\left[h_\alpha(G)h_\beta(b)\right]^2 \leqslant \sum_{e\geqslant 2,\text{ even}}^{d}\left(\frac{e}{2d}\right)^{e/2} \leqslant 1.$$

□

### 4.3 Generalizing to $D > 2$

In this section, we prove Lemma 4.4 for arbitrary $D$. In this case, we have polynomial equations $g_s(x) = \langle G_s, x^{\otimes D} \rangle = b_s$ for $s \in [m]$, where $G_s \in (\mathbb{R}^n)^{\otimes D}$.

For Hermite indices $\alpha \in \mathbb{N}^{m \times n \times \cdots \times n}$ and $\beta \in \mathbb{N}^m$, we calculate $\mathbb{E}_{(G,b)\sim\nu_p}\left[h_\alpha(G)h_\beta(b)\right]$. Here, we view $\alpha$ as a labeled directed $D$-uniform hypergraph with edges labeled $1, \ldots, m$, and define $\Delta \in \mathbb{N}^n$ as the total degree of vertex $i \in [n]$. Note that $|\Delta| = \sum_{i=1}^n \Delta_i = D|\alpha|$. The following lemma is almost identical to Lemma 4.6.

**Lemma 4.13.** *For $D \geqslant 2$, indices $\alpha \in \mathbb{N}^{m \times n \times \cdots \times n}$, $\beta \in \mathbb{N}^m$, and $c > 0$, define $\Delta \in \mathbb{N}^n$ such that $\Delta_i$ is the total degree of vertex $i$ when viewing $\alpha$ as a labeled $D$-uniform hypergraph. Then, if $\Delta_i$ is even for all $i \in [n]$ and $\beta_s \leqslant |\alpha^s|$, $|\alpha^s| + \beta_s \equiv 0 \pmod 2$ for all $s \in [m]$, then*

$$\mathbb{E}_{(G,b)\sim\nu_p}\left[h_\alpha(G)h_\beta(b)\right] = n^{-D|\alpha|/2}\prod_{s=1}^m \xi_{|\alpha^s|,\beta_s}(c).$$

*Otherwise, $\mathbb{E}_{(G,b)\sim\nu_p}\left[h_\alpha(G)h_\beta(b)\right] = 0$.*

*Proof.* Similar to the proof of Lemma 4.6, we apply Lemma 4.8 with $v = z^{\otimes D}$ (a vector in $\mathbb{R}^{n^D}$),

$$\begin{aligned}\mathbb{E}_{(G,b)\sim\nu_p}\left[h_\alpha(G)b^\beta\right] &= \mathbb{E}_{z,b}\left[\prod_{s=1}^m (z^{\otimes D})^{\alpha^s} h_{|\alpha^s|}(cb_s)h_{\beta_s}(b_s)\right] \\ &= n^{-D|\alpha|/2}\prod_{s=1}^m \xi_{|\alpha^s|,\beta_s}(c),\end{aligned}$$

since $|\Delta| = \sum_{i=1}^n \Delta_i = D|\alpha|$. This completes the proof. □

The proof of Lemma 4.4 for arbitrary $D$ is almost identical to the case of $D = 2$, except for counting the number of graphs with even degrees. The following is the generalization of Lemma 4.11.

**Lemma 4.14.** *Let $D, d, e, n \in \mathbb{N}$ such that $D > 2$ and $0 \leqslant e \leqslant d \leqslant \frac{2n}{D}$. Consider directed $D$-uniform hypergraphs with $e$ unlabeled edges (parallel edges and self-loops allowed) such that the vertices have even degrees and have distinct labels in $[n]$. The number of such graphs is upper bounded by*

$$O(Dn)^{\frac{De}{2}} e^{(\frac{D}{2}-1)e},$$

*if $De$ is even. Otherwise, there are no such graphs.*

*Proof.* Note that $De$ is the total degree, which must be even. The number of vertices $v$ can range from 1 to $\frac{De}{2}$. In order to perform the counting, we view a hypergraph $H$ as a bipartite factor graph $(V, F, E)$ with left-hand side vertex set $V$ that contains the vertices of $H$, and right-hand side vertices $F$ that contains a vertex for each hyperedge. Note, in particular, that the right-degree of the bipartite graph is $D$. And all left-degrees have to be even in the hypergraphs we intend to count.

We directly analyze the number of $v$-vertex graphs.

1. We choose $v$ labels from $[n]$, giving us $\binom{n}{v}$.

2. We choose the left-degrees: the degrees must be even and sum to $De$. This is the same as the number of ways $v$ positive integers add up to $\frac{De}{2}$, which is $\binom{\frac{De}{2}-1}{\frac{De}{2}-v}$.

3. We add edges between $V$ and $F$ while ensuring that the degrees are consistent. To do so, we construct a vertex set $V'$ by including $\deg(i)$ copies of vertex $i \in V$, hence $|V'| = De$. Then, we add edges between $F$ and $V'$ such that each $f \in F$ has degree $D$ and each $i \in V'$ has degree 1. Since the factor vertices are unlabeled, there are at most $\frac{(De)!}{e!}$ ways to do so.

Then, combining the above and summing $v$ from 1 to $\frac{De}{2}$,

$$\frac{(De)!}{e!} \sum_{v=1}^{\frac{De}{2}} \binom{n}{v} \binom{\frac{De}{2}-1}{\frac{De}{2}-v} \leqslant \frac{(De)!}{e!} \binom{n+\frac{De}{2}-1}{\frac{De}{2}}.$$

Using the fact that $\frac{De}{2} \leqslant n$ and Stirling's approximation, we can upper bound the above by

$$O(Dn)^{\frac{De}{2}} e^{(\frac{D}{2}-1)e}. \qquad \square$$

The contributions from assigning labels to edges is exactly the same as Lemma 4.12, except with coefficient $n^{-De}$. Thus, we are in position to prove Lemma 4.4.

*Proof of Lemma 4.4.* We sum over all contributions of $|\alpha| = e$ from 1 to $d$. Setting $c = o\left(\frac{1}{d\sqrt{m}}\right)$, we can ignore the odd terms; in fact, if $D$ is odd, the odd terms are exactly zero since $De$ must be even. Thus, the total contribution is

$$\sum_{\substack{\alpha,\beta: \\ 1 \leqslant |\alpha|+|\beta| \leqslant d}} \mathbb{E}_{\nu_P} \left[h_\alpha(G) h_\beta(\boldsymbol{b})\right]^2 \leqslant \sum_{e \text{ even}} O\left(\frac{De}{n}\right)^{\frac{De}{2}} \left(\frac{m}{e}\right)^{\frac{e}{2}} \leqslant 1,$$

when $m \leqslant O_D\left(\frac{n^D}{d^{D-1}}\right)$. This completes the proof. $\square$

# 5 Algorithmic Thresholds at Degree 2

In this section, we give a short proof of the following theorem that gives a sharp threshold on the number of quadratic equations $m$ required for the existence of degree-2 SoS refutations.

**Theorem 5.1.** *For any homogeneous quadratic polynomials $g_1, g_2, \ldots, g_m$ in $x_1, x_2, \ldots, x_n$ and real numbers $b_1, b_2, \ldots, b_m$, let $\mathsf{SOS}_2(\mathcal{P})$ be the degree-2 SoS relaxation of the system of constraints $\{g_i(x) = b_i\}_{i \leqslant m}$. Specifically, let $G_i \in \mathbb{R}^{n \times n}$ be matrices such that $g_i(x) = x^\top G_i x$ for each $i \in [m]$. Then, the degree-2 SoS relaxation is the following SDP:*

$$X \succeq 0, \ \operatorname{tr}(G_i X) = b_i \ \text{ for all } 1 \leqslant i \leqslant m. \tag{12}$$

*Suppose each coefficient of $g_i$ is chosen to be an independent draw from the standard Gaussian distribution $\mathcal{N}(0,1)$. Then, there is an absolute constant $C$ such that if $m \geqslant \frac{n^2}{4} + Cn \log n$, the semidefinite program above is infeasible with probability at least 0.49. On the other hand if $m \leqslant \frac{n^2}{4} - Cn \log n$ then the semidefinite program above is feasible with probability at least $1 - \frac{1}{n}$.*

Our proof is an immediate application of a classical work [ALMT14] on understanding phase transitions for convex programs with random data that relies on deep results from conic integral geometry [SW08]. In particular, our proof relies on the following *approximate kinematic formula*.

**Fact 5.2** ([ALMT14, Theorem I]). *Fix a tolerance $\eta \in (0,1)$. Let $C$ and $K$ be convex cones in $\mathbb{R}^N$, and let $Q \in \mathbb{R}^{N\times N}$ be a uniformly random (i.e. Haar distributed) orthogonal matrix. Then,*

$$\delta(C)+\delta(K) \leqslant N - O(\sqrt{N}\log(1/\eta)) \implies \mathop{\mathbf{Pr}}_{Q}[C \cap QK \neq \{0\}] \leqslant \eta;$$
$$\delta(C)+\delta(K) \geqslant N + O(\sqrt{N}\log(1/\eta)) \implies \mathop{\mathbf{Pr}}_{Q}[C \cap QK \neq \{0\}] \geqslant 1-\eta.$$

*Here, $QK = \{Qz \mid z \in K\}$ is the rotation of the cone $K$ by $Q$ and $\delta(C), \delta(K)$ are* statistical dimensions *of the cones $C, K$ respectively.*

We will not define statistical dimension formally in this work but note that the statistical dimension of a subspace of dimension $r$ is $r$ and that of the cone of positive semidefinite $n \times n$ matrices is $\frac{1}{4}n(n+1)$ (see Table 3.1 of [ALMT14]). For background and proofs, we refer the reader to [ALMT14].

*Proof of Theorem 5.1.* Let $S_+$ be the open convex cone of positive definite matrices. Let $K$ be the linear span of the symmetric matrices $G_1, \ldots, G_m$ viewed as $\frac{n(n+1)}{2}$ dimensional vectors. Let $K^\perp$ be the orthogonal complement of $K$ in $\mathbb{R}^{n^2}$.

Since $0 < m < \frac{n(n+1)}{2}$, $K$ and $K^\perp$ have dimension $m$ and $\frac{n(n+1)}{2} - m$ with probability 1 over the draw of the $G_i$s. Thus, the statistical dimension of $K$, $K^\perp$ is $m$ and $\frac{n(n+1)}{2} - m$ respectively. Observe that because the coefficients of $g_i$s are independent standard Gaussians, $G_i$s are standard Gaussian vectors, and $K, K^\perp$ are random (rotations of) subspaces of their dimension. The statistical dimension of $S_+$ is $\frac{1}{4}n(n+1)$.

Applying Fact 5.2 to $K$ and $K^\perp$ with $\eta = \frac{1}{n}$ yields that there is a constant $C > 0$ such that:

1. **Case 1:** If $m \geqslant \frac{n^2}{4} + Cn\log n$, then, there is a positive definite matrix $M_1$ in $K$.
2. **Case 2:** If $m \leqslant \frac{n^2}{4} - Cn\log n$, with probability at least $1 - 1/n$, there is a positive definite matrix $M_2$ in $K^\perp$.

Let's now condition on the existence of $M_1/M_2$ in the two cases and analyze the SDP (12).

**Case 1:** Suppose for the sake of contradiction that there is a PSD $Y$ such that $\langle G_i, Y\rangle = b_i$ for every $i \in [m]$. Let $M_1 = \sum_i c_i G_i \in K$ for $c_i \in \mathbb{R}$. Then, $\langle M_1, Y\rangle = \sum_i c_i b_i$. Now, the LHS is non-negative since $M_1, Y$ are both positive semidefinite. The RHS $\sum_i c_i b_i$, on the other hand, is distributed as a standard scalar Gaussian and is thus $< 0$ with probability $1/2$. Thus, there can be no such $Y$ with probability at least $1/2$.

**Case 2:** Let $M_2$ be the positive definite matrix such that $\langle M_2, G_i\rangle = 0$ for every $i \in [m]$. Let $Y \in \mathbb{R}^{n\times n}$ be any solution to $\langle G_i, Y\rangle = b_i$ for every $i$. Such a $Y$ exists since $G_i$s are linearly independent with probability 1. Then, observe that for some large enough scaling $R$, $RM_2 + Y$ is positive semidefinite and is feasible for the SDP (12).

This completes the proof. ∎

# 6 Sum-of-Squares Lower Bounds at Degree 4

In this section, we show that there is an $m = n^2/\operatorname{poly}(\log n)$ such that for random homogeneous quadratic polynomials $g_1, g_2, \ldots, g_m$ of degree 2, the constraint system $\{g_i(x) = 0\}_{i \leqslant m}$ does not admit a degree-4 sum-of-squares refutation. Specifically, we will establish the following dual version of such a claim:

**Theorem 6.1.** *Fix $m = m(n) \leqslant n^2/\operatorname{poly}(\log n)$. Let $g_1, g_2, \ldots, g_m$ be homogeneous degree-2 polynomials in $x_1, x_2, \ldots, x_n$ such that each coefficient of each $g_i$ is an independent draw of the standard Gaussian distribution $\mathcal{N}(0,1)$. Then, with probability $1 - o(1)$, there exists a degree-4 pseudo-distribution $\mu$ on $x_1, x_2, \ldots, x_n$ consistent with the constraint system $\{g_i(x) = 0\}_{i \leqslant m}$.*

We will prove Theorem 6.1 by giving an explicit construction of a pseudo-distribution $\mu$ satisfying the requirements of the theorem. Our construction of $\mu$ will rely on the standard technique of pseudo-calibration and will use the planted distribution constructed in the previous section. Our analysis adapts the high-level analysis strategy invented in [GJJ+20] who proved a sum-of-squares lower bound for optimizing the Sherrington-Kirkpatrick Hamiltonian. The details of this strategy in our setting are somewhat different.

**Candidate pseudo-distribution.** We construct a candidate pseudo-distribution $\mu$ based on the pseudo-calibration method, using the planted distribution $\nu_P$ in Definition 4.3 (with $D = 2$ and $c = 0$). In a nutshell, the pseudo-calibration method is a mechanical way to construct each entry of the candidate pseudo-moment matrix based on $\nu_P$.

**Definition 6.2** (Candidate pseudo-distribution)**.** *Fix $m = m(n) \leqslant n^2/\operatorname{poly}(\log n)$ and truncation threshold $\tau = \operatorname{poly}(\log n)$. Given $G$ sampled from the null distribution $\nu_N$, we define the pseudo-distribution $\mu$ over $\{\pm\frac{1}{\sqrt{n}}\}^n$ (as a function of $G$) by describing the pseudo-expectation of all degree $\leqslant 4$ monomials: for $I \subseteq [n]$ and $|I| \leqslant 4$,*

$$\widetilde{\mathbb{E}}_\mu[x^I] := \sum_{\substack{\alpha \in \mathbb{N}^{m \times n \times n} \\ |\alpha| \leqslant \tau}} \mathbb{E}_{(G',z)\sim\nu_P}\left[z^I h_\alpha(G')\right] \cdot \frac{h_\alpha(G)}{\alpha!}\,.$$

Note that we have the "normalized" booleanity constraint $x_i^2 = \frac{1}{n}$. Our final construction that yields Theorem 6.1 will be obtained by a small perturbation of the construction in Definition 6.2.

To analyze this construction, it is helpful to study a matrix – the moment matrix – associated with the pseudo-distribution.

**The Moment Matrix.** The moment matrix $\mathcal{M}$ of $\mu$ is a matrix indexed by subsets $I, J \subseteq [n]$ of size $\leqslant 2$ and entries defined by:

$$\mathcal{M}(I,J) := \widetilde{\mathbb{E}}_\mu[x^{I+J}] = \sum_{\substack{\alpha \in \mathbb{N}^{m \times n \times n} \\ |\alpha| \leqslant \tau}} \mathbb{E}_{(G',z)\sim\nu_P}\left[z^{I+J} h_\alpha(G')\right] \cdot \frac{h_\alpha(G)}{\alpha!}\,.$$

We can explicitly compute the coefficient of the Hermite polynomial $h_\alpha(G)$ in the above expression for $\mathcal{M}(I,J)$ as follows. Again, we will use $s$ to denote an index in $[m]$ and $i, j$ to denote

indices in $[n]$. By the computation we did in the context of our low-degree lower bounds, specifically Lemma 4.6 (setting $\beta = 0$ and $c = 0$), we obtain that for any $I, J \subseteq [n]$ and any $\alpha \in \mathbb{N}^{m\times n\times n}$,

$$\lambda_{\alpha,I,J} := \frac{1}{\alpha!}\mathbb{E}_{(G',z)\sim\nu_P}\left[z^{I+J}h_\alpha(G')\right] = (-1)^{|\alpha|/2}n^{-|\alpha|-\frac{|I|+|J|}{2}}\prod_{s=1}^{m}(|\alpha^s|-1)!!\cdot\frac{1}{\alpha!} \tag{13}$$

if $|\alpha^s|$ is even for all $s \in [m]$ and $\Delta_i + I_i + J_i$ is even for all $i \in [n]$ (here we denote $I_i := \mathbf{1}\{i \in I\}$), and 0 otherwise (recall that $\Delta \in \mathbb{N}^n$ where $\Delta_i := \sum_{s=1}^m \sum_{j=1}^n \alpha_{ij}^s + \alpha_{ji}^s$, interpreted as the total degree of vertex $i$). Thus, we have

$$\mathcal{M}(I,J) := \sum_{\substack{\alpha:|\alpha|\leqslant\tau \\ |\alpha^s| \text{ even}, \Delta_i+I_i+J_i \text{ even}}} \lambda_{\alpha,I,J}h_\alpha(G).$$

Note the $1/\alpha!$ factor in (13) is there because we use the unnormalized Hermite polynomials. By an upper bound on the double factorial (Fact 2.3),

$$|\lambda_{\alpha,I,J}| \leqslant n^{-|\alpha|-\frac{|I|+|J|}{2}}\left(\frac{|\alpha|}{2}\right)^{|\alpha|/2}. \tag{14}$$

Keep in mind that $\mathcal{M}$ will only approximately satisfy the conditions of a pseudo-moment, e.g. $\mathcal{M}(\varnothing,\varnothing) \approx 1$ and $\mathcal{M}(\{i\},\{i\}) \approx \frac{1}{n}$. However, we will show that we can "fix" the moment matrix such that it represents a valid pseudo-distribution and satisfies all constraints. Note that the positivity property, i.e., $\widetilde{\mathbb{E}}_\mu[q^2] \geqslant 0$ for every degree-2 polynomial $q$ is equivalent to the positive semidefiniteness of the moment matrix $\mathcal{M}$ of $\mu$.

**Lemma 6.3.** *There exist constants $C_1, C_2 > 0$ such that if $m = n^2/\log^{C_1} n$ and $\tau = \log^{C_2} n$, then there exists a correction matrix $\mathcal{E}$ such that $\mathcal{M} - \mathcal{E}$ satisfies all constraints $\{g_s(x) = 0\}_{s\leqslant m}$ and that $\mathcal{M} - \mathcal{E} \succeq 0$.*

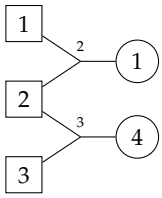
This lemma is the bulk of the proof of Theorem 6.1 and requires a relatively technically involved argument. In order to prove PSDness of $\mathcal{M}$ we need to analyze its spectrum. This is somewhat challenging as the matrix has dependent random entries. Our proof relies on a strategy invented in previous works (starting with [BHK+16] and built further in [HKP+17, GJJ+20]) that decomposes moment matrices built via pseudo-calibration into a sum of structured random matrices (called *graph matrices*) that are helpful in spectral analysis. We start with a brief background of graph matrices specialized to our setting before giving an outline of our proof.

### 6.1 Background on graph matrices

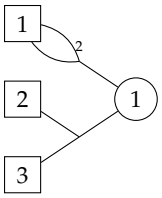
Our notations and definitions follow that of [AMP20, GJJ+20] who also studied with graphical matrices when the input data is random Gaussian.

We represent each Hermite index $\alpha \in \mathbb{N}^{m\times n\times n}$ as a *3-uniform hypergraph* with two types of vertices: circles ◯ and squares □. Each square $\boxed{i}$ has a label $i \in [n]$, and each circle $\textcircled{s}$ has a label $s \in [m]$. A nonzero entry $\alpha_{ij}^s$ is represented by a hyperedge $(\boxed{i}, \boxed{j}, \textcircled{s})$. See Figure 1 for an example. Note that the order of $\boxed{i}$ and $\boxed{j}$ matters since we allow $\alpha_{ij}^s \neq \alpha_{ji}^s$, but for simplicity we don't draw it out explicitly.

Next, we define *ribbons* and *shapes* (see Definitions 2.9–2.12 in [GJJ+20]). Denote $\mathcal{S} := \{\boxed{i} : i \in [n]\}$ and $\mathcal{C} := \{\textcircled{s} : s \in [m]\}$. A ribbon $R$ is simply a hypergraph $(V(R), E(R))$ of some $\alpha$ (as in




(a) $\alpha^1_{12} = 2$ and $\alpha^4_{23} = 3$.




(b) $\alpha^1_{11} = 2$ and $\alpha^1_{23} = 1$.

Figure 1: Examples of $\alpha \in \mathbb{N}^{m \times n \times n}$ represented as hypergraphs.

Figure 1) with a set of "left" and "right" vertices $A_R, B_R \subseteq V(R)$. Each ribbon defines a matrix with a single entry.

**Definition 6.4** (Ribbons)**.** *A ribbon is a 3-uniform hypergraph $R = (V(R), E(R), A_R, B_R)$ such that $V(R) \subseteq \mathcal{S} \cup \mathcal{C}$ contains labeled square and circle vertices, and $A_R, B_R \subseteq V(R)$ (not necessarily disjoint). The edges in $E(R)$ are labeled and must be connected to two square vertices and one circle vertex.*

**Definition 6.5** (Matrix of a ribbon)**.** *Let a ribbon $R = (V(R), E(R), A_R, B_R)$, and let $\alpha \in \mathbb{N}^{m \times n \times n}$ be the multiset represented by $(V(R), E(R))$. The matrix of a ribbon $M_R$, indexed by subsets of $\mathcal{S} \cup \mathcal{C}$, is defined as*

$$M_R(I, J) = \begin{cases} h_\alpha(G) & I = A_R, J = B_R, \\ 0 & \text{otherwise}. \end{cases}$$

The *shape* is a ribbon with the labels of each vertex removed, i.e. ribbons with the same hypergraph structure but different labels have the same shape.

**Definition 6.6** (Shape)**.** *A shape is a 3-uniform hypergraph $a = (V(a), E(a), U_a, V_a)$ where $V(a)$ contains unlabeled circle and square vertices and $U_a, V_a \subseteq V(a)$ (not necessarily disjoint). The edges in $E(a)$ are labeled and must be connected to two square vertices and one circle vertex.*

*We call $U_a, V_a$ the "left" and "right" vertices. Moreover, define $W_a := V(a) \setminus (U_a \cap V_a)$ to be the "middle" vertices of the shape and $W_{\text{iso}}$ to be the isolated vertices in $W_a$.*

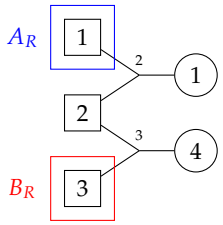
**Definition 6.7** (Graph matrix)**.** *The matrix of a shape $M_a$ is defined as*

$$M_a := \sum_{R\text{: ribbon of shape } a} M_R.$$

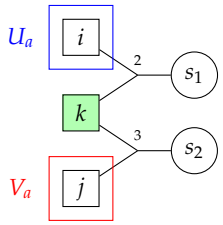
Ribbons and shapes are best explained by examples. Consider the ribbon $R$ and shape $a$ in Figure 2. The matrix $M_R$ has entries $M_R(I, J) = h_2(G^1_{12})h_3(G^4_{23})$ if $I = \{1\}$, $J = \{3\}$, and 0 otherwise. The graph matrix $M_a$ is a sum of all ribbons of shape $a$, including $R$. Thus, $M_a(\{i\}, \{j\}) = \sum_{k \in [n], k \neq i,j} \sum_{s_1 \neq s_2 \in [m]} h_2(G^{s_1}_{ik})h_3(G^{s_2}_{kj})$ for $i \neq j$.

**Definition 6.8** (Transpose of a shape)**.** *The transpose of a shape $a = (V(a), E(a), U_a, V_a)$ is defined as $a^\top := (V(a), E(a), V_a, U_a)$. This implies that $M_a = (M_{a^\top})^\top$.*

**Graph matrix norm bounds.** We will require spectral norm bounds of graph matrices. We can directly use the norm bounds from [AMP20], which are obtained using the trace power method. First, define the weights of square and circle vertices: $w(\square) = 1$ and $w(\bigcirc) = \log_n(m)$. This is




(a) Ribbon $R$.




(b) Shape $a$. $S_{\min} = \{\boxed{k}\}$.

Figure 2: Example of a ribbon and shape. The minimum vertex separator of a shape is colored green.

defined such that for any shape $a$ and any subsets $S, C$ of square and circle vertices, $n^{w(S)+w(C)} = n^{|S|}m^{|C|}$, which is roughly the number of ways you can label $S, C$; such quantities naturally arise in trace moment calculations.

Next, we define the *minimum vertex separator*:

**Definition 6.9** (Minimum vertex separator)**.** *For a shape $a$, a set $S \subseteq V(a)$ is a vertex separator if all paths from $U_a$ to $V_a$ pass through $S$. A minimum vertex separator $S_{\min}$ is the smallest weight vertex separator.*

See Figure 2b for example; in our figures the minimum vertex separator is colored green. Note that by definition, $U_a \cap V_a$ must be in the minimum vertex separator. Using the norm bounds from [AMP20, Corollary 8.16] and the same calculations from [GJJ+20, Appendix A]), we have

**Proposition 6.10.** *With probability over $1 - o(1)$, for all shapes $a$ the graph matrix satisfies*

$$\|M_a\| \leqslant (|V(a)| \cdot |E(a)| \cdot \log n)^{O(|V(a)|+|E(a)|)} \cdot n^{\frac{w(V(a))-w(S_{\min})+w(W_{\text{iso}})}{2}} = \widetilde{O}\left(n^{\frac{w(V(a))-w(S_{\min})+w(W_{\text{iso}})}{2}}\right).$$

## 6.2 Proof overview of Lemma 6.3

Since the proof is rather technical, we first provide an overview of the proof and defer the technical details to the Appendix. At a high-level, our strategy resembles that of [GJJ+20] who proved a sum-of-squares lower bound for the problem of certifying the optimum value of the Sherrington-Kirkpatrick Hamiltonian. However, there are important differences to adapt this strategy to our setting as we describe in Remark 6.26.

At a high-level, our strategy works in two steps which are rather common in the analyses of moment matrices arising in several prior works on SoS lower bounds using pseudo-calibration. In the first step, we will prove that the moment matrix $\mathcal{M}$ is positive semidefinite and approximately (but not exactly) satisfies the polynomial constraints. In the second step, we will modify the pseudo-distribution $\mu$ so as to satisfy the constraints exactly and further show that this correction is small and does not affect the analysis of PSDness.

**Decomposition of $\mathcal{M}$.** Observe that the coefficients $\lambda_{\alpha,I,J}$ in (13) only depend on the shapes. Thus, we can write $\mathcal{M}$ as

$$\mathcal{M} = \sum_{a\text{: shape}} \lambda_a M_a.$$

We will first identify combinatorial conditions on the shapes defining the graphical matrices that appear with nonzero coefficients in the above expansion. The shapes with $\lambda_a \neq 0$ need to satisfy the following conditions,

**Definition 6.11.** *Let $\mathcal{L}$ be the set of shapes $a$ such that*

1. *$U_a, V_a$ contain only square vertices and $|U_a|, |V_a| \leqslant 2$,*
2. *$\deg(\boxed{i}) + \mathbf{1}\{\boxed{i} \in U_a\} + \mathbf{1}\{\boxed{i} \in V_a\}$ is even for all $\boxed{i} \in V(a)$,*
3. *$\deg(\textcircled{s})$ is even for all $\textcircled{s} \in V(a)$,*
4. *$|E(a)| \leqslant \tau$,*
5. *There are no isolated vertices in $W_a$.*

In words, Condition 1 is because $\mathcal{M}$ only contains moments of degree $\leqslant 4$; Condition 2 ensures that $\Delta_i + I_i + I_j$ is even; Condition 3 ensures that $|\alpha^s|$ is even; Condition 4 ensures that $|\alpha| \leqslant \tau$; and finally Condition 5 is simply because shapes with isolated vertices don't appear in the decomposition (there can only be isolated vertices in $U_a \cap V_a$).

**Remark 6.12.** For any $a \in \mathcal{L}$, the conditions in Definition 6.11 also imply that $|E(a)|$ is even and the total degree of square vertices is a multiple of 4.

Thus, we can decompose $\mathcal{M}$ into shapes in $\mathcal{L}$:

$$\mathcal{M} = \sum_{a \in \mathcal{L}} \lambda_a M_a.$$

Next, observe that we can break $\mathcal{M}$ into blocks indexed by $(k, \ell) \in \{0, 1, 2\}^2$. The $(k, \ell)$ block $\mathcal{M}_{k\ell}$ is $\binom{n}{k} \times \binom{n}{\ell}$ whose rows are indexed by subsets $\binom{[n]}{k}$ and columns are indexed by subsets $\binom{[n]}{\ell}$. Clearly, shapes $a$ with $|U_a| = k$, $|V_a| = \ell$ contribute to $\mathcal{M}_{k\ell}$ only. Moreover, $|U_a| + |V_a|$ must be even because the total degree of the square vertices must be even (each hyperedge contributes two). Thus, the blocks $\mathcal{M}_{01}$, $\mathcal{M}_{10}$, $\mathcal{M}_{12}$, and $\mathcal{M}_{21}$ are zero, i.e. all odd moments are zero. Thus, $\mathcal{M}$ has the following structure

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{00} & 0 & \mathcal{M}_{02} \\ 0 & \mathcal{M}_{11} & 0 \\ \mathcal{M}_{20} & 0 & \mathcal{M}_{22} \end{bmatrix},$$

where $\mathcal{M}_{00}$ is a scalar, and $\mathcal{M}_{02} = \mathcal{M}_{20}^\top$ is a vector and has the same entries as $\mathcal{M}_{11}$.

We need to show that $\mathcal{M}$, with some small modifications, is positive semidefinite and satisfies all constraints in the quadratic system.

**Proving PSDness.** We parameterize $m = n^{2-\varepsilon}$ for $\varepsilon = \frac{C \log\log n}{\log n}$ for a sufficiently large constant $C$ in the analysis that follows. We know that $\mathcal{M}$ can be expanded as a sum of graphical matrices indexed by shapes $a$ in $\mathcal{L}$ with coefficient $\lambda_a$. We first identify the shapes that contribute scaled identity matrices in the diagonal blocks. We call these shapes the *trivial shapes*; see Figure 3 for examples.


(a) $|\lambda_a| = n^{-1}$, $\|M_a\| = 1$. (b) $|\lambda_a| = n^{-2}$, $\|M_a\| = 1$.

Figure 3: Trivial shapes: $U_a = V_a$, $E(a) = \varnothing$, and $M_a = \mathbb{I}$.

**Definition 6.13** (Trivial shape)**.** *A shape $a$ is trivial if $U_a = V_a$, $W_a = \varnothing$, and $E(a) = \varnothing$. Its associated matrix $M_a = \mathbb{I}$.*

In other words, the trivial shapes correspond to the Hermite indices $\alpha = \vec{0}$ and $|I| = |J|$. For a trivial shape $a_{\mathrm{triv},k}$ with $|U_{a_{\mathrm{triv},k}}| = |V_{a_{\mathrm{triv},k}}| = k$, its matrix $\lambda_{a_{\mathrm{triv},k}} M_{a_{\mathrm{triv},k}} = n^{-k} \cdot \mathbb{I}$ is a component in $\mathcal{M}_{kk}$. Crucially, it is full rank and has minimum singular value $n^{-k}$, hence we can *charge* other shapes that have small norm to the trivial shapes. We call this procedure a *charging scheme*.

**Negligible shapes.** We can charge several shapes to the trivial shapes if the contribution from those shapes are dominated by the scaled identity matrices from the trivial shapes; we call all shapes that can be charged this way *negligible*.

**Definition 6.14.** *We say a shape is negligible if $|E(a)| \neq 0$ and*

$$\|\lambda_a M_a\| \leqslant n^{-\frac{|U_a|+|V_a|}{2}} \cdot n^{-\Omega(\varepsilon|E(a)|)}.$$

Intuitively, for a negligible shape $a$ in block $\mathcal{M}_{kk}$ (meaning $|U_a| = |V_a| = k$), its contribution $\|\lambda_a M_a\| \ll n^{-k}$, which is the minimum singular value of $\lambda_{a_{\mathrm{triv},k}} M_{a_{\mathrm{triv},k}}$.

In Section A.1, we will identify a simple criterion to determine whether a shape is negligible or not (Lemma A.3), then we will prove that $M_{a_{\mathrm{triv},k}}$ dominates all negligible shapes, hence forming a PSD component in $\mathcal{M}$:

**Lemma 6.15.** *For $k = 0, 1, 2$, let $\mathcal{L}_{\mathrm{negl},k}$ be the set of negligible shapes in block $\mathcal{M}_{kk}$, and let $\mathcal{E}_{\mathrm{negl},k} := \sum_{a \in \mathcal{L}_{\mathrm{negl},k}} \lambda_a M_a$. There exist constants $c_1, c_2 > 0$ such that if the threshold $\tau \leqslant n^{c_1 \varepsilon}$, then*

$$\|\mathcal{E}_{\mathrm{negl},k}\| \leqslant n^{-k-c_2\varepsilon}.$$

*This implies that*

$$\lambda_{a_{\mathrm{triv},k}} M_{a_{\mathrm{triv},k}} + \mathcal{E}_{\mathrm{negl},k} \succ 0.$$

Note that in the case $k = 0$, we have $\mathcal{M}_{00} = 1 + o(1)$. This is consistent with the calculations of low-degree hardness in Section 4. Note also that $\mathcal{M}$ must have a non-trivial null space due to the constraints, hence there must be non-negligible shapes in $\mathcal{L}$ which we deal with next.

The same analysis also shows the following norm bounds,

**Lemma 6.16.** *There exists a constant $c_1 > 0$ such that if the threshold $\tau \leqslant n^{c_1 \varepsilon}$, then for any $k, \ell$, $\|\mathcal{M}_{k\ell}\| \leqslant n^{-\frac{k+\ell}{4}}$.*

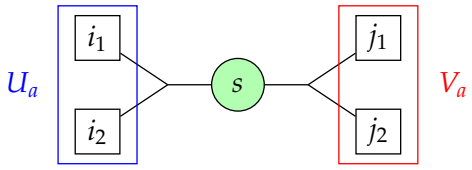



Figure 4: Spider $a_{\text{spider}}$.

**Connected shapes and spider.** We look at the shapes in $\mathcal{M}_{11}$ and $\mathcal{M}_{22}$ that are connected, meaning there is path from $U_a$ to $V_a$ and $S_{\min} \neq \varnothing$. We show in Section A.2 that there is only *one* connected non-trivial shape that is not negligible, namely the *spider*; see Figure 4 for illustration.

**Lemma 6.17.** *If $a \in \mathcal{L}$ is a connected shape and not a trivial shape nor a spider, then $a$ is negligible.*

Next, we handle the spider $a_{a_{\text{spider}}}$. The main insight is that $M_{a_{\text{spider}}}$ is "almost" in the null space of $\mathcal{M}$, i.e. $\mathcal{M}M_{a_{\text{spider}}} \approx 0$. Then, we use the following result (see also [GJJ+20, Fact 3.1]); we give a short proof for completeness.

**Lemma 6.18.** *Suppose a matrix $A$ satisfies $\mathcal{M}A = 0$, then $\mathcal{M} - A \succeq 0$ implies $\mathcal{M} \succeq 0$.*

*Proof.* For any vector $x$, let $y$ be its projection onto the column space of $\mathcal{M}$. We have $y \perp \text{Null}(\mathcal{M})$ and $y^\top Ay = 0$. Then, $x^\top \mathcal{M}x = y^\top \mathcal{M}y = y^\top(\mathcal{M} - A)y$. Thus, $\mathcal{M} - A \succeq 0$ implies $x^\top \mathcal{M}x \geqslant 0$ for all $x$, which means $\mathcal{M} \succeq 0$. □

Intuitively, Lemma 6.18 allows us to add/remove any component of $\mathcal{M}$ which is in the null space of $\mathcal{M}$. In our case, we can thus remove the component $\lambda_{a_{\text{spider}}} M_{a_{\text{spider}}}$ from $\mathcal{M}$ modulo some small error $\mathcal{E}_{\text{spider}}$. More specifically, in Section A.3, we will show the following,

**Lemma 6.19.** *Suppose $\mathcal{M}$ exactly satisfies all constraints $\{g_s(x) = 0\}_{s \leqslant m}$. Then there exists a matrix $A$ such that $\mathcal{M}A = 0$ and*

$$\lambda_{a_{\text{spider}}} M_{a_{\text{spider}}} = A + \mathcal{E}_{00} + \mathcal{E}_{20} + \mathcal{E}_{20}^\top + \mathcal{E}_{22},$$

*where $\mathcal{E}_{00}$, $\mathcal{E}_{20}$, $\mathcal{E}_{20}^\top$, $\mathcal{E}_{22}$ are errors in blocks $\mathcal{M}_{00}$, $\mathcal{M}_{20}$, $\mathcal{M}_{02}$, $\mathcal{M}_{22}$ respectively, and $|\mathcal{E}_{00}| = \widetilde{O}(n^{-3})$, $\|\mathcal{E}_{20}\| = \widetilde{O}(n^{-5/2})$, and $\|\mathcal{E}_{22}\| = \widetilde{O}(n^{-2-\varepsilon})$.*

Thus, we have

$$\begin{aligned} \mathcal{M}' := \mathcal{M} - A &= \mathcal{M} - \lambda_{a_{\text{spider}}} M_{a_{\text{spider}}} + \mathcal{E}_{\text{spider}} \\ &= \begin{bmatrix} \mathcal{M}_{00} + \mathcal{E}_{00} & 0 & \mathcal{M}_{20}^\top + \mathcal{E}_{20}^\top \\ 0 & \mathcal{M}_{11} & 0 \\ \mathcal{M}_{20} + \mathcal{E}_{20} & 0 & \mathcal{M}'_{22} + \mathcal{E}_{22} \end{bmatrix}, \end{aligned} \tag{15}$$

where $\mathcal{M}'_{22}$ is the block $\mathcal{M}_{22}$ with the spider removed.

Then, by Lemma 6.18, it suffices to prove that $\mathcal{M} - A \succeq 0$. Next, we turn to the shapes in $\mathcal{M}_{20}, \mathcal{M}_{02}$ and the disconnected shapes in $\mathcal{M}_{22}$.

**Disconnected shapes.** Several disconnected shapes in $\mathcal{L}$ (with $S_{\min} = \varnothing$) are not negligible. We note that all shapes in $\mathcal{M}_{11}$ must be connected due to the conditions in Definition 6.11. We will show that all disconnected shapes can be captured in a positive semidefinite component while introducing negligible errors.

We first introduce the following definition,

**Definition 6.20** (One-sided shape). *We say a shape is one-sided if either $U_a$ or $V_a$ is empty and there is no isolated component disconnected from $U_a$ or $V_a$. If $V_a = \varnothing$, we call it a left one-sided shape; if $U_a = \varnothing$, we call it a right one-sided shape.*

Note that the transpose of a left one-sided shape is a right one-sided shape, and further any disconnected shape in $\mathcal{M}_{22}$ contains a left and right one-sided shape. The main observation is that for any disconnected shape $a = (a_1, a_2^\top)$, $M_a \approx M_{a_1} M_{a_2}^\top$.

**Lemma 6.21.** *For a disconnected shape $a = (a_1, a_2^\top)$ where $a_1, a_2$ are left one-sided shapes,*

$$M_a = M_{a_1} M_{a_2}^\top + \mathcal{E}_{\text{collapse}(a_1, a_2^\top)}$$

*where $\mathcal{E}_{\text{collapse}(a_1, a_2^\top)}$ consists of shapes obtained from* collapsing *$a_1$ and $a_2^\top$. Moreover, all such collapsed shapes are negligible.*

The collapsed shapes are a result of graph matrix multiplication; see details in Section A.4. We will show that all disconnected shapes can be captured in a PSD component. The intuition is that since $\lambda_a = \lambda_{a_1}\lambda_{a_2}$, the term $\lambda_a M_a \approx (\lambda_{a_1} M_{a_1})(\lambda_{a_2} M_{a_2})^\top$.

**Lemma 6.22.** *Consider the first column of $\mathcal{M}'$: $(\mathcal{M}_{00}, 0, \mathcal{M}_{20})$, and let $v := (1, 0, \frac{\mathcal{M}_{20}}{\mathcal{M}_{00}})$. The matrix $\mathcal{M}_{00} \cdot vv^\top$ captures all disconnected shapes in $\mathcal{M}_{22}$ modulo some error consisting of negligible shapes.*

At this point, we can conclude that all shapes in $\mathcal{M}$ are accounted for and thus $\mathcal{M}$ is positive semidefinite. However, it does not exactly satisfy the constraints. We now proceed to prove that we can correct $\mathcal{M}$ with a small perturbation.

**Fixing the pseudo-distribution.** We show that we can "fix" $\mathcal{M}$ such that $\mathcal{M}_{\text{final}} = \mathcal{M} + \mathcal{E}$ satisfies all constraints exactly and that $\mathcal{E}$ is negligible. Suppose we view the pseudo-expectation $\widetilde{\mathbb{E}}$ as a flattened vector, then there exists a matrix $Q$ such that $Q\widetilde{\mathbb{E}} = 0$ if and only if $\widetilde{\mathbb{E}}$ satisfies all constraints. Here we assume that $\widetilde{\mathbb{E}}$ only includes even-degree monomials, since odd-degree monomials are zero already and don't need to be fixed.

To begin, in Section A.5 we show that if the truncation threshold $\tau$ in Definition 6.2 is chosen appropriately, then $\mathcal{M}$ already approximately satisfies the constraints, i.e. the norm of the "error vector" $\|Q\widetilde{\mathbb{E}}\|_2 \approx 0$.

**Lemma 6.23.** *There exist constants $C, C_1, c_2, c_3 > 0$ such that if $\varepsilon \geqslant \frac{C \log\log n}{\log n}$ and $\frac{C_1}{\varepsilon} \leqslant \tau \leqslant n^{c_2\varepsilon}$, then $\|Q\widetilde{\mathbb{E}}\|_2 \leqslant n^{-c_3\varepsilon\tau}$.*

Next, we fix $\widetilde{\mathbb{E}}$ by projecting it to the null space of $Q$,

$$\widetilde{\mathbb{E}}_{\text{final}} := \widetilde{\mathbb{E}} - Q^\top (QQ^\top)^\dagger Q\widetilde{\mathbb{E}},$$

where $(QQ^\top)^\dagger$ is the pseudo-inverse of $QQ^\top$ since it is not invertible. Clearly, $Q\widetilde{\mathbb{E}}_{\text{final}} = 0$.

Finally, to bound the norm of the correction $\|Q^\top (QQ^\top)^\dagger Q\widetilde{\mathbb{E}}\|_2$, it suffices to upper bound $\|Q\|$ and $\|(QQ^\top)^\dagger\|$. For $\|(QQ^\top)^\dagger\|$, we need to *lower bound* the smallest *nonzero* singular value of $Q$. We prove the following in Section A.6,

**Lemma 6.24.** *There exists a constant $C$ such that for $\varepsilon \geqslant \frac{C \log\log n}{\log n}$, $\|Q\| \leqslant \widetilde{O}(n)$ and the smallest nonzero eigenvalue of $QQ^\top$ is $\Omega(n^2)$.*

Lemma 6.23 and Lemma 6.24 immediately imply the following

**Lemma 6.25.** *There exist constants $C, C_1, c_2, c_3 > 0$ such that if $\varepsilon \geqslant \frac{C \log\log n}{\log n}$ and $\frac{C_1}{\varepsilon} \leqslant \tau \leqslant n^{c_2\varepsilon}$, then there exists a matrix $\mathcal{E}_{\text{fix}}$ that corrects the nonzero blocks of $\mathcal{M}$ such that $\mathcal{M} + \mathcal{E}_{\text{fix}}$ satisfies all constraints $\{g_s(x) = 0\}_{s \leqslant m}$ and that $\|\mathcal{E}_{\text{fix}}\| \leqslant n^{-\Omega(\varepsilon\tau)}$.*

*Proof.* $\|Q\| \cdot \|(QQ^\top)^\dagger\| \leqslant \widetilde{O}(1/n)$ due to Lemma 6.24. Thus, the correction $\|Q^\top (QQ^\top)^\dagger Q\widetilde{\mathbb{E}}\|_2 \leqslant \|Q\| \cdot \|(QQ^\top)^\dagger\| \cdot \|Q\widetilde{\mathbb{E}}\|_2 \leqslant n^{-\Omega(\varepsilon\tau)}$. □

**Putting things together.** We are ready to prove Lemma 6.3. The proof is essentially a summary of the results in this overview.

*Proof of Lemma 6.3.* The candidate moment matrix given by Definition 6.2 can be written as a sum of graph matrices of shapes in $\mathcal{L}$: $\mathcal{M} = \sum_{a \in \mathcal{L}} \lambda_a M_a$, and has the following structure,

$$\mathcal{M} = \begin{bmatrix} \mathcal{M}_{00} & 0 & \mathcal{M}_{02} \\ 0 & \mathcal{M}_{11} & 0 \\ \mathcal{M}_{20} & 0 & \mathcal{M}_{22} \end{bmatrix}.$$

By Lemma 6.25, we can correct the moment matrix so that $\mathcal{M}_{\text{final}} := \mathcal{M} + \mathcal{E}_{\text{fix}}$ satisfies all constraints with error $\|\mathcal{E}_{\text{fix}}\| \leqslant n^{-\Omega(\varepsilon\tau)}$.

Next, by Lemma 6.19, there exists a matrix $A$ such that $\mathcal{M}_{\text{final}} A = 0$ and that the spider term $\lambda_{a_{\text{spider}}} M_{a_{\text{spider}}}$ equals $A$ plus some errors:

$$\mathcal{M}' := \mathcal{M}_{\text{final}} - A = \begin{bmatrix} \mathcal{M}_{00} + \mathcal{E}_{00} & 0 & \mathcal{M}_{20}^\top + \mathcal{E}_{20}^\top \\ 0 & \mathcal{M}_{11} + \mathcal{E}_{11} & 0 \\ \mathcal{M}_{20} + \mathcal{E}_{20} & 0 & \mathcal{M}'_{22} + \mathcal{E}_{22} \end{bmatrix},$$

where $\mathcal{M}'_{22}$ is the block $\mathcal{M}_{22}$ with the spider removed, and the errors $|\mathcal{E}_{00}| = \widetilde{O}(n^{-3})$, $\|\mathcal{E}_{20}\| = \widetilde{O}(n^{-5/2})$, $\|\mathcal{E}_{22}\| = \widetilde{O}(n^{-2-\varepsilon})$, and $\|\mathcal{E}_{11}\| = n^{-\Omega(\varepsilon\tau)}$. Now, due to Lemma 6.18 it suffices to prove that $\mathcal{M}'$ is positive semidefinite.

Next, let $u := (1, 0, \frac{\mathcal{M}_{20} + \mathcal{E}_{20}}{\mathcal{M}_{00} + \mathcal{E}_{00}})$, the first column of $\mathcal{M}'$ divided by the first entry, and consider the matrix $(\mathcal{M}_{00} + \mathcal{E}_{00}) \cdot uu^\top$ (note that $\mathcal{M}_{00} + \mathcal{E}_{00} = 1 + o(1)$). This matrix is approximately the matrix $\mathcal{M}_{00} \cdot vv^\top$ in Lemma 6.22 that captures the disconnected shapes in $\mathcal{M}_{22}$:

$$\mathcal{M}' = (\mathcal{M}_{00} + \mathcal{E}_{00}) \cdot uu^\top + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathcal{M}_{11} + \mathcal{E}_{11} & 0 \\ 0 & 0 & \mathcal{M}''_{22} + \mathcal{E}''_{22} \end{bmatrix},$$

where $\mathcal{M}''_{22}$ is $\mathcal{M}'_{22}$ with the disconnected shapes removed, and $\mathcal{E}''_{22}$ contains error terms including negligible shapes, $\mathcal{M}_{20}\mathcal{E}_{20}^\top$, and $\mathcal{E}_{00}\mathcal{M}_{20}\mathcal{M}_{20}^\top$. By Lemma 6.16, the latter two have norms $\widetilde{O}(n^{-3}) \ll n^2$.

Finally, both $\mathcal{M}_{11} + \mathcal{E}_{11}$ and $\mathcal{M}''_{22} + \mathcal{E}''_{22}$ now only contain the trivial shapes and negligible shapes, hence by Lemma 6.15 they are PSD. This proves that $\mathcal{M}' \succeq 0$, which completes the proof. □

**Remark 6.26** (Comparison to the proof strategy of [GJJ+20])**.** Our proof is conceptually similar and builds heavily on the analysis in [GJJ+20] with some key differences. In [GJJ+20], the goal is to work with a special form of "rank 1" polynomial constraints $\{\langle x, g_i\rangle^2 = 1\}_{i\leqslant m}$ where the $g_i$s are random vectors (the "affine planes" problem). As a result, the construction of pseudo-distribution leads to a moment matrix with a different set of shapes playing a prominent role – 2-uniform graphs as opposed to 3-uniform hypergraphs in our case. As a result, several components in the proof (including the spectral norm bounds, the characterization of negligible shapes and spiders) are different.

Our analysis also requires dealing with certain disconnected shapes a bit differently by "charging" them to an appropriate extra PSD component. This actually leads to an important quantitative difference: in the result of [GJJ+20], the sum-of-squares lower bound (that works for $n^{O(1)}$-degree as against just degree 4 in our work) succeeds only for $m \leqslant n^{3/2-\varepsilon}$. This is despite the fact that low-degree hardness even for the rank-1 random polynomial above suggests a threshold of $m \leqslant n^{2-\varepsilon}$. In contrast, our analysis provides a nearly optimal lower bound at degree 4 that matches the prediction of low-degree hardness.

## Acknowledgments

We thank anonymous reviewers for their comments and suggestions. We also thank Alperen Ergür, Amit Sahai and Aayush Jain for illuminating discussions and pointing us to relevant related work. Finally, we would like to thank Sidhanth Mohanty and Jeff Xu for discussions on low-degree hardness and SoS lower bounds in general.

## References

[AGK21] Jackson Abascal, Venkatesan Guruswami, and Pravesh K. Kothari. *Strongly Refuting All Semi-Random Boolean CSPs*, page 454–472. Society for Industrial and Applied Mathematics, USA, 2021. 2

[ALMT14] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp. Living on the edge: Phase transitions in convex programs with random data. *Information and Inference: A Journal of the IMA*, 3(3):224–294, 2014. 5, 34

[AMP20] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. *arXiv preprint arXiv:1604.03423*, 2020. 10, 36, 37, 38

[AOW15] Sarah R Allen, Ryan ODonnell, and David Witmer. How to refute a random csp. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 689–708. IEEE, 2015. 2, 4

[AS17] Prabhanjan Ananth and Amit Sahai. Projective arithmetic functional encryption and indistinguishability obfuscation from degree-5 multilinear maps. In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, *Advances in Cryptology – EUROCRYPT 2017*, pages 152–181, Cham, 2017. Springer International Publishing. 2

[BBH+20] Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. *arXiv preprint arXiv:2009.06107*, 2020. 15

[BBKK18] Boaz Barak, Zvika Brakerski, Ilan Komargodski, and Pravesh K Kothari. Limits on low-degree pseudorandom generators (or: Sum-of-squares meets program obfuscation). In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 649–679. Springer, 2018. 3, 7

[BC11] Peter Bürgisser and Felipe Cucker. On a problem posed by Steve Smale. *Ann. of Math. (2)*, 174(3):1785–1836, 2011. 2

[BCK15] Boaz Barak, Siu On Chan, and Pravesh K. Kothari. Sum of squares lower bounds from pairwise independence. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 97–106, New York, NY, USA, 2015. Association for Computing Machinery. 2

[BGL17] Vijay V. S. P. Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Sum-of-squares certificates for maxima of random tensors on the sphere. In *APPROX-RANDOM*, volume 81 of *LIPIcs*, pages 31:1–31:20. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017. 4

[BHJ+19] Boaz Barak, Samuel B Hopkins, Aayush Jain, Pravesh Kothari, and Amit Sahai. Sum-of-squares meets program obfuscation, revisited. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 226–250. Springer, 2019. 3, 7

[BHK+16] Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *FOCS*, pages 428–437. IEEE Computer Society, 2016. 36

[BHK+19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019. 6, 7, 10, 15

[BM16] Boaz Barak and Ankur Moitra. Noisy tensor completion via the sum-of-squares hierarchy. In *COLT*, volume 49 of *JMLR Workshop and Conference Proceedings*, pages 417–445. JMLR.org, 2016. 2

[BP31] A. Bloch and G. Pólya. On the Roots of Certain Algebraic Equations. *Proc. London Math. Soc. (2)*, 33(2):102–114, 1931. 1

[BP08] Carlos Beltrán and Luis Miguel Pardo. On Smale's 17th problem: a probabilistic positive solution. *Found. Comput. Math.*, 8(1):1–43, 2008. 2

[BS09] Carlos Beltrán and Michael Shub. Complexity of Bezout's theorem. VII. Distance estimates in the condition metric. *Found. Comput. Math.*, 9(2):179–195, 2009. 1

[Can10] Emmanuel J. Candès. The power of convex relaxation: The surprising stories of matrix completion and compressed sensing. In *SODA*, page 1321. SIAM, 2010. 2

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

[dKNS20] Tommaso d'Orsi, Pravesh K. Kothari, Gleb Novikov, and David Steurer. Sparse pca: Algorithms, adversarial perturbations and certificates. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 553–564, 2020. 2

[DKWB19] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse pca. *arXiv preprint arXiv:1907.11635*, 2019. 6

[DM15] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 523–562, Paris, France, 03–06 Jul 2015. PMLR. 7

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009. 6

[Fei02] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 534–543. ACM, New York, 2002. 2

[Fei07] Uriel Feige. Refuting smoothed 3cnf formulas. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 407–417, 2007. 2

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):1–37, 2017. 15

[FKP19] Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient algorithm design. *Foundations and Trends® in Theoretical Computer Science*, 14(1-2):1–221, 2019. 12, 13, 14

[GJJ+20] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for sherrington-kirkpatrick via planted affine planes. *arXiv preprint arXiv:2009.01874*, 2020. 6, 10, 27, 35, 36, 38, 41, 44, 57, 59

[GJW20] David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. *arXiv preprint arXiv:2004.12063*, 2020. 6

[Gol00] Oded Goldreich. Candidate one-way functions based on expander graphs. *IACR Cryptology ePrint Archive*, 2000:63, 2000. 2

[Gro11] David Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011. 2

[HKP+16] Samuel B. Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. In *SODA*, pages 1079–1095. SIAM, 2016. 7

[HKP+17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017. 6, 10, 15, 36

[Hop18] Samuel Hopkins. Statistical inference and the sum of squares method. *PhD thesis, Cornell University*, 2018. 6, 15

[HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390. IEEE, 2017. 6, 15

[JH16] Cédric Josz and Didier Henrion. Strong duality in Lasserre's hierarchy for polynomial optimization. *Optim. Lett.*, 10(1):3–10, 2016. 14

[JLS21] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from well-founded assumptions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 60–73, 2021. 2

[Kac49] M. Kac. On the Average Number of Real Roots of a Random Algebraic Equation (II). *Proc. London Math. Soc. (2)*, 50(6):390–408, 1949. 1

[Kan85] R. Kannan. Solving systems of linear equations over polynomials. *Theoret. Comput. Sci.*, 39(1):69–88, 1985. 18

[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 SIGACT Symposium on Theory of Computing*, pages 132–145, 2017. 2

[Kos02] Eric Kostlan. On the expected number of real roots of a system of random polynomial equations. In *Foundations of computational mathematics (Hong Kong, 2000)*, pages 149–188. World Sci. Publ., River Edge, NJ, 2002. 2

[Kri64] Jean-Louis Krivine. Anneaux préordonnés. *Journal d'analyse mathématique*, 12(1):307–326, 1964. 12

[KWB19] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. *arXiv preprint arXiv:1907.11636*, 2019. 6, 9, 15

[KZ13] Gady Kozma and Ofer Zeitouni. On common roots of random Bernoulli polynomials. *Int. Math. Res. Not. IMRN*, (18):4334–4347, 2013. 1

[Lai17] Pierre Lairez. A deterministic algorithm to compute approximate roots of polynomial systems in polynomial average time. *Found. Comput. Math.*, 17(5):1265–1292, 2017. 2

[Las01] Jean B. Lasserre. New positive semidefinite relaxations for nonconvex quadratic programs. In *Advances in convex analysis and global optimization (Pythagorion, 2000)*, volume 54 of *Nonconvex Optim. Appl.*, pages 319–331. Kluwer Acad. Publ., Dordrecht, 2001. 14

[Lin16] Huijia Lin. Indistinguishability obfuscation from constant-degree graded encoding schemes. In *Proceedings, Part I, of the 35th Annual International Conference on Advances in Cryptology — EUROCRYPT 2016 - Volume 9665*, page 28–57, Berlin, Heidelberg, 2016. Springer-Verlag. 2

[Lin17] Huijia Lin. Indistinguishability obfuscation from sxdh on 5-linear maps and locality-5 prgs. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology – CRYPTO 2017*, pages 599–629, Cham, 2017. Springer International Publishing. 2

[LT17] Huijia Lin and Stefano Tessaro. Indistinguishability obfuscation from trilinear maps and block-wise local prgs. Cryptology ePrint Archive, Report 2017/250, 2017. https://eprint.iacr.org/2017/250. 2

[LV17] Alex Lombardi and Vinod Vaikuntanathan. Limits on the locality of pseudorandom generators and applications to indistinguishability obfuscation. In Yael Kalai and Leonid Reyzin, editors, *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part I*, volume 10677 of *Lecture Notes in Computer Science*, pages 119–137. Springer, 2017. 3, 7

[MPW15] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *STOC*, pages 87–96. ACM, 2015. 7

[MRX20] Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: degree-2 to degree-4. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 840–853, 2020. 6

[Nes00] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, volume 33 of *Appl. Optim.*, pages 405–440. Kluwer Acad. Publ., Dordrecht, 2000. 14

[Par00] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000. 14

[Pit97] Toniann Pitassi. Algebraic propositional proof systems. In *Descriptive complexity and finite models (Princeton, NJ, 1996)*, volume 31 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 215–244. Amer. Math. Soc., Providence, RI, 1997. 13

[PS17] Aaron Potechin and David Steurer. Exact tensor completion with sum-of-squares. In *COLT*, volume 65 of *Proceedings of Machine Learning Research*, pages 1619–1673. PMLR, 2017. 2

[Rec11] Benjamin Recht. A simpler approach to matrix completion. *Journal of Machine Learning Research*, 12:3413–3430, 2011. 2

[RRS17] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold. In *STOC*, pages 121–131. ACM, 2017. 2, 4

[Sho87] N. Z. Shor. Quadratic optimization problems. *Izv. Akad. Nauk SSSR Tekhn. Kibernet.*, (1):128–139, 222, 1987. 14

[Shu09] Michael Shub. Complexity of Bezout's theorem. VI. Geodesics in the condition (number) metric. *Found. Comput. Math.*, 9(2):171–178, 2009. 1

[SS93a] M. Shub and S. Smale. Complexity of Bezout's theorem. II. Volumes and probabilities. In *Computational algebraic geometry (Nice, 1992)*, volume 109 of *Progr. Math.*, pages 267–285. Birkhäuser Boston, Boston, MA, 1993. 1

[SS93b] Michael Shub and Steve Smale. Complexity of Bézout's theorem. I. Geometric aspects. *J. Amer. Math. Soc.*, 6(2):459–501, 1993. 1

[SS93c] Michael Shub and Steve Smale. Complexity of Bezout's theorem. III. Condition number and packing. volume 9, pages 4–14. 1993. Festschrift for Joseph F. Traub, Part I. 1

[SS94] M. Shub and S. Smale. Complexity of Bezout's theorem. V. Polynomial time. volume 133, pages 141–164. 1994. Selected papers of the Workshop on Continuous Algorithms and Complexity (Barcelona, 1993). 1

[SS96] Michael Shub and Steve Smale. Complexity of Bezout's theorem. IV. Probability of success; extensions. *SIAM J. Numer. Anal.*, 33(1):128–148, 1996. 1

[Ste74] Gilbert Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Mathematische Annalen*, 207(2):87–97, 1974. 12

[SW08] Rolf Schneider and Wolfgang Weil. *Stochastic and integral geometry*. Springer Science & Business Media, 2008. 34

[SW20] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *arXiv preprint arXiv:2008.02269*, 2020. 6, 9

[Sze39] Gabor Szegö. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939. 11

[Wei20] Alexander S Wein. Optimal low-degree hardness of maximum independent set. *arXiv preprint arXiv:2010.06563*, 2020. 6

[Wik] Wikipedia. Smale's 17 problems. 2

# A Omitted Proofs

## A.1 Contributions from negligible shapes

In this section, we look at what shapes are negligible, and we show that the total contribution of the negligible shapes is dominated by the trivial shapes.

By the graph matrix norm bounds (Proposition 6.10) and (14),

$$|\lambda_a| \cdot \|M_a\| \leqslant n^{-|E(a)|-\frac{|U_a|+|V_a|}{2}} \cdot n^{\frac{w(V(a))-w(S_{\min})+w(W_{\text{iso}})}{2}} \cdot (|E(a)| \cdot \log n)^{O(|E(a)|)}, \tag{16}$$

since $|V(a)| \leqslant O(|E(a)|)$ for $a \in \mathcal{L}$.

In the following lemma, we identify a quantity that indicates if a shape is negligible or not.

**Lemma A.1.** *For a shape $a \in \mathcal{L}$, define*

$$\varphi(a) := |E(a)| - \frac{w(V(a)) - w(S_{\min}) + w(W_{\text{iso}})}{2}. \tag{17}$$

*Then, there exist constants $C, c_1 > 0$ such that for $\varepsilon = \frac{C \log\log n}{\log n}$ and $\tau = n^{c_1\varepsilon}$, if a shape $a \in \mathcal{L}$ satisfies $|E(a)| \neq 0$ and*

$$\varphi(a) \geqslant \frac{\varepsilon}{8}|E(a)|,$$

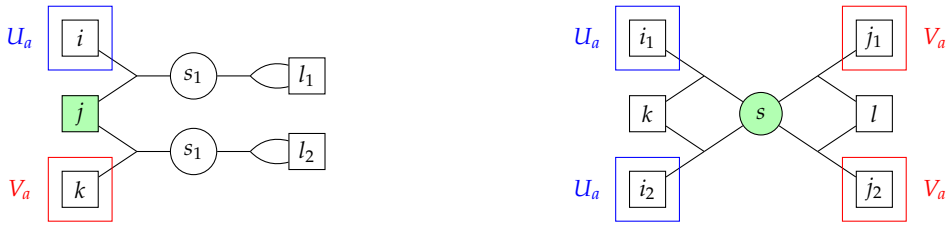
*then it is negligible.*

*Proof.* First, $W_{\text{iso}} = \varnothing$ due to Condition 5 of $\mathcal{L}$ (in Definition 6.11). If we choose $C$ to be sufficiently large and $c_1$ sufficiently small, then since $|E(a)| \leqslant \tau = n^{c_1\varepsilon}$, by (16) we have

$$\|\lambda_a M_a\| \leqslant n^{-\frac{|U_a|+|V_a|}{2}-\varphi(a)} n^{O(\varepsilon|E(a)|)} \leqslant n^{-\frac{|U_a|+|V_a|}{2}} \cdot n^{-c_2\varepsilon|E(a)|}$$

for some constant $c_2$. This satisfies the definition of negligible shapes (Definition 6.14), thus $a$ is negligible. □

Figure 5 includes some negligible shapes for illustration.




(a) $\varphi(a) = \varepsilon = \frac{\varepsilon}{4}|E(a)|$. (b) $\varphi(a) = 1$.

Figure 5: Negligible shapes.

Next, we analyze the quantity $\varphi(a)$. To do so, we first introduce some more notations. For a shape $a$, let $S_a, C_a$ be the set of square and circle vertices respectively. Let $\widetilde{U}_a := U_a \setminus (U_a \cap V_a)$ and $\widetilde{V}_a := V_a \setminus (U_a \cap V_a)$, and let $\widetilde{S}_{\min} := S_{\min} \setminus (U_a \cap V_a)$. The conditions of $\mathcal{L}$ in Definition 6.11 ensure that

1. The degree of any square vertex in $\widetilde{U}_a \cup \widetilde{V}_a$ must be odd; minimum degree 1,

2. The degree of any square vertex in $U_a \cap V_a$ must be even; minimum degree 0,

3. The degree of any other vertex must be even; minimum degree 2.

**Definition A.2** (Large-degree vertex)**.** *We say that a vertex has large degree if its degree is larger than the minimum degree (and must be larger by at least 2 due to the conditions in Definition 6.11).*

The following lemma lets us determine whether a shape in $\mathcal{L}$ is negligible or not,

**Lemma A.3.** *For a shape $a \in \mathcal{L}$, let $\delta_s, \delta_c$ be the number of large-degree square and circle vertices,*

$$\varphi(a) \geqslant -\frac{1}{4}(|\widetilde{U}_a| + |\widetilde{V}_a|) + \frac{\varepsilon}{4}|E(a)| + \frac{1}{2}w(\widetilde{S}_{\min}) + \frac{1}{2}\delta_s + \left(1 - \frac{\varepsilon}{2}\right)\delta_c\,.$$

*Proof.* First, by Condition 5 of $\mathcal{L}$, $W_{\text{iso}} = \varnothing$. Moreover, $U_a \cap V_a$ must be in the minimum vertex separator, so their contributions in (17) cancel out. Then, using $w(\square) = 1$ and $w(\bigcirc) = 2 - \varepsilon$, we can rewrite $\varphi(a)$ as

$$\varphi(a) = |E(a)| - \frac{1}{2}|\widetilde{S}_a| - \left(1 - \frac{\varepsilon}{2}\right)|C_a| + \frac{1}{2}w(\widetilde{S}_{\min})\,,$$

where $\widetilde{S}_a$ is the set of square vertices excluding $U_a \cap V_a$.

By Condition 2 of $\mathcal{L}$, each square vertex in $\widetilde{U}_a, \widetilde{V}_a$ must have degree at least 1, and each square vertex in $\widetilde{S}_a \setminus (\widetilde{U}_a \cup \widetilde{V}_a)$ must have degree at least 2 (they are not isolated). Each large degree vertex introduces at least two more. Thus, the total number of hyperedges

$$|E(a)| \geqslant \frac{1}{2}\left(|\widetilde{U}_a| + |\widetilde{V}_a| + 2|\widetilde{S}_a \setminus (\widetilde{U}_a \cup \widetilde{V}_a)| + 2\delta_s\right) = |\widetilde{S}_a| - \frac{1}{2}\left(|\widetilde{U}_a| + |\widetilde{V}_a|\right) + \delta_s\,.$$

Moreover, each circle vertex must have degree at least 2. Thus,

$$|E(a)| \geqslant 2|C_a| + 2\delta_c.$$

Combining the two, we get

$$\begin{aligned}\varphi(a) &= |E(a)| - \left(\frac{1}{2}|E(a)| + \frac{1}{4}(|\widetilde{U}_a| + |\widetilde{V}_a|) - \frac{1}{2}\delta_s\right) - \left(1 - \frac{\varepsilon}{2}\right)\left(\frac{1}{2}|E(a)| - \delta_c\right) + \frac{1}{2}w(\widetilde{S}_{\min}) \\ &\geqslant -\frac{1}{4}(|\widetilde{U}_a| + |\widetilde{V}_a|) + \frac{\varepsilon}{4}|E(a)| + \frac{1}{2}w(\widetilde{S}_{\min}) + \frac{1}{2}\delta_s + \left(1 - \frac{\varepsilon}{2}\right)\delta_c\,.\end{aligned}$$

This completes the proof. ∎

We now prove Lemma 6.15. We first derive a bound on the number of negligible shapes.

**Lemma A.4.** *Let $\ell \geqslant 2$. The number of shapes with exactly $\ell$ edges and no isolated vertices is at most $\ell^{O(\ell)}$.*

*Proof.* Since each hyperedge connects to 3 vertices, there can be at most $3\ell$ vertices. If there are $v$ vertices, then there are $v^{3\ell}$ ways to assign edges. Thus, in total, the number of shapes is at most $\sum_v v^{3\ell} \leqslant (3\ell) \cdot (3\ell)^{3\ell} \leqslant \ell^{O(\ell)}$. ∎

**Lemma A.5** (Restatement of Lemma 6.15)**.** *For $k \in \{0,1,2\}$, let $\mathcal{L}_{\mathrm{negl},k}$ be the set of negligible shapes in block $\mathcal{M}_{kk}$, and let $\mathcal{E}_{\mathrm{negl},k} := \sum_{a\in\mathcal{L}_{\mathrm{negl},k}} \lambda_a M_a$. There exist constants $c_1, c_2 > 0$ such that if the threshold $\tau \leqslant n^{c_1\varepsilon}$, then*

$$\|\mathcal{E}_{\mathrm{negl},k}\| \leqslant n^{-k-c_2\varepsilon}\,.$$

*This implies that*

$$\lambda_{a_{\mathrm{triv},k}} M_{a_{\mathrm{triv},k}} + \mathcal{E}_{\mathrm{negl},k} \succ 0\,.$$

*Proof.* The shapes in $\mathcal{L}_{\mathrm{negl},k}$ can have number of hyperedges ranging from 2 to $\tau$. By Lemma A.4, Definition 6.14 of the negligible shapes, and the triangle inequality,

$$\|\mathcal{E}_{\mathrm{negl},k}\| \leqslant \sum_{a\in\mathcal{L}_{\mathrm{negl},k}} |\lambda_a| \cdot \|M_a\| \leqslant \sum_{\ell=2}^{\tau} \ell^{O(\ell)} \cdot n^{-k-\Omega(\varepsilon\ell)} \leqslant n^{-k} \sum_{\ell=2}^{\tau} n^{-c_2'\varepsilon\ell}\,,$$

for some constant $c_2' > 0$ provided that $c_1$ is small enough. The summation is upper bounded by $n^{-c_2\varepsilon}$ for some $c_2$. Thus, $\|\mathcal{E}_{\mathrm{negl},k}\| \leqslant n^{-k-c_2\varepsilon} \leqslant o(n^{-k})$, much smaller than the minimum singular value of $\lambda_{a_{\mathrm{triv},k}} M_{a_{\mathrm{triv},k}}$. This completes the proof. □

The same analysis shows that the norm of block $\mathcal{M}_{k\ell}$ is dominated by the shape with the largest norm.

**Lemma A.6** (Restatement of Lemma 6.16)**.** *There exists a constant $c_1 > 0$ such that if the threshold $\tau \leqslant n^{c_1\varepsilon}$, then for any $k, \ell$, $\|\mathcal{M}_{k\ell}\| \leqslant n^{-\frac{k+\ell}{4}}$.*

*Proof.* By Lemma A.3, $\varphi(a) \geqslant -\frac{|U_a|+|V_a|}{4} - \frac{\varepsilon}{4}|E(a)|$. The same calculations show that

$$\|\mathcal{M}_{k\ell}\| \leqslant \sum_{\substack{a\in\mathcal{L}:\\ |U_a|=k, |V_a|=\ell}} \|\lambda_a M_a\| \leqslant n^{-\frac{k+\ell}{4}}\,.$$

□

## A.2 Non-trivial non-spider connected shapes are negligible

We say that a shape is connected if there is path from $U_a$ to $V_a$. For connected shapes, we will show that except for *one* shape, all connected shapes can be charged to the trivial shapes. We call that shape a *spider*; see Figure 4.

We first show the following result about the structure of connected shapes.

**Lemma A.7.** *For connected shapes $\alpha \in \mathcal{L}$, suppose $|U_a| = |V_a| = 2$, $U_a \cap V_a = \varnothing$, and the minimum vertex separator contains only one square vertex, then that vertex must have degree at least 4.*

*Proof.* Suppose for contradiction $S_{\min}$ contains one square vertex $\boxed{i}$ and it has degree 2 (it cannot be isolated). Then, consider the left and right sides of the graph separated by this vertex. Since $\boxed{i}$ must contribute exactly one degree to each side, the total degree of square vertices on each side must be odd (each hyperedge contributes two degrees). However, by the conditions in Definition 6.11, the degree of any $\boxed{j} \notin U_a \cup V_a$ must be even, whereas the degree of $\boxed{j} \in U_a \cup V_a$ must be odd, thus due to $|U_a| = |V_a| = 2$ the total degree must be even. This is a contradiction. Thus, $\deg(\boxed{i})$ must be larger than 2, which means it must be at least 4. □

Next, we show that all non-spider non-trivial connected shapes are negligible. Recall that a shape is negligible if $\varphi(a) \geqslant \frac{\varepsilon}{8}|E(a)|$, and this can be determined by Lemma A.3.

**Lemma A.8** (Restatement of Lemma 6.17)**.** *If $a \in \mathcal{L}$ is a connected shape and not a trivial shape nor a spider, then $a$ is negligible.*

*Proof.* First, observe that for connected shapes, we must have $|U_a| = |V_a| \leqslant 2$. We split into several cases and apply Lemma A.3,

1. $U_a = V_a$. In this case, $\widetilde{U}_a, \widetilde{V}_a = \varnothing$, and since $a$ is not a trivial shape, $|E(a)| > 0$. Thus, $\varphi(a) \geqslant \frac{\varepsilon}{4}|E(a)|$.
2. $|U_a| = |V_a| = 1$ and $U_a \cap V_a = \varnothing$. Such shapes must be connected, so $w(S_{\min})$ is at least 1, which cancels out with $-\frac{1}{4}(|\widetilde{U}_a| + |\widetilde{V}_a|) = -\frac{1}{2}$. Thus, $\varphi(a) \geqslant \frac{\varepsilon}{4}|E(a)|$.
3. $|U_a| = |V_a| = 2$ and $|U_a \cap V_a| = 1$. Since $|\widetilde{U}_a| = |\widetilde{V}_a| = 1$ and $\widetilde{U}_a, \widetilde{V}_a$ must be connected, this is the same as the previous case.
4. $|U_a| = |V_a| = 2$, $U_a \cap V_a = \varnothing$, and $w(S_{\min}) = 1$ or $w(S_{\min}) \geqslant 2$. First, if $S_{\min}$ contains just one square vertex, then by Lemma A.7 it must have large degree, hence $\frac{1}{2}w(S_{\min}) + \frac{1}{2}\delta_s \geqslant 1$, canceling out the term $-\frac{1}{4}(|U_a| + |V_a|) = -1$. On the other hand, if $w(S_{\min}) \geqslant 2$, then clearly it already cancels out the $-1$.
5. $|U_a| = |V_a| = 2$, $U_a \cap V_a = \varnothing$, and $w(S_{\min}) = 2 - \varepsilon$. In this case, $S_{\min}$ contains exactly one circle vertex, and $\varphi(a) \geqslant -1 + \frac{\varepsilon}{4}|E(a)| + (1 - \varepsilon/2) = \varepsilon(\frac{1}{4}|E(a)| - \frac{1}{2})$. Now, if $a$ is a spider, then $|E(a)| = 2$ and $\varphi(a) = 0$. Fortunately, for other non-spider shapes, $|E(a)| \geqslant 4$, which means $\varphi(a) \geqslant \frac{\varepsilon}{8}|E(a)|$.

Thus, except the spider, all non-trivial connected shapes have $\varphi(a) \geqslant \frac{\varepsilon}{8}|E(a)|$. □

### A.3 The spider is close to the null space of $\mathcal{M}$

As described in the proof overview, we first identify the null space of the moment matrix $\mathcal{M}$ that satisfies all constraints exactly: for any $I \subseteq [n]$ ($|I| \leqslant 2$) and $s \in [m]$,

$$0 = \widetilde{\mathbb{E}}[x^I(x^\top G_s x)] = \sum_{i \neq j} G^s_{ij} \widetilde{\mathbb{E}}[x^I x_i x_j] + \frac{1}{n} \sum_i G^s_{ii} \widetilde{\mathbb{E}}[x^I],$$

where we use $x_i^2 = \frac{1}{n}$. We can represent the above using a matrix $L_2 = M_{a_1} + \frac{1}{n} M_{a_2}$ (drawn as graph matrices in Figure 6), with rows indexed by $s \in [m]$ and columns indexed by $I \subseteq [n]$. It is easy to see that the $(s, I)$ entry of $L_2\mathcal{M}$ is exactly $\widetilde{\mathbb{E}}[x^I(x^\top G_s x)]$. One can view $L_2$ as a "check matrix", i.e. if $\mathcal{M}$ exactly satisfies the constraints, then $L_2\mathcal{M} = 0$.

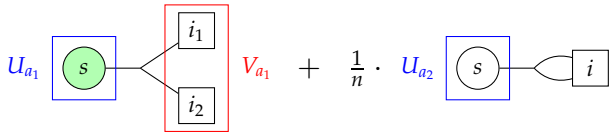



Figure 6: $L_2 = M_{a_1} + \frac{1}{n} M_{a_2}$.

Next, we prove that $M_{a_{\text{spider}}}$ is close to the null space of $\mathcal{M}$.

**Lemma A.9** (Restatement of Lemma 6.19). *Suppose $\mathcal{M}$ exactly satisfies all constraints $\{g_s(x) = 0\}_{s\leqslant m}$. Then there exists a matrix $A$ such that $\mathcal{M}A = 0$ and*

$$\lambda_{a_{\text{spider}}} M_{a_{\text{spider}}} = A + \mathcal{E}_{00} + \mathcal{E}_{20} + \mathcal{E}_{20}^\top + \mathcal{E}_{22},$$

*where $\mathcal{E}_{00}$, $\mathcal{E}_{20}$, $\mathcal{E}_{20}^\top$, $\mathcal{E}_{22}$ are errors in blocks $\mathcal{M}_{00}$, $\mathcal{M}_{20}$, $\mathcal{M}_{02}$, $\mathcal{M}_{22}$ respectively, and $|\mathcal{E}_{00}| = \widetilde{O}(n^{-3})$, $\|\mathcal{E}_{20}\| = \widetilde{O}(n^{-5/2})$, and $\|\mathcal{E}_{22}\| = \widetilde{O}(n^{-2-\varepsilon})$.*

*Proof.* Consider the matrix $L_2$ and shapes $a_1, a_2$ in Figure 6. Using the graph matrix norm bounds (Proposition 6.10), $\|M_{a_1}\| = \widetilde{O}(n)$ and $\|\frac{1}{n}M_{a_2}\| = \widetilde{O}(n^{\frac{1}{2}-\frac{\varepsilon}{2}})$. Now, we consider $L_2^\top L_2$:

$$L_2^\top L_2 = M_{a_1}^\top M_{a_1} + \frac{1}{n}(M_{a_2}^\top M_{a_1} + M_{a_1}^\top M_{a_2}) + \frac{1}{n^2}M_{a_2}^\top M_{a_2} = M_{a_1}^\top M_{a_1} + E'_{20} + E'^\top_{20} + E'_{00}.$$

where $\|E'_{20}\| \leqslant \widetilde{O}(n^{\frac{3}{2}-\frac{\varepsilon}{2}})$ and $\|E'_{00}\| \leqslant \widetilde{O}(n^{1-\varepsilon})$. For the first term $M_{a_1}^\top M_{a_1}$, for $i_1 \neq i_2$ and $j_1 \neq j_2$,

$$M_{a_1}^\top M_{a_1}(\{i_1, i_2\}, \{j_1, j_2\}) = \sum_{s\in[m]} G^s_{i_1 i_2} G^s_{j_1 j_2}.$$

Represented using graph matrix multiplication, $M_{a_1}^\top M_{a_1}$ is a sum of shapes in Figure 7. Note that the last two graphs come from the term $M_{a_1}^\top M_{a_1}(\{i_1, i_2\}, \{i_1, i_2\}) = \sum_{s\in[m]}(G^s_{i_1 i_2})^2$ and using the fact $h_1(z)^2 = (z^2-1)+1 = h_2(z) + h_0(z)$.

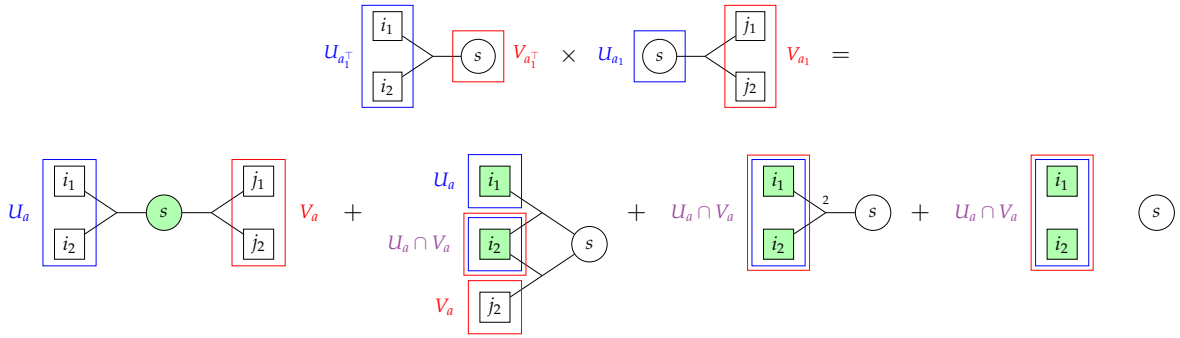



Figure 7: Expansion of $M_{a_1}^\top M_{a_1}$.

Observe that the first shape in Figure 7 is the spider, which is the dominating term in the expansion: $\|M_{a_{\text{spider}}}\| = \widetilde{O}(n^2)$, whereas the rest of the shapes have norms $\widetilde{O}(n^{2-\varepsilon})$. Since $\lambda_{a_{\text{spider}}} = n^{-4}$, we can rewrite the spider term

$$\lambda_{a_{\text{spider}}} M_{a_{\text{spider}}} = n^{-4}(M_{a_1}^\top M_{a_1} + E'_{22}) = n^{-4}L_2^\top L_2 + \mathcal{E}_{00} + \mathcal{E}_{20} + \mathcal{E}_{20}^\top + \mathcal{E}_{22},$$

where $\|\mathcal{E}_{00}\| = \widetilde{O}(n^{-3})$, $\|\mathcal{E}_{20}\| = \widetilde{O}(n^{-5/2})$, and $\|\mathcal{E}_{22}\| = \widetilde{O}(n^{-2-\varepsilon})$. ∎

Thus, since $\mathcal{M} \cdot L_2^\top L_2 = 0$, by Lemma 6.18 it suffices to show that $\mathcal{M} - n^{-4}L_2^\top L_2 \succeq 0$. This essentially kills the spider term $\lambda_{a_{\text{spider}}} M_{a_{\text{spider}}}$ in $\mathcal{M}$.

## A.4 Disconnected shapes are captured in a PSD component

Disconnected shapes are the shapes where $U_a, V_a$ are disconnected, i.e. there is no minimum vertex separator. Many of these shapes are not negligible compared to the trivial shapes. However, we will prove that such shapes can be captured in a PSD component of $\mathcal{M}$.

Given a disconnected shape $a = (a_1, a_2^\top)$ where $a_1, a_2$ are left one-sided shapes, we first analyze the multiplication $M_{a_1} M_{a_2}^\top$ (recall that $M_{a_1}, M_{a_2}$ are both vectors). Consider the example in Figure 8. For $i_1 \neq i_2 \neq j_1 \neq j_2$, the entry of $M_{a_1} M_{a_2}^\top$ is

$$\begin{aligned}(M_{a_1} M_{a_2}^\top)(\{i_1,i_2\},\{j_1,j_2\}) &= \left(\sum_{\substack{k\neq i_1,i_2\\ s_1\in[m]}} G^{s_1}_{i_1k} G^{s_1}_{ki_2}\right)\cdot\left(\sum_{\substack{\ell\neq j_1,j_2\\ s_2\in[m]}} G^{s_2}_{j_1j_2} G^{s_2}_{\ell\ell}\right)\\ &= \sum_{\substack{k\neq \ell\notin\{i_1,i_2,j_1,j_2\}\\ s_1\neq s_2}} G^{s_1}_{i_1k} G^{s_1}_{ki_2} G^{s_2}_{j_1j_2} G^{s_2}_{\ell\ell} + \sum_{\substack{k\neq \ell\notin\{i_1,i_2,j_1,j_2\}\\ s_1= s_2}} G^{s_1}_{i_1k} G^{s_1}_{ki_2} G^{s_1}_{j_1j_2} G^{s_1}_{\ell\ell}\\ &\quad + \sum_{\substack{k=\ell\notin\{i_1,i_2,j_1,j_2\}\\ s_1\neq s_2}} G^{s_1}_{i_1k} G^{s_1}_{ki_2} G^{s_2}_{j_1j_2} G^{s_2}_{kk} + \sum_{\substack{k=\ell\notin\{i_1,i_2,j_1,j_2\}\\ s_1= s_2}} G^{s_1}_{i_1k} G^{s_1}_{ki_2} G^{s_1}_{j_1j_2} G^{s_1}_{kk} + \cdots.\end{aligned}$$

This expansion introduces several graph matrices, the first is the disconnected shape $a$. The first two terms are drawn out in Figure 8. For other entries such as $i := i_1 = j_1$ and $i \neq i_2 \neq j_2$,

$$\begin{aligned}(M_{a_1} M_{a_2}^\top)(\{i,i_2\},\{i,j_2\}) &= \sum_{\substack{k\neq \ell\notin\{i,i_2,j_2\}\\ s_1\neq s_2}} G^{s_1}_{ik} G^{s_1}_{ki_2} G^{s_2}_{ij_2} G^{s_2}_{\ell\ell} + \sum_{\substack{k\neq \ell\notin\{i,i_2,j_2\}\\ s_1= s_2}} G^{s_1}_{ik} G^{s_1}_{ki_2} G^{s_1}_{ij_2} G^{s_1}_{\ell\ell}\\ &\quad + \sum_{\substack{k=\ell\notin\{i,i_2,j_2\}\\ s_1\neq s_2}} G^{s_1}_{ik} G^{s_1}_{ki_2} G^{s_2}_{ij_2} G^{s_2}_{kk} + \sum_{\substack{k=\ell\notin\{i,i_2,j_2\}\\ s_1= s_2}} G^{s_1}_{ik} G^{s_1}_{ki_2} G^{s_1}_{ij_2} G^{s_1}_{kk} + \cdots.\end{aligned}$$

These terms correspond to the collapsed shapes: each collapsed shape is obtained by iteratively merging one vertex from the left one-sided shape with one vertex (of the same type) from the right one-sided shape. The first and second terms above are drawn out in Figure 8.

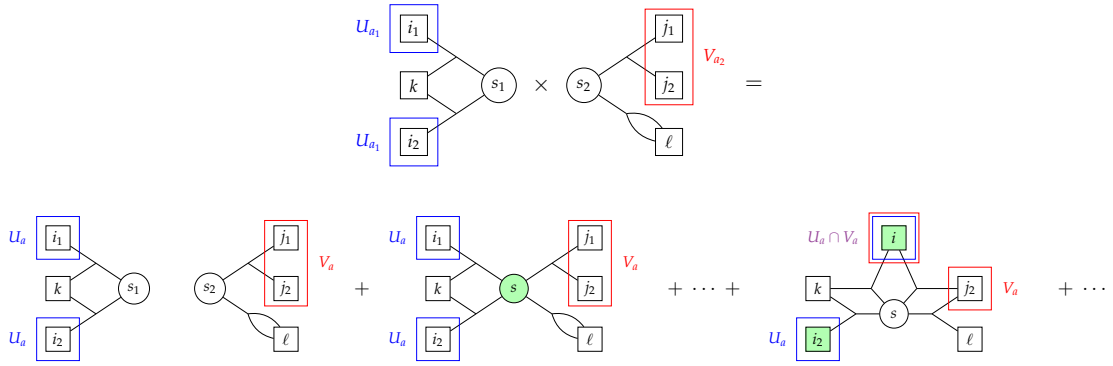



Figure 8: Shapes introduced by multiplying $M_{a_1} M_{a_2}^\top$. The first is a disconnected shape; the second is by collapsing $s_1, s_2$ (setting $s_1 = s_2$); the third is by collapsing $i_1, j_1$ and $s_1, s_2$ (setting $s_1 = s_2$ and $i_1 = j_1$).

Now, we proceed to prove Lemma 6.21.

**Lemma A.10** (Restatement of Lemma 6.21)**.** *For a disconnected shape $a = (a_1, a_2^\top)$ where $a_1, a_2$ are left one-sided shapes,*

$$M_a = M_{a_1} M_{a_2}^\top + \mathcal{E}_{\text{collapse}(a_1, a_2^\top)}$$

*where $\mathcal{E}_{\text{collapse}(a_1, a_2^\top)}$ consists of shapes obtained from* collapsing *$a_1$ and $a_2^\top$. Moreover, all such collapsed shapes are negligible.*

*Proof.* By the discussions above, expanding the matrix $M_{a_1} M_{a_2}^\top$ results in a summation of several shapes, one of which is the disconnected shape $a = (a_1, a_2^\top)$. Now, it suffices to show that all collapsed shapes are negligible.

For a collapsed shape $a$, since it is connected, $S_{\min}$ must contain at least one vertex. Moreover, the merged vertices must have large degree (recall Definition A.2), hence $\delta_s$ or $\delta_c$ must be at least 1. Since $|U_a| + |V_a| = 4$, by Lemma A.3 we must have $\varphi(a) \geqslant \frac{\varepsilon}{4}|E(a)|$, meaning that $a$ is negligible.

Some care is required if the collapsed shape has parallel edges, which may happen if the endpoints of two different hyperedges collapse. We handle this by breaking the parallel edge into a sum of graph matrices. For example, suppose a shape collapsed from $a_1$ and $a_2^\top$ has two parallel edges $e$ labeled 1. Then by $h_1(z)^2 = h_2(z) + 1$, we get a sum of two shapes $b_1, b_2$ (with the same coefficient $\lambda_{a_1}\lambda_{a_2}$), where $b_1$ has the same edge $e$ labeled 2, and $b_2$ has no edge (and may have an isolated vertex). Clearly, $b_1$ is negligible, but to show that $b_2$ is also negligible requires some work.

Let $a = (a_1, a_2^\top)$ be the disconnected shape, and consider a shape $b$ with isolated vertices collapsed from $a_1$ and $a_2^\top$. This collapsed shape introduces an error $\lambda_a M_b$, and we must show that $\|\lambda_a M_b\| \leqslant O(n^{-2-\Omega(\varepsilon|E(b)|)})$. Let $E_{\text{del}}$ be the set of deleted edges, and note that $|\lambda_a| = |\lambda_b| \cdot O(n^{-|E_{\text{del}}|})$ and $\|\lambda_b M_b\| = \widetilde{O}(n^{-2-\varphi(b)})$ (recall the definition of $\varphi$ in (17)). Thus, it suffices to show that

$$\varphi(b) + |E_{\text{del}}| \geqslant \Omega(\varepsilon|E(b)|).$$

Let $S_{\text{iso}}, C_{\text{iso}}$ be the set of isolated square and circle vertices respectively, and let $b' = b \setminus (S_{\text{iso}} \cup C_{\text{iso}})$, the shape without the isolated vertices. Clearly, $\varphi(b) = \varphi(b') - w(W_{\text{iso}}) = \varphi(b') - |S_{\text{iso}}| - (2-\varepsilon)|C_{\text{iso}}|$, and further by Lemma A.3, $\varphi(b') \geqslant -1 + \frac{\varepsilon}{4}|E(b)| + \frac{1}{2}w(S_{\min})$. Next, observe that the vertices in $W_{\text{iso}}$ must have degree $\geqslant 4$ before the edges were removed. We consider two cases:

1. The isolated vertices were originally connected to circle vertices only: In this case, we have $w(S_{\min}) \geqslant 2 - \varepsilon$. Further, $|E_{\text{del}}| \geqslant \frac{1}{2} \cdot 4|S_{\text{iso}}|$ and $|E_{\text{del}}| \geqslant 4|C_{\text{iso}}| + 2$. Thus,

$$\varphi(b) + |E_{\text{del}}| \geqslant -1 + \frac{\varepsilon}{4}|E(b)| + \frac{1}{2}(2-\varepsilon) - \frac{1}{2}|E_{\text{del}}| - (2-\varepsilon)\frac{|E_{\text{del}}| - 2}{4} + |E_{\text{del}}| \geqslant \frac{\varepsilon}{4}|E(b)|\,.$$

2. The isolated vertices were originally connected to some square vertices: In this case, we have $w(S_{\min}) \geqslant 1$. Observe that the originally connecting square vertices must contribute at least 2 deleted edges. Thus, $|E_{\text{del}}| \geqslant 2|S_{\text{iso}}| + 2$ and $|E_{\text{del}}| \geqslant 4|C_{\text{iso}}|$, and we have

$$\varphi(b) + |E_{\text{del}}| \geqslant -1 + \frac{\varepsilon}{4}|E(b)| + \frac{1}{2} - \frac{|E_{\text{del}}| - 2}{2} - (2-\varepsilon)\frac{|E_{\text{del}}|}{4} + |E_{\text{del}}| \geqslant \frac{\varepsilon}{4}|E(b)|\,.$$

In both cases, the collapsed shape is negligible. This completes the proof. ∎

Finally, we handle the disconnected shapes with an additional disconnected component. Consider $a = (a_1, a_2)$ where $a_1$ is a left one-sided shape and $a_2$ is the disconnected component. Note

that $a_2$ must be a shape in $\mathcal{L}$ and $M_{a_2}$ is a scalar which is negligible due to Lemma A.3. Moreover, $\lambda_{(a_1,a_2)} = \lambda_{a_1}\lambda_{a_2}$. The matrix $M_a$ can be written as $M_a = M_{a_1}M_{a_2} - \mathcal{E}_{\text{collapse}(a_1,a_2)}$, where $\mathcal{E}_{\text{collapse}(a_1,a_2)}$ consists of shapes obtained by collapsing $a_1, a_2$. Now summing up all possible $a_2$'s for a fixed $a_1$, we get

$$\lambda_{a_1}M_{a_1} + \sum_{a_2}\lambda_{(a_1,a_2)}M_{(a_1,a_2)} = \lambda_{a_1}M_{a_1}\left(1 + \sum_{a_2}\lambda_{a_2}M_{a_2}\right) - \lambda_{a_1}\sum_{a_2}\lambda_{a_2}\mathcal{E}_{\text{collapse}(a_1,a_2)}$$

Observe that $1 + \sum_{a_2}\lambda_{a_2}M_{a_2}$ is simply $\mathcal{M}_{00}$! The same procedure can be done for shapes $(a_1, a_2, a_3^\top)$ where $a_1, a_3$ are left one-sided shapes. This essentially shows that all disconnected components can be absorbed into the shape without that component.

**Lemma A.11** (Restatement of Lemma 6.22)**.** *Consider the first column of $\mathcal{M}$: $(\mathcal{M}_{00}, 0, \mathcal{M}_{20})$, and let $v := (1, 0, \frac{\mathcal{M}_{20}}{\mathcal{M}_{00}})$. The matrix $\mathcal{M}_{00} \cdot vv^\top$ captures all disconnected shapes in $\mathcal{M}_{22}$ modulo some error consisting of negligible shapes.*

*Proof.* Let $\mathcal{L}_{\text{left}}, \mathcal{L}_{\text{right}}$ be the set of left and right one-sided shapes in $\mathcal{L}$. Also, note that $\mathcal{M}_{00} = 1 + o(1)$. From the discussion above, we can write the vector

$$\mathcal{M}_{20} = \mathcal{M}_{00}\sum_{a\in\mathcal{L}_{\text{left}}}\lambda_a M_a + \sum_{a\in\mathcal{L}_{\text{left}}}\sum_{b\in\mathcal{L}}\lambda_a\lambda_b\mathcal{E}_{\text{collapse}(a,b)}\,.$$

Similarly, the sum of disconnected shapes in $\mathcal{M}_{22}$ can be written this way:

$$\mathcal{M}_{00}\sum_{a_1,a_2\in\mathcal{L}_{\text{left}}}\lambda_{a_1}\lambda_{a_2}M_{a_1,a_2} + \sum_{a_1,a_2\in\mathcal{L}_{\text{left}}}\sum_{b\in\mathcal{L}}\lambda_{a_1}\lambda_{a_2}\lambda_b\mathcal{E}_{\text{collapse}((a_1,a_2),b)}\,,$$

where the first term

$$\sum_{a_1,a_2\in\mathcal{L}_{\text{left}}}\lambda_{a_1}\lambda_{a_2}M_{a_1,a_2} = \left(\sum_{a\in\mathcal{L}_{\text{left}}}\lambda_a M_a\right)\left(\sum_{a\in\mathcal{L}_{\text{left}}}\lambda_a M_a\right)^\top + \mathcal{E}_{\text{collapsed}}$$

with $\mathcal{E}_{\text{collapsed}}$ consists of negligible collapsed shapes due to Lemma 6.21.

Then, consider the first column of $\mathcal{M}$: $(\mathcal{M}_{00}, 0, \mathcal{M}_{20})$, and let $v := (1, 0, \frac{\mathcal{M}_{20}}{\mathcal{M}_{00}})$. Clearly, the matrix $\mathcal{M}_{00} \cdot vv^\top$ captures all disconnected shapes in $\mathcal{M}_{22}$ modulo some negligible collapsed shapes. □

## A.5 Truncation error is small

We first prove that the candidate moment matrix $\mathcal{M}$ given by the pseudo-calibration method already approximately satisfies the constraints $x^\top G_s x = 0$ with very small error. Specifically, we show that $\widetilde{\mathbb{E}}[x^I(x^\top G_s x)]$ is close to 0 for any $I \subseteq [n]$, $|I| \leqslant 2$.

Adopting the notation of [GJJ+20], we use $Q\widetilde{\mathbb{E}}$ to denote the error, where $\widetilde{\mathbb{E}}$ is treated as a dimension $\binom{n}{\leqslant 4}$ vector, $Q$ is a matrix with rows indexed by $(I, s)$ for $I \subseteq [n], s \in [m]$, and

$$Q\widetilde{\mathbb{E}}(I, s) := \widetilde{\mathbb{E}}\left[x^I(x^\top G_s x)\right].$$

Note that we only work with degree-4 SoS, so $|I| \leqslant 2$.

**Lemma A.12** (Restatement of Lemma 6.23)**.** *There exist constants $C, C_1, c_2, c_3 > 0$ such that if $\varepsilon \geqslant \frac{C \log\log n}{\log n}$ and $\frac{C_1}{\varepsilon} \leqslant \tau \leqslant n^{c_2\varepsilon}$, then $\|Q\widetilde{\mathbb{E}}\|_2 \leqslant n^{-c_3\varepsilon\tau}$.*

*Proof.* By Definition 2.9,

$$\begin{aligned}\widetilde{\mathbb{E}}\left[x^I(x^\top G_s x)\right] &= \sum_{i,j\in[n]} G^s_{ij}\widetilde{\mathbb{E}}\left[x^I x_i x_j\right] \\ &= \sum_{i,j\in[n]}\sum_{\alpha:|\alpha|\leqslant\tau} G^s_{ij}\cdot h_\alpha(G)\cdot\frac{1}{\alpha!}\mathbb{E}_{(G',z)\sim\nu_P}\left[z^I z_i z_j h_\alpha(G')\right].\end{aligned} \tag{18}$$

Next, using the recurrence of Hermite polynomials, we have $xh_k(x) = h_{k+1}(x) + kh_{k-1}(x)$ for all $k \in \mathbb{N}$ (assuming $h_{-1}(x) = 0$). For simplicity of presentation, let us single out an entry $(s,i,j)$ and let $k := \alpha^s_{ij}$. We look at the terms $h_k(x)$ and $h_{k+2}(x)$,

$$\begin{aligned}&\frac{1}{k!}x\cdot h_k(x)h_k(x') + \frac{1}{(k+2)!}x\cdot h_{k+2}(x)h_{k+2}(x') \\ =\ &\frac{1}{k!}\left(h_{k+1}(x)+kh_{k-1}(x)\right)h_k(x') + \frac{1}{(k+2)!}\left(h_{k+3}(x)+(k+2)h_{k+1}(x)\right)h_{k+2}(x').\end{aligned}$$

Now, the coefficient of $h_{k+1}(x)$ is

$$\frac{1}{(k+1)!}h_{k+1}(x)\cdot\left((k+1)h_k(x')+h_{k+2}(x')\right) = \frac{1}{(k+1)!}h_{k+1}(x)\cdot x'h_{k+1}(x'),$$

again using the recurrence of Hermite polynomials. Intuitively, this allows us to rewrite a sum of $xh_k(x)h_k(x')$ as a sum of $x'h_k(x')h_k(x)$. Thus, (18) can be rewritten as

$$\begin{aligned}Q\widetilde{\mathbb{E}}(I,s) := \widetilde{\mathbb{E}}\left[x^I(x^\top G_s x)\right] &= \sum_{\beta:|\beta|\leqslant\tau-1}\frac{h_\beta(G)}{\beta!}\sum_{i,j\in[n]}\mathbb{E}_{(G',z)\sim\nu_P}\left[z^I z_i z_j G'^s_{ij}h_\beta(G')\right] + \varepsilon_\tau(I,s) \\ &= \sum_{\beta:|\beta|\leqslant\tau-1}\frac{h_\beta(G)}{\beta!}\mathbb{E}_{(G',z)\sim\nu_P}\left[z^I h_\beta(G')(zG'^s z^\top)\right] + \varepsilon_\tau(I,s) \\ &= \varepsilon_\tau(I,s),\end{aligned}$$

here we see the importance of the planted distribution: any $(G',z)\sim\nu_P$ satisfies $zG'^s z^\top = 0$.

Finally, we analyze the remaining error term $\varepsilon_\tau(I,s)$. Denote $\alpha_{+sij}\in\mathbb{N}^{m\times n\times n}$ as the index $\alpha$ with the entry $\alpha^s_{ij}$ incremented by 1. Since $|\alpha|$ must be even, $|\beta|$ must be odd, and the error only consists of terms $h_\beta(G)$ where $\beta = \alpha_{+sij}$ and $|\alpha| = \tau$.

$$\varepsilon_\tau(I,s) = \sum_{\alpha:|\alpha|=\tau}\sum_{i,j\in[n]}\frac{h_{\alpha_{+sij}}(G)}{\alpha!}\cdot\mathbb{E}_{\nu_P}\left[z^{I+\{i,j\}}h_\alpha(G')\right] = \sum_{\alpha:|\alpha|=\tau}\sum_{i,j\in[n]}\lambda_{\alpha,I,\{i,j\}}h_{\alpha_{+sij}}(G). \tag{19}$$

First, the magnitude of $\lambda_{\alpha,I,\{i,j\}}$ (recall equation (13)) can be upper bounded by

$$|\lambda_{\alpha,I,\{i,j\}}| \leqslant n^{-|\alpha|+|I|/2+1}(|\alpha|-1)!! \leqslant n^{-\tau+2}\tau^{\tau/2}$$

here we use the fact that $|I|\leqslant 2$, $|\alpha| = \tau$, and $(2k-1)!! \leqslant (2k)^k$.

Next, fix $I, s, i, j$. The quantity $\sum_{|\alpha|=\tau} \lambda_{\alpha,I+\{i,j\}} h_{\alpha_{+sij}}(G)$ is a sum of graph matrices over shapes with $\tau$ edges. By Lemma A.4, there are at most $\tau^{O(\tau)}$ such shapes. Since $(\alpha, I, \{i,j\})$ must satisfy the conditions in Definition 6.11 so that $\lambda_{\alpha,I,\{i,j\}}$ is nonzero, we can use Lemma A.3 to upper bound

$$\left| \sum_{|\alpha|=\tau} \lambda_{\alpha,I+\{i,j\}} h_{\alpha_{+sij}}(G) \right| \leqslant n^{-\frac{\varepsilon\tau}{4}+O(1)} \tau^{O(\tau)} .$$

Summing over all $i, j$, we get $|\varepsilon_\tau(I,s)| \leqslant n^{-\Omega(\varepsilon\tau)+O(1)}$ if $\tau \leqslant n^{c_2\varepsilon}$ for a small enough constant $c_2$. Moreover, if $\tau \geqslant \frac{C_1}{\varepsilon}$ for a large enough constant $C_1$, then $\|\varepsilon_\tau\|_2 \leqslant n^{-\Omega(\varepsilon\tau)}$. This requires $\varepsilon \geqslant \frac{C \log\log n}{\log n}$ for some constant $C$. This completes the proof. □

We remark that a result similar to Lemma 6.23 can be also obtained using [GJJ+20, Lemma 7.7]. In general, due to the pseudo-calibration method, if the truncation threshold $\tau$ is not too small, then the candidate moment matrix already approximately satisfies all constraints with tiny error.

### A.6 Bounds on the norm and nonzero singular values of $Q$

Since $Q\widetilde{\mathbb{E}} = 0$ if and only if $\widetilde{\mathbb{E}}$ exactly satisfies all constraints, the natural "fix" is

$$\widetilde{\mathbb{E}}_{\text{fix}} := \widetilde{\mathbb{E}} - Q^\top (QQ^\top)^\dagger Q\widetilde{\mathbb{E}} .$$

$(QQ^\top)^\dagger$ is the *pseudo inverse*. Clearly, $Q\widetilde{\mathbb{E}}_{\text{fix}} = 0$.

We assume that $\widetilde{\mathbb{E}}$ only contains the even degree monomials since the odd monomials are zero and don't need to be fixed. Moreover, we assume that the $G_s$'s are symmetrized so that $G^s_{ij} = G^s_{ji}$; this has no effect on the results and will greatly simplify the presentation. The entries $G^s_{ij}$ and $G^s_{ii}$ will thus have different scaling, but this is only a constant factor difference.

Recall that the rows of $Q$ are indexed by $(s, I)$ where $|I| = 0$ or 2.

**$|I| = 0$ case.** We first look at the entries of $Q\widetilde{\mathbb{E}}$ corresponding to $I = \varnothing$:

$$Q\widetilde{\mathbb{E}}(\varnothing, s) = \widetilde{\mathbb{E}}[x^\top G_s x] = 2\sum_{i<j} G^s_{ij} \widetilde{\mathbb{E}}[x_i x_j] + \frac{1}{n} \sum_{i\in[n]} G^s_{ii} .$$

Here we use $x_i^2 = \frac{1}{n}$. We can see that this is same as the analysis in Section A.3, and the above can be represented by the matrix $L_2$ (see Figure 6).

**Lemma A.13.** *$\|L_2\| = \widetilde{O}(n)$ and $L_2 L_2^\top$ has minimum eigenvalue $\Omega(n^2)$.*

*Proof.* $\|L_2\| = \widetilde{O}(n)$ is immediate from graph matrix norm bounds. For the minimum singular value, observe that since $m \ll n^2$, $L_2$ is a dense rectangular matrix and every entry is independent: $L_2((s,\varnothing),\{i,j\}) = G^s_{ij}$. Standard techniques in random matrix theory (such as an $\varepsilon$-net argument) show that $L_2 L_2^\top$ is full rank and has minimum eigenvalue $\Omega(n^2)$ with high probability. □

$|I| = 2$ **case.** Suppose $I = \{k, \ell\}$ with $k \neq \ell$, we have

$$Q\widetilde{\mathbb{E}}(I, s) = \widetilde{\mathbb{E}}[x^I (x^\top G_s x)] = 2 \sum_{i<j: i\neq j\neq k\neq \ell} G^s_{ij} \widetilde{\mathbb{E}}[x_i x_j x_k x_\ell] + \frac{1}{n} \sum_{\substack{j: i\neq j\neq \ell \\ i=k}} G^s_{kj} \widetilde{\mathbb{E}}[x_j x_\ell]$$
$$+ \frac{1}{n} \sum_{i=j\neq k\neq \ell} G^s_{ii} \widetilde{\mathbb{E}}[x_k x_\ell] + \cdots$$

The expansion corresponds to the shapes in Figure 9 (first 3 terms are drawn out). We denote the sum as $L_4$.

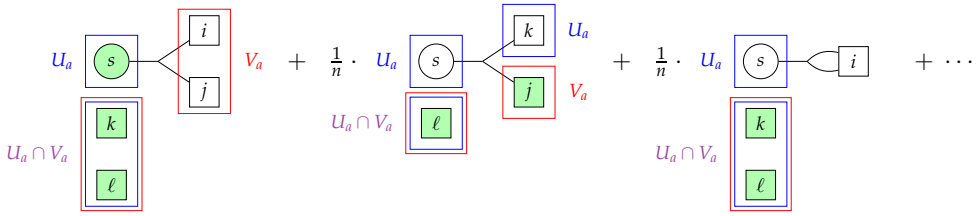



Figure 9: Shapes in $L_4$.

Before we dive into the analysis, we first define a shape $a^*$ drawn in Figure 10. This shape appears in the expansion of $L_4 L_4^\top$ and will play a crucial role in our analysis.

**Definition A.14** (Shape $a^*$)**.** *We define $a^*$ as the shape drawn in Figure 10. The matrix $M_{a^*}$ has entries*

$$M_{a^*}(\{s_1, i_1, j_1\}, \{s_2, i_2, j_2\}) = G^{s_1}_{i_2 j_2} G^{s_2}_{i_1 j_1},$$

*if $s_1 \neq s_2$ and $i_1 \neq j_1 \neq i_2 \neq j_2$, and 0 otherwise.*

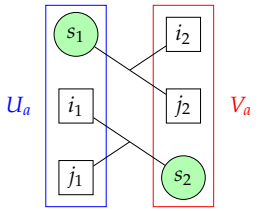



Figure 10: Special shape $a^*$. $\|M_{a^*}\| = \widetilde{O}(n^2)$.

**Analysis of $L_4 L_4^\top$.** Let $a_1, a_2, a_3$ be the first three shapes in $L_4$ drawn in Figure 9. $\|M_{a_1}\| = \widetilde{O}(n)$, $\|\frac{1}{n} M_{a_2}\| \leqslant \widetilde{O}(n^{\frac{1}{2}-\frac{\epsilon}{2}})$, $\|\frac{1}{n} M_{a_3}\| = \widetilde{O}(n^{\frac{1}{2}-\frac{\epsilon}{2}})$, and the rest of the terms have norm $o(1)$. Thus, our analysis will focus on $M_{a_1}$. We show the following lemma,

**Lemma A.15.** *There exists a matrix $A_1$ such that*

$$L_4 L_4^\top = \Theta(n^2) \cdot \mathbb{I} + A_1 A_1^\top + M_{a^*} + \mathcal{E}_2$$

*where $\|\mathcal{E}_2\| = \widetilde{O}(n^{2-\frac{\epsilon}{2}})$.*

*Proof.* First, $L_4 L_4^\top$ can be written as

$$L_4 L_4^\top = M_{a_1} M_{a_1}^\top + \mathcal{E}_1$$

where $\|\mathcal{E}_1\| = \widetilde{O}(n^{\frac{3}{2}-\frac{\varepsilon}{2}})$. Thus, it suffices to analyze the matrix $M_{a_1} M_{a_1}^\top$.

Let us write out the entries of $M_{a_1}$ explicitly: $M_{a_1}(\{s,k',\ell'\},\{i,j,k,\ell\}) = G^s_{ij}$ if $\{k',\ell'\} = \{k,\ell\}$ and 0 otherwise. In other words, it is nonzero only when $\{k',\ell'\} \subset \{i,j,k,\ell\}$. Then, we can write the entries of $M_{a_1} M_{a_1}^\top$ explicitly,

$$(M_{a_1} M_{a_1}^\top)(\{s_1,i_1,j_1\},\{s_2,i_2,j_2\}) = \begin{cases} \sum_{k\neq\ell\neq i_1\neq j_1} G^{s_1}_{k\ell} G^{s_2}_{k\ell} & \text{if } \{i_1,j_1\} = \{i_2,j_2\}, \\ \sum_{k\notin\{i_1,i_2,j_1\}} G^{s_1}_{i_2 k} G^{s_2}_{i_1 k} & \text{if } j_1 = j_2 \neq i_1 \neq i_2, \\ G^{s_1}_{i_2 j_2} G^{s_2}_{i_1 j_1} & \text{if } i_1 \neq i_2 \neq j_1 \neq j_2. \end{cases}$$

The above can be represented as a sum of several graph matrices. We split into different cases; each case corresponds to a shape:

- $s_1 = s_2$ (diagonal blocks of $M_{a_1} M_{a_1}^\top$):
  - Case $\{i_1,j_1\} = \{i_2,j_2\}$: for this shape there is an identity component, $M_a = \Theta(n^2)\cdot\mathbb{I} + \mathcal{E}$ where the error $\|\mathcal{E}\| = \widetilde{O}(n)$.
  - Case $j_1 = j_2 \neq i_1 \neq i_2$: for this shape $\|M_a\| = \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$.
  - Case $i_1 \neq i_2 \neq j_1 \neq j_2$: this shape can be decomposed into a PSD component plus some errors: $M_a = A_1 A_1^\top + \mathcal{E}$ where $\|\mathcal{E}\| = \widetilde{O}(n)$.
- $s_1 \neq s_2$ (off-diagonal blocks of $M_{a_1} M_{a_1}^\top$):
  - Case $\{i_1,j_1\} = \{i_2,j_2\}$: for this shape $\|M_a\| = \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$.
  - Case $j_1 = j_2 \neq i_1 \neq i_2$: for this shape $\|M_a\| = \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$.
  - Case $i_1 \neq i_2 \neq j_1 \neq j_2$: this shape is exactly $a^*$ in Figure 10.

Therefore, we can write

$$M_{a_1} M_{a_1}^\top = \Theta(n^2)\cdot\mathbb{I} + A_1 A_1^\top + M_{a^*} + \mathcal{E}_2$$

where $\|\mathcal{E}_2\| = \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$. This completes the proof. □

**Remark A.16.** We will later show that $L_4 L_4^\top$ has a non-trivial null space. Thus, the shape $a^*$ must exist in the expansion of $M_{a_1} M_{a_1}^\top$; without it, $L_4 L_4^\top$ would be full rank, which is a contradiction.

**Null space of $L_4 L_4^\top$.** Consider any $\widetilde{\mathbb{E}}$ and fix $s_1 < s_2 \in [m]$. Observe that

$$\widetilde{\mathbb{E}} \sum_{k,\ell} (x^\top G_{s_1} x) G^{s_2}_{k\ell} x_k x_\ell - \widetilde{\mathbb{E}} \sum_{k,\ell} (x^\top G_{s_2} x) G^{s_1}_{k\ell} x_k x_\ell = 0\,.$$

Treating $\widetilde{\mathbb{E}}$ as a vector, this can be written as $\widetilde{\mathbb{E}}^\top L_4^\top N_{s_1,s_2} = 0$. Since this holds for all vectors $\widetilde{\mathbb{E}}$, $N_{s_1,s_2}$ is in the null space of $L_4^\top$. Collecting the vectors for all pairs $s_1 < s_2$, we get a matrix $N$ such that $L_4^\top N = 0$.

Similar to the analysis of $L_4$, we look at the dominating component $M_{b_1}$ of $N$; $M_{b_1}$ has norm $\widetilde{O}(n)$ whereas the other term has norm $\widetilde{O}(n^{\frac{1}{2}-\frac{\varepsilon}{2}})$. The rows of $M_{b_1}$ are indexed by $\{s,i,j\}$ and the columns are indexed by $\{s_1,s_2\}$:

$$M_{b_1}(\{s,i,j\},\{s_1,s_2\}) = \begin{cases} G_{ij}^{s_2} & \text{if } s = s_1, \\ -G_{ij}^{s_1} & \text{if } s = s_2, \\ 0 & \text{otherwise.} \end{cases}$$

Next, we prove the following result for $NN^\top$,

**Lemma A.17.** *There exists a matrix $A_2$ such that*

$$NN^\top = A_2A_2^\top - M_{a^*} + \mathcal{E}_3$$

*where $\|\mathcal{E}_3\| \leqslant \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$.*

*Proof.* It suffices to consider $M_{b_1}M_{b_1}^\top$.

$$\left(M_{b_1}M_{b_1}^\top\right)(\{s_1,i_1,j_1\},\{s_2,i_2,j_2\}) = \begin{cases} -G_{i_1j_1}^{s_2}G_{i_2j_2}^{s_1} & \text{if } s_1 \neq s_2 \\ \sum_{s_3\neq s_1} G_{i_1j_1}^{s_3}G_{i_2j_2}^{s_3} & \text{if } s_1 = s_2 \end{cases}$$

We can also write $M_{b_1}M_{b_1}^\top$ as a sum of graph matrices:

- $s = s_1 = s_2$ (diagonal blocks of $M_{b_1}M_{b_1}^\top$): it is clear that $\sum_{s_3\neq s} G_{i_1j_1}^{s_3}G_{i_2j_2}^{s_3}$ is a PSD component. Thus, we can write this component as $A_2A_2^\top$ for some matrix $A_2$.
- $s_1 \neq s_2$ (off-diagonal blocks of $M_{b_1}M_{b_1}^\top$):
  - Case $i_1 \neq i_2 \neq j_1 \neq j_2$: this shape is exactly $a^*$ but with a crucial negative sign.
  - Other cases: these shapes have norms bounded by $\widetilde{O}(n^{2-\frac{\varepsilon}{2}})$.

Thus, we have

$$NN^\top = A_2A_2^\top - M_{a^*} + \mathcal{E}_3$$

where $\|\mathcal{E}_3\| \leqslant \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$. □

**Proof of Lemma 6.24.** Combining Lemma A.15 and Lemma A.17, we see that the term $M_{a^*}$ cancels out. This implies that $L_4L_4^\top + NN^\top$ is full rank and has minimum eigenvalue $\Omega(n^2)$. Now, we are ready to prove Lemma 6.24.

**Lemma A.18** (Restatement of Lemma 6.24)**.** *There exists a constant $C$ such that for $\varepsilon \geqslant \frac{C\log\log n}{\log n}$, $\|Q\| \leqslant \widetilde{O}(n)$ and the smallest nonzero eigenvalue of $QQ^\top$ is $\Omega(n^2)$.*

*Proof.* $Q = L_2 + L_4$. By the graph matrix norm bounds, we have $\|Q\| \leqslant \widetilde{O}(n)$.

Next, we lower bound the minimum eigenvalue of $QQ^\top$. Observe that

$$QQ^\top = \begin{bmatrix} L_2L_2^\top & L_2L_4^\top \\ L_4L_2^\top & L_4L_4^\top \end{bmatrix}.$$

For $L_2 L_2^\top$, Lemma A.13 shows that it has minimum eigenvalue $\Omega(n^2)$.

For $L_4 L_4^\top$, by Lemma A.15 and Lemma A.17 we have

$$L_4 L_4^\top + NN^\top = \Theta(n^2) \cdot \mathbb{I} + A_1 A_1^\top + A_2 A_2^\top + \mathcal{E}_4 \,,$$

where $\|\mathcal{E}_4\| \leqslant \widetilde{O}(n^{2-\frac{\varepsilon}{2}})$. This means that $L_4 L_4^\top + NN^\top$ is full rank and has minimum eigenvalue $\Omega(n^2)$.

For the off-diagonal block $L_2 L_4^\top$, although both $\|L_2\|$ and $\|L_4\| = \widetilde{O}(n)$, note that $L_2$ and $M_{a_1}$ (the dominating component of $L_4$) have disjoint rows and columns in $Q$, meaning that $L_2 M_{a_1}^\top$ does not contribute to $L_2 L_4^\top$. Then, since $\|L_4 - M_{a_1}\| \leqslant \widetilde{O}(n^{\frac{1}{2}-\frac{\varepsilon}{2}})$, we have $\|L_2 L_4^\top\| \leqslant \widetilde{O}(n^{\frac{3}{2}-\frac{\varepsilon}{2}})$.

We have shown that $QQ^\top$ plus an orthogonal matrix is full rank and has minimum eigenvalue $\Omega(n^2)$. This implies that the minimum nonzero eigenvalue of $QQ^\top$ is $\Omega(n^2)$. This completes the proof. □