Sum-of-Squares Lower Bounds for Sherrington-Kirkpatrick via Planted Affine Planes

Mrinalkanti Ghosh* Fernando Granha Jeronimo† Chris Jones‡
Aaron Potechin§ Goutham Rajendran¶

September 7, 2020

Abstract

The Sum-of-Squares (SoS) hierarchy is a semi-definite programming meta-algorithm that captures state-of-the-art polynomial time guarantees for many optimization problems such as Max-k-CSPs and Tensor PCA. On the flip side, a SoS lower bound provides evidence of hardness, which is particularly relevant to average-case problems for which NP-hardness may not be available.

In this paper, we consider the following average case problem, which we call the *Planted Affine Planes* (PAP) problem: Given m random vectors $d_1, \ldots, d_m$ in $\mathbb{R}^n$, can we prove that there is no vector $v \in \mathbb{R}^n$ such that for all $u \in [m]$, $\langle v, d_u \rangle^2 = 1$? In other words, can we prove that m random vectors are not all contained in two parallel hyperplanes at equal distance from the origin? We prove that for $m \leq n^{3/2-\varepsilon}$, with high probability, degree-$n^{\Omega(\varepsilon)}$ SoS fails to refute the existence of such a vector v.

When the vectors $d_1, \ldots, d_m$ are chosen from the multivariate normal distribution, the PAP problem is equivalent to the problem of proving that a random n-dimensional subspace of $\mathbb{R}^m$ does not contain a boolean vector. As shown by Mohanty–Raghavendra–Xu [STOC 2020], a lower bound for this problem implies a lower bound for the problem of certifying energy upper bounds on the Sherrington-Kirkpatrick Hamiltonian, and so our lower bound implies a degree-$n^{\Omega(\varepsilon)}$ SoS lower bound for the certification version of the Sherrington-Kirkpatrick problem.

*TTIC. mkghosh@ttic.edu. Supported in part by NSF grant CCF-1816372.
†University of Chicago. granha@uchicago.edu. Supported in part by NSF grant CCF-1816372.
‡University of Chicago. csj@uchicago.edu. Supported in part by NSF grant CCF-2008920.
§University of Chicago. potechin@uchicago.edu. Supported in part by NSF grant CCF-2008920.
¶University of Chicago. goutham@uchicago.edu. Supported in part by NSF grant CCF-1816372.

Contents

1 Introduction

The Sum-of-Squares (SoS) hierarchy is a semi-definite programming (SDP) hierarchy which provides a meta-algorithm for polynomial optimization [Las15]. Given a polynomial objective function and a system of polynomial equalities and inequalities as constraints, the SoS framework specifies a family of increasingly "larger" SDP programs, where each program provides a convex relaxation to the polynomial optimization problem. The family is indexed by a size parameter D called the SoS degree. Roughly speaking, the larger the SoS degree D, the tighter the relaxation, but also, the greater the computational time required to solve the convex program, with $D = O(1)$ corresponding to polynomial time and $D = n$ able to exactly solve an optimization problem on n boolean variables. Due to the versatility of polynomials in modeling computational problems, the SoS hierarchy can be applied to a vast range of optimization problems. It has been shown to be quite successful in this regard, as it captures state-of-the-art approximation guarantees for many problems such as Sparsest Cut [ARV04], MaxCut [GW95], Tensor PCA [HSS15] and all Max-k-CSPs [Rag08].

The success of SoS for optimization confers on it an important role as an algorithmic tool. For this reason, on the flip side, understanding the degree range for which SoS *fails* to provide a desired guarantee to a computational problem can be useful to the algorithm designer in two ways. Firstly and more concretely, since SoS is a proof system capturing a broad class of algorithmic reasoning [FKP19], an SoS lower bound can inform the algorithm designer not only of the minimum degree required within the SoS hierarchy, but also to avoid methods of proof that are captured by low-degree SoS reasoning. Secondly, an SoS lower bound can serve as strong evidence for computational hardness [HKP$^+$17, Hop18], even though it is not a formal guarantee against all types of algorithms. This hardness evidence is particularly relevant to average-case problems for which we do not have NP-hardness (see, e.g., the SoS lower bound on the Planted Clique problem [BHK$^+$16]).

Our main results concern the performance of SoS on the following basic optimization problem

$$\mathsf{OPT}(W) := \max_{x \in \{\pm 1\}^n} x^\intercal W x, \tag{1}$$

where W is a symmetric matrix in $\mathbb{R}^{n \times n}$. This problem arises in the fields of computer science and statistical physics, though the terminology can sometimes differ. Computer scientists might regard $x \in \{\pm 1\}^n$ as encoding a bipartition of $[n] = \{1, 2, \ldots, n\}$. Note that by taking W to be a graph Laplacian [HLW06, Section 4] the problem is equivalent to the MaxCut problem, a well-known NP-hard problem in the worst case [Kar72]. A statistical physicist might regard x as encoding spin values in a spin-glass model. The matrix $-W$ is regarded as the *Hamiltonian* of the underlying physical system, where entry $-W_{i,j}$ models the interaction between spin x_i and x_j (with $-W_{i,j} \geq 0$ being ferromagnetic and $-W_{i,j} < 0$ being anti-ferromagnetic). Then, the optimized x corresponds to the minimum-energy, or ground, state of the system.

Instead of considering $\mathsf{OPT}(W)$ for a worst-case W, one can consider the average-case problem in which W is sampled according to some distribution. One of the simplest models of random matrices is the Gaussian Orthogonal Ensemble, denoted $\mathrm{GOE}(n)$ for n-by-n matrices and defined as follows.

Definition 1.1. *The Gaussian Orthogonal Ensemble, denoted* $\mathrm{GOE}(n)$*, is the distribution of* $\frac{1}{\sqrt{2}}(A + A^\intercal)$ *where* A *is a random* $n \times n$ *matrix with i.i.d. standard Gaussian entries.*

Taking $W \sim \mathrm{GOE}(n)$ for the optimization problem $\mathsf{OPT}(W)$ of Eq. (1) gives rise to the so-

called Sherrington–Kirkpatrick (SK) Hamiltonian [SK75]. Note that $\mathrm{GOE}(n)$ is a particular kind of Wigner matrix ensemble, thereby satisfying the semicircle law, which in this case establishes that the largest eigenvalue of W is $(2+\mathrm{o}_n(1))\cdot\sqrt{n}$ with probability $1-\mathrm{o}_n(1)$. Thus, a trivial spectral bound establishes $\mathsf{OPT}(W) \le (2+\mathrm{o}_n(1))\cdot n^{3/2}$ with probability $1-\mathrm{o}_n(1)$. However, in a foundational work based on a variational argument [Par79], Parisi conjectured that

$$\mathop{\mathbb{E}}_{W\sim\mathrm{GOE}(n)}[\mathsf{OPT}(W)] \approx 2\cdot P^*\cdot n^{3/2},$$

where $P^* \approx 0.7632$ is now referred to as the Parisi constant. In a breakthrough result, Talagrand [Tal06] gave a rigorous proof of Parisi's conjecture [1]. The question then became, "is there a polynomial-time algorithm that given $W\sim\mathrm{GOE}(n)$ computes an x achieving close to $\mathsf{OPT}(W)$?" As it turns out, the answer was essentially shown to be yes by Montanari [Mon19]!

The natural question we study is that of certification: "is there an efficient algorithm to certify an upper bound on $\mathsf{OPT}(W)$ for any input W, that improves upon the trivial spectral bound?" In particular, we can ask how well SoS does as a certification algorithm. The natural upper bound of $(2+\mathrm{o}_n(1))\cdot n^{3/2}$ obtained via the spectral norm of W is also the value of the degree-2 SoS relaxation [MS16]. Two independent recent works of Mohanty–Raghavendra–Xu [MRX19] and Kunisky–Bandeira [KB19] show that degree-4 SoS does not perform much better, and a heuristic argument from [BKW19] suggests that even degree-$(n/\log n)$ SoS cannot certify anything stronger than the trivial spectral bound. Thus we ask,

Can higher-degree SoS certify better upper bounds for the Sherrington–Kirkpatrick problem, hopefully closer to the true bound $2\cdot P^*\cdot n^{3/2}$*?*

Our Results. We answer the question above negatively by showing that even at degree as large as n^δ, SoS cannot improve upon the basic spectral algorithm. More precisely, we have the following theorem which is our first main result and our most important contribution.

Theorem 1.2. *[Main I] There exists a constant $\delta > 0$ such that, w.h.p. for $W\sim\mathrm{GOE}(n)$, there is a degree-n^δ SoS solution for the Sherrington–Kirkpatrick problem with value at least $(2-\mathrm{o}_n(1))\cdot n^{3/2}$.*

In light of the result of Montanari [Mon19], the situation is intriguing. Montanari showed that for all $\varepsilon > 0$, there is a $\mathrm{O}_\varepsilon(n^2)$ time randomized algorithm that given a random W drawn from the Gaussian Orthogonal Ensemble, outputs an x such that $x^TWx \ge (1-\varepsilon)\mathsf{OPT}(W)$. The correctness of the algorithm assumes a widely-believed conjecture from statistical physics known as the full replica symmetry breaking assumption. However, we show an integrality gap for SoS.

Based on this, it is an interesting question whether SoS, together with an appropriate rounding scheme, is optimal for the Sherrington-Kirkpatrick problem. On the one hand, the situation could be similar to the Feige-Schechtman integrality gap instance for MaxCut [FS02]. For the Feige-Schechtman integrality gap instance, SoS fails to certify the value of the optimal solution. However, applying hyperplane rounding to the SoS solution gives an almost-optimal solution for these instances. It could be the case that there is a rounding scheme which takes an SoS solution for the Sherrington-Kirkpatrick problem on a random W and returns an almost optimal solution x. On the other hand, we currently don't know what this rounding scheme would be.

In order to prove Theorem 1.2, we first introduce a new average-case problem we call Planted Affine Planes (PAP) for which we directly prove a SoS lower bound. We then use the PAP lower

[1] The results of Talagrand [Tal06] were for the Sherrington–Kirkpatrick and mixed p-spin systems with p even. In [Pan14], Panchenko generalized these results to arbitrary mixed p-spin systems (also including odd p).

bound to prove a lower bound on the Sherrington–Kirkpatrick problem. The PAP problem can be informally described as follows (see Definition 2.1 for the formal definition).

Definition 1.3 (Informal statement of PAP)**.** *Given m random vectors $d_1, \dots, d_m$ in $\mathbb{R}^n$, can we prove that there is no vector $v \in \mathbb{R}^n$ such that for all $u \in [m]$, $\langle v, d_u \rangle^2 = 1$? In other words, can we prove that m random vectors are not all contained in two parallel hyperplanes at equal distance from the origin?*

This problem, when we restrict v to a Boolean vector in $\{\pm\frac{1}{\sqrt{n}}\}^n$, can be encoded as the feasibility of the polynomial system

$$\exists v \in \mathbb{R}^n \text{ s.t.} \qquad \forall i \in [n],\ v_i^2 = \frac{1}{n},$$
$$\forall u \in [m],\ \langle v, d_u \rangle^2 = 1.$$

Hence it is a ripe candidate for SoS. However, we show that SoS fails to refute a random instance. The Boolean restriction on v actually makes the lower bound result stronger since SoS cannot refute even a smaller subset of vectors in $\mathbb{R}^n$. In this work, we will consider two different random distributions, namely when $d_1, \dots, d_m$ are independent samples from the multivariate normal distribution and when they are independent samples from the uniform distribution on the boolean hypercube.

Theorem 1.4 (Main II)**.** *For both the Gaussian and Boolean settings, there exists a constant $c > 0$ such that for all $\varepsilon > 0$ and $\delta \le c\varepsilon$, for $m \le n^{3/2-\varepsilon}$, w.h.p. there is a feasible degree-n^δ SoS solution for Planted Affine Planes.*

It turns out that the Planted Affine Plane problem introduced above is closely related to the following "Boolean vector in a random subspace" problem, which we call the Planted Boolean Vector problem, introduced by [MRX19] in the context of studying the performance of SoS on computing the Sherrington–Kirkpatrick Hamiltonian.

The Planted Boolean Vector problem is to certify that a random subspace of $\mathbb{R}^n$ is far from containing a boolean vector. Specifically, we want to certify an upper bound for

$$\mathsf{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b,$$

where V is a uniformly random p-dimensional subspace[2] of $\mathbb{R}^n$, and Π_V is the projector onto V. In brief, the relationship to the Planted Affine Plane problem is that the PAP vector v represents the coefficients on a linear combination for the vector b in the span of a basis of V.

An argument of [MRX19] shows that, when $p \ll n$, w.h.p., $\mathsf{OPT}(V) \approx \frac{2}{\pi}$, whereas they also show that w.h.p. assuming $p \ge n^{0.99}$, there is a degree-4 SoS solution with value $1 - o_n(1)$. They ask whether or not there is a polynomial time algorithm that can certify a tighter bound; we rule out SoS-based algorithms for a larger regime both in terms of SoS degree and the dimension p of the random subspace.

Theorem 1.5. *[Main III] There exists a constant $c > 0$ such that, for all $\varepsilon > 0$ and $\delta \le c\varepsilon$, for $p \ge n^{2/3+\varepsilon}$, w.h.p. over V there is a degree-n^δ SoS solution for Planted Boolean Vector of value 1.*

[2] V can be specified by a basis, which consists of p i.i.d. samples from $\mathcal{N}(0, I)$.

Our Approach. We now provide a brief high-level description of our approach (see Section 3 for a more detailed overview). The bulk of our technical contribution lies in the SoS lower bound for the Planted Affine Planes problem, Theorem 1.4. We then show that Planted Affine Planes in the Gaussian setting is equivalent to the Planted Boolean Vector problem. The reduction from Sherrington-Kirkpatrick to the Planted Boolean Vector problem is due to Mohanty–Raghavendra–Xu [MRX19].

As a starting point to the PAP lower bound, we employ the general techniques introduced by Barak et al. [BHK+16] for SoS lower bounds. We use their pseudocalibration machinery to produce a good candidate SoS solution $\widetilde{\mathbb{E}}$. The operator $\widetilde{\mathbb{E}}$ unfortunately does not exactly satisfy the PAP constraints "$\langle v, d_u\rangle^2 = 1$", it only satisfies them up to a tiny error. We use an interesting and rather generic approach to round $\widetilde{\mathbb{E}}$ to a nearby pseudoexpectation operator $\widetilde{\mathbb{E}}'$ which does exactly satisfy the constraints.

For degree D, the candidate SoS solution can be viewed as a (pseudo) moment matrix $\mathcal{M}$ with rows and columns indexed by subsets $I, J \subset [n]$ with size bounded by $D/2$ and with entries

$$\mathcal{M}[I, J] := \widetilde{\mathbb{E}}[v^I v^J].$$

The matrix $\mathcal{M}$ is a random function of the inputs $d_1, \ldots, d_m$, and the most challenging part of the analysis consists of showing that $\mathcal{M}$ is positive semi-definite (PSD) with high probability.

Similarly to [BHK+16], we decompose $\mathcal{M}$ as a linear combination of graph matrices, i.e., $\mathcal{M} = \sum_\alpha \lambda_\alpha \cdot M_\alpha$, where M_α is the graph matrix associated with shape α. In brief, each graph matrix aggregates all terms with shape α in the Fourier expansions of the entries of $\mathcal{M}$ – the shape α is informally a graph with labeled edges with size bounded by $\mathrm{poly}(D)$. A graph matrix decomposition of $\mathcal{M}$ is particularly handy in the PSD analysis since the operator norm of individual graph matrices M_α is (with high probability) determined by simple combinatorial properties of the graph α. One technical difference from [BHK+16] is that our graph matrices have two types of vertices □ and ◯; these graph matrices fall into the general framework developed by Ahn et al. in [AMP20].

To show that the matrix $\mathcal{M}$ is PSD, we need to study the graph matrices that appear with nonzero coefficients in the decomposition. The matrix $\mathcal{M}$ can be split into blocks and each diagonal block contains in the decomposition a (scaled) identity matrix. From the graph matrix perspective, this means that certain "trivial" shapes appear in the decomposition, with appropriate coefficients. If we could bound the norms of all other graph matrices that appear against these trivial shapes and show that, together, they have negligible norm compared to the sum of these scaled identity blocks, then we would be in good shape.

Unfortunately, this approach will not work. The kernel of the matrix $\mathcal{M}$ is nontrivial, as a consequence of satisfying the PAP constraints "$\langle v, d_u\rangle^2 = 1$", and hence there is no hope of showing that the contribution of all nontrivial shapes in the decomposition of $\mathcal{M}$ has small norm. Indeed, certain shapes α appearing in the decomposition of $\mathcal{M}$ are such that $\|\lambda_\alpha \cdot M_\alpha\|$ is large. As it turns out, all such shapes have a simple graphical substructure, and so we call these shapes *spiders*.

To get around the null space issue, we restrict ourselves to $\mathrm{Null}(\mathcal{M})^\perp$, which is the complement of the nullspace of $\mathcal{M}$. We show that the substructure present in a spider implies that the spider is close to the zero matrix in $\mathrm{Null}(\mathcal{M})^\perp$. Because of this, we can almost freely add and subtract M_α for spiders α while preserving the action of $\mathcal{M}$ on $\mathrm{Null}(\mathcal{M})^\perp$. Our strategy is to "kill" the spiders by subtracting off $\lambda_\alpha \cdot M_\alpha$ for each spider α. But because M_α is only approximately in $\mathrm{Null}(\mathcal{M})^\perp$, this strategy could potentially introduce new graph matrix terms, and in particular it could introduce new spiders. To handle this, we recursively kill them while carefully analyzing

how the coefficients of all the graph matrices change. After all spiders are killed, the resulting moment matrix becomes

$$\sum_{0\le k\le D/2} \frac{1}{n^k}\cdot I_k + \sum_{\gamma\colon \text{non-spiders}} \lambda'_\gamma \cdot M_\gamma,$$

for some new coefficients λ'_γ. Here, I_k is the matrix which has an identity in the kth block and the remaining entries 0. Using a novel charging argument, we finally show that the latter term is negligible compared to the former term, thus establishing $\mathcal{M} \succeq 0$.

Summary of Related Work and Our Contributions. We now summarize the existing work on these problems and our contributions. Degree-4 SoS lower bounds on the Sherrington-Kirkpatrick Hamiltonian problem were proved independently by Mohanty–Raghavendra–Xu [MRX19] and Kunisky–Bandeira [KB19] whereas we prove an improved degree-n^δ SoS lower bound for some constant $\delta > 0$. Our result is obtained by reducing the Sherrington-Kirkpatrick problem to the "Boolean Vector in a Random Subspace" problem which is equivalent to our new Planted Affine Planes problem on the normal distribution. The reduction from Sherrington-Kirkpatrick problem to the "Boolean Vector in a Random Subspace" is due to Mohanty–Raghavendra–Xu [MRX19]. The results of Mohanty–Raghavendra–Xu [MRX19] and Kunisky–Bandeira [KB19] build on a degree-2 SoS lower bounds of Montanari and Sen [MS16]. Regarding upper bounds, Montanari [Mon19] gave an efficient randomized message passing algorithm to estimate $\mathsf{OPT}(W)$ in the SK problem within a $(1-\varepsilon)$ factor under the full replica symmetry breaking assumption.

Degree-4 SoS lower bounds on the "Boolean Vector in a Random Subspace" problem for $p \ge n^{0.99}$ were proved by Mohanty–Raghavendra–Xu in [MRX19] where this problem was introduced. We improve the dependence on p to $p \ge n^{2/3+\varepsilon}$ for any $\varepsilon > 0$ and obtain a stronger degree-$n^{c\varepsilon}$ SoS lower bound for some absolute constant $c > 0$.

2 Technical Preliminaries

In this section we record problem statements, then define and discuss the main objects in our SoS lower bound: pseudoexpectation operators, the moment matrix, and graph matrices.

For a vector or variable $v \in \mathbb{R}^n$, and $I \subseteq [n]$, we use the notation $v^I := \prod_{i\in I} v_i$. When a statement holds with high probability (w.h.p.), it means it holds with probability $1 - o_n(1)$. In particular, there is no requirement for small n.

2.1 Problem statements

We introduce the Planted Affine Planes problem over a distribution $\mathcal{D}$.

Definition 2.1 (Planted Affine Planes (PAP) problem)**.** *Given $d_1, \dots, d_m \sim \mathcal{D}$ where each d_u is a vector in $\mathbb{R}^n$, determine whether there exists $v \in \{\pm\frac{1}{\sqrt{n}}\}^n$ such that*

$$\langle v, d_u\rangle^2 = 1,$$

for every $u \in [m]$.

Our results hold for the Gaussian setting $\mathcal{D} = \mathcal{N}(0, I)$ and the boolean setting where $\mathcal{D}$ is uniformly sampled from $\{\pm 1\}^n$, though we conjecture (Section 8) that similar SoS bounds hold under more general conditions on $\mathcal{D}$.

Observe that in both settings the solution vector v is restricted to be Boolean (in the sense that the entries are either $\frac{1}{\sqrt{n}}$ or $\frac{-1}{\sqrt{n}}$) and an SoS lower bound for this restricted version of the problem is stronger than when v can be an arbitrary vector from $\mathbb{R}^n$.

The Sherrington–Kirkpatrick (SK) problem comes from the spin-glass model in statistical physics [SK75].

Definition 2.2 (Sherrington-Kirkpatrick problem)**.** *Given $W \sim \mathrm{GOE}(n)$, compute*

$$\mathsf{OPT}(W) := \max_{x \in \{\pm 1\}^n} x^\intercal W x.$$

The Planted Boolean Vector problem was introduced by Mohanty–Raghavendra–Xu [MRX19], where it was called the "Boolean Vector in a Random Subspace".

Definition 2.3 (Planted Boolean Vector problem)**.** *Given a uniformly random p-dimensional subspace V of $\mathbb{R}^n$ in the form of a projector Π_V onto V, compute*

$$\mathsf{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b.$$

2.2 Sum-of-Squares solutions

We will work with two equivalent definitions of a degree-D SoS solution: a pseudoexpectation operator and a moment matrix. We tailor these definitions to our setting of feasibility of systems of polynomial equality constraints given by the common zero set of a collection of polynomials $\mathcal{P}$ on $\pm\frac{1}{\sqrt{n}}$ Boolean variables $v_1, \ldots, v_n$. For a degree-D solution to be well defined, we need D to be at least the maximum degree of a polynomial in $\mathcal{P}$. Let $\mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ be the subset of polynomials of degree at most D from the polynomial ring $\mathbb{R}(v_1, \ldots, v_n)$. We denote the degree of a polynomial $f \in \mathbb{R}(v_1, \ldots, v_n)$ by $\deg(f)$.

2.2.1 Pseudoexpectation operator

We formally define the pseudoexpectation operators used in our setting.

Definition 2.4 (Pseudoexpectation)**.** *Given a finite collection of "constraint" polynomials $\mathcal{P}$ of degree at most D on $\pm\frac{1}{\sqrt{n}}$ Boolean variables $v_1, \ldots, v_n$, a degree-D pseudoexpectation operator $\widetilde{\mathbb{E}}$ is an operator $\widetilde{\mathbb{E}} \colon \mathbb{R}^{\leq D}(v_1, \ldots, v_n) \to \mathbb{R}$ satisfying:*

1. $\widetilde{\mathbb{E}}[1] = 1$,
2. *$\widetilde{\mathbb{E}}$ is an $\mathbb{R}$-linear operator, i.e., $\widetilde{\mathbb{E}}[f + g] = \widetilde{\mathbb{E}}[f] + \widetilde{\mathbb{E}}[g]$ for every $f, g \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$,*
3. *$\widetilde{\mathbb{E}}[f^2] \geq 0$ for every $f \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with $\deg(f^2) \leq D$.*
4. *$\widetilde{\mathbb{E}}[(v_i^2 - \frac{1}{n}) \cdot f] = 0$ for all $i \in [n]$ and for every $f \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with $\deg(f) \leq D - 2$, and*
5. *$\widetilde{\mathbb{E}}[g \cdot f] = 0$ for every $g \in \mathcal{P}, f \in \mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with $\deg(f \cdot g) \leq D$.*

Note that $\widetilde{\mathbb{E}}$ behaves similarly to an expectation operator restricted to $\mathbb{R}^{\leq D}(v_1, \ldots, v_n)$ with the caveat that $\widetilde{\mathbb{E}}$ is only guaranteed to be non-negative on sum-of-squares polynomials.

The degree-D SoS algorithm checks feasibility of a polynomial system by checking whether or not a degree-D pseudoexpectation operator exists. To show an SoS lower bound, one must construct a pseudoexpectation operator.

2.2.2 Moment matrix

We define the moment matrix associated with a degree-D pseudoexpectation $\widetilde{\mathbb{E}}$.

Definition 2.5 (Moment Matrix of $\widetilde{\mathbb{E}}$)**.** *The moment matrix $\mathcal{M} = \mathcal{M}(\widetilde{\mathbb{E}})$ associated to a pseudoexpectation $\widetilde{\mathbb{E}}$ is a $\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}$ matrix with rows and columns indexed by subsets of $I, J \subseteq [n]$ of size at most $D/2$ and defined as*

$$\mathcal{M}[I, J] := \widetilde{\mathbb{E}}\left[v^I \cdot v^J\right].$$

To show that a candidate pseudoexpectation satisfies Item 3 in Definition 2.4, we will rely on the following standard fact.

Fact 2.6. *In the definition of pseudoexpectation, Definition 2.4, the condition in Item 3 is equivalent to $\mathcal{M} \succeq 0$.*

2.3 Graph matrices

To study $\mathcal{M}$, we decompose it using the framework of *graph matrices*. Originally developed in the context of the planted clique problem, graph matrices are random matrices whose entries are symmetric functions of an underlying random object – in our case, the set of vectors $d_1, \ldots, d_m$. We take the general presentation and results from [AMP20]. For our purposes, the following definitions are sufficient.

The graphs that we study have two types of vertices, circles ◯ and squares □. We let $\mathcal{C}_m$ be a set of m circles labeled 1 through m, which we denote by $\textcircled{1}, \textcircled{2}, \ldots, \textcircled{m}$, and let $\mathcal{S}_n$ be a set of n squares labeled 1 through n, which we denote by $\boxed{1}, \boxed{2}, \ldots, \boxed{n}$. We will work with bipartite graphs with edges between circles and squares, which have positive integer labels on the edges. When there are no multiedges (the graph is simple), such graphs are in one-to-one correspondence with Fourier characters on the vectors d_u. An edge between $\textcircled{u}$ and $\boxed{i}$ with label l represents $h_l(d_{u,i})$ where $\{h_k\}$ is the Fourier basis (e.g. Hermite polynomials).

$$\text{simple graph with labeled edges} \quad \Longleftrightarrow \quad \prod_{\substack{\textcircled{u} \in \mathcal{C}_m, \\ \boxed{i} \in \mathcal{S}_n}} h_{l(\textcircled{u}, \boxed{i})}(d_{u,i})$$

An example of a Fourier polynomial as a graph with labeled edges is given in Fig. 1. Unlabeled edges are implicitly labeled 1.

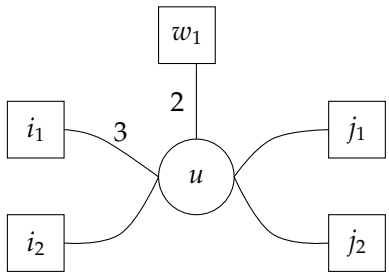

Figure 1: The Fourier polynomial $h_3(d_{u,i_1})h_1(d_{u,i_2})h_2(d_{u,w_1})h_1(d_{u,j_1})h_1(d_{u,j_2})$ represented as a graph.

Define the degree of a vertex v, denoted $\deg(v)$, to be the sum of the labels incident to v, and $|E|$ to be the sum of all labels. For intuition it is mostly enough to work with simple graphs, in which case these quantities make sense as the edge multiplicities in an implicit multigraph.

Definition 2.7 (Proper)**.** *We say an edge-labeled graph is proper if it has no multiedges.*

The definitions allow for "improper" edge-labeled multigraphs which simplify multiplying graph matrices (Section 5.2 and Section 7).

Definition 2.8 (Matrix indices)**.** *A matrix index is a set A of elements from $\mathcal{C}_m \cup \mathcal{S}_n$.*

We let $A(\boxed{i})$ or $A(\textcircled{u})$ be 0 or 1 to indicate if the vertex is in A.

Definition 2.9 (Ribbons)**.** *A ribbon is an undirected, edge-labeled graph $R = (V(R), E(R), A_R, B_R)$, where $V(R) \subseteq \mathcal{C}_m \cup \mathcal{S}_n$ and A_R, B_R are two matrix indices (possibly not disjoint) with $A_R, B_R \subseteq V(R)$, representing two distinguished sets of vertices. Furthermore, all edges in $E(R)$ go between squares and circles.*

We think of A_R and B_R as being the "left" and "right" sides of R, respectively. We also define the set of "middle vertices" $C_R := V(R) \setminus (A_R \cup B_R)$. If $e \notin E(R)$, then we define its label $l(e) = 0$. We also abuse notation and write $l(\boxed{i}, \textcircled{u})$ instead of $l(\{\boxed{i}, \textcircled{u}\})$.

Akin to the picture above, each ribbon corresponds to a Fourier polynomial. This Fourier polynomial lives inside a single entry of the matrix M_R. In the definition below, the $h_k(x)$ are the Fourier basis corresponding to the respective setting. In the Gaussian case, they are the (unnormalized) Hermite polynomials, and in the boolean case, they are just the parity function, represented by

$$h_0(x) = 1, \qquad h_1(x) = x, \qquad h_k(x) = 0 \ \ (k \geq 2)$$

Definition 2.10 (Matrix for a ribbon)**.** *The matrix M_R has rows and columns indexed by subsets of $\mathcal{C}_m \cup \mathcal{S}_n$, with a single nonzero entry defined by*

$$M_R[I, J] = \begin{cases} \prod_{\substack{e \in E(R), \\ e = \{\boxed{i}, \textcircled{u}\}}} h_{l(e)}(d_{u,i}) & I = A_R, J = B_R \\ 0 & \textit{Otherwise} \end{cases}$$

Next we describe the shape of a ribbon, which is essentially the ribbon when we have forgotten all the vertex labels and retained only the graph structure and the distinguished sets of vertices.

Definition 2.11 (Index shapes)**.** *An index shape is a set U of formal variables. Furthermore, each variable is labeled as either a "circle" or a "square".*

We let $U(\boxed{i})$ and $U(\textcircled{u})$ be either 0 or 1 for whether $\boxed{i}$ or $\textcircled{u}$, respectively, is in U.

Definition 2.12 (Shapes)**.** *A shape is an undirected, edge-labeled graph $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ where $V(\alpha)$ is a set of formal variables, each of which is labeled as either a "circle" or a "square". U_α and V_α are index shapes (possibly with variables in common) such that $U_\alpha, V_\alpha \subseteq V(\alpha)$. The edge set $E(\alpha)$ must only contain edges between the circle variables and the square variables.*

We'll also use $W_\alpha := V(\alpha) \setminus (U_\alpha \cup V_\alpha)$ to denote the "middle vertices" of the shape.

Remark 2.13. *We will abuse notation and use* $\boxed{i}, \boxed{j}, \textcircled{u}, \textcircled{v}, \ldots$ *for both the vertices of ribbons and the vertices of shapes. If they are ribbon vertices, then the vertices are elements of* $\mathcal{C}_m \cup \mathcal{S}_n$ *and if they are shape vertices, then they correspond to formal variables with the appropriate type.*

Definition 2.14 (Trivial shape)**.** *Define a shape* α *to be trivial if* $U_\alpha = V_\alpha$, $W_\alpha = \emptyset$ *and* $E(\alpha) = \emptyset$.

Definition 2.15 (Transpose of a shape)**.** *The transpose of a shape* $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ *is defined to be the shape* $\alpha^\intercal = (V(\alpha), E(\alpha), V_\alpha, U_\alpha)$.

For a shape α and an injective map $\sigma : V(\alpha) \to \mathcal{C}_m \cup \mathcal{S}_n$, we define the realization $\sigma(\alpha)$ as a ribbon in the natural way, by labeling all the variables using the map σ. We also require σ to be type-preserving i.e. it takes square variables to $\mathcal{S}_n$ and circle variables to $\mathcal{C}_m$. The ribbons that result are referred to as *ribbons of shape* α; notice that this partitions the set of all ribbons according to their shape[3][4].

Finally, given a shape α, the graph matrix M_α consists of all Fourier characters for ribbons of shape α.

Definition 2.16 (Graph matrices)**.** *Given a shape* $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$*, the graph matrix* M_α *is*

$$M_\alpha = \sum_{R \text{ is a ribbon of shape } \alpha} M_R$$

The moment matrix for PAP will turn out to be defined using graph matrices M_α whose left and right sides only have square vertices, and no circles. However, in the course of the analysis we will factor and multiply graph matrices with circle vertices in the left or right.

2.4 Norm bounds

The spectral norm of a graph matrix is determined, up to logarithmic factors, by relatively simple combinatorial properties of the graph. For a subset $S \subseteq \mathcal{C}_m \cup \mathcal{S}_n$, we define the weight $w(S) := (\#\text{ circles in } S) \cdot \log_n(m) + (\#\text{ squares in } S)$. Observe that $n^{w(S)} = m^{\#\text{ circles in } S} \cdot n^{\#\text{ squares in } S}$.

Definition 2.17 (Minimum vertex separator)**.** *For a shape* α*, a set* $S_{\min}$ *is a minimum vertex separator if all paths from* U_α *to* V_α *pass through* $S_{\min}$ *and* $w(S_{\min})$ *is minimized over all such separating sets.*

Let W_{iso} denote the set of isolated vertices in W_α. Then essentially the following norm bound holds for all shapes α with high probability (a formal statement can be found in Appendix A):

$$\|M_\alpha\| \leq \widetilde{O}\left(n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}\right)$$

In fact, the only probabilistic property required of the inputs $d_1, \ldots, d_m$ by our proof is that the above norm bounds hold for all shapes that arise in the analysis. We henceforth assume that the norm bounds in Lemma A.3 (for the Gaussian case) and Lemma A.1 (for the boolean case) hold.

[3]Partitions up to equality of shapes, where two shapes are equal if there is a type-preserving bijection between their variables that converts one shape to the other. When we operate on sets of shapes below, we implicitly use each distinct shape only once.

[4]Note that in our definition two realizations of a shape may give the same ribbon.

3 Proof Strategy

Here we explain in more detail the ideas for the Planted Affine Planes lower bound. Towards the proof of Theorem 1.4, fix a constant $\varepsilon > 0$ and a random instance $d_1, \dots, d_m$ with $n \le m \le n^{3/2-\varepsilon}$. We will construct a pseudoexpectation operator and show that it is PSD up to degree $D = 2 \cdot n^{\delta}$ with high probability.

We start by pseudocalibrating to obtain a pseudoexpectation operator $\widetilde{\mathbb{E}}$. The operator $\widetilde{\mathbb{E}}$ will exactly satisfy the "booleanity" constraints "$v_i^2 = \frac{1}{n}$" though it may not exactly satisfy the constraints "$\langle v, d_u \rangle^2 = 1$" due to truncation error in the pseudocalibration. Taking the truncation parameter n^{τ} to be larger than the degree D of the SoS solution, i.e., $\delta \ll \tau$, the truncation error is small enough that we can round $\widetilde{\mathbb{E}}$ to a nearby $\widetilde{\mathbb{E}}'$ that exactly satisfies the constraints. This is formally accomplished by viewing $\widetilde{\mathbb{E}} \in \mathbb{R}^{\binom{[n]}{\le D}}$ as a vector and expressing the constraints as a matrix Q such that $\widetilde{\mathbb{E}}$ satisfies the constraints iff it lies in the null space of Q. The choice of $\widetilde{\mathbb{E}}'$ is then the projection of $\widetilde{\mathbb{E}}$ to $\mathrm{Null}(Q)$. The end result is that we construct a moment matrix $\mathcal{M}_{fix} = \mathcal{M} + \mathcal{E}$ that exactly satisfies the constraints such that $\|\mathcal{E}\|$ is tiny. This step is done in Section 7.

After performing pseudocalibration, in both settings, we will have essentially the graph matrix decomposition

$$\mathcal{M} = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha = \sum_{\substack{\text{shapes } \alpha: \\ \deg(\boxed{i}) + U(\boxed{i}) + V(\boxed{i}) \text{ even}, \\ \deg(\textcircled{u}) \text{ even}}} \frac{1}{n^{\frac{|U_\alpha| + |V_\alpha|}{2}}} \cdot \left(\prod_{\textcircled{u} \in V(\alpha)} h_{\deg(\textcircled{u})}(1) \right) \cdot \frac{M_\alpha}{n^{|E(\alpha)|/2}}$$

Here $h_k(1)$ is in both settings the k-th Hermite polynomial, evaluated on 1.

In this decomposition of $\mathcal{M}$, the trivial shapes will be the dominant terms which we will use to bound the other terms. Recall that a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is trivial if $U_\alpha = V_\alpha$, $W_\alpha = \emptyset$ and $E(\alpha) = \emptyset$. These shapes contribute scaled identity matrices on different blocks of the main diagonal of $\mathcal{M}$, with trivial shape α contributing an identity matrix with coefficient $n^{-|U_\alpha|}$. Two trivial shapes are illustrated in Fig. 2.

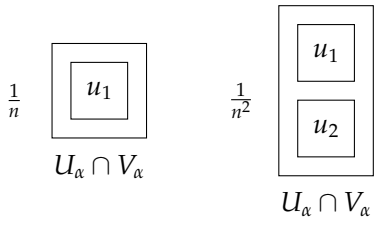

Figure 2: Two examples of trivial shapes.

Let $\mathcal{M}_{\text{triv}}$ be this diagonal matrix of trivial shapes in the above decomposition of $\mathcal{M}$. To prove that $\mathcal{M} \succeq 0$, we attempt the simple strategy of showing that the norm of all other terms can be "charged" against this diagonal matrix $\mathcal{M}_{\text{triv}}$. For several shapes this strategy is indeed viable. To illustrate, let's consider one such shape α depicted in Fig. 3.

This graph matrix has $|\lambda_\alpha| = \Theta(\frac{1}{n^5})$. Using the graph matrix norm bounds, with high probability the norm of this graph matrix is $\tilde{O}(n^2 m)$: there are four square vertices and two circle vertices which are not in the minimum vertex separator. Thus, for this shape α, with high proba-

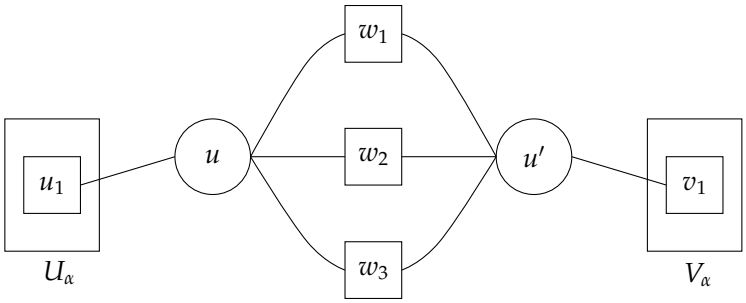

Figure 3: Picture of basic non-spider shape α.

bility $|\lambda_\alpha| \, \|M_\alpha\|$ is $\tilde{O}\left(\frac{m}{n^3}\right)$ and thus $\lambda_\alpha M_\alpha \preceq \frac{1}{n} Id$ (which is the multiple of the identity appearing in the corresponding block of $\mathcal{M}_{\text{triv}}$).

Unfortunately, as pointed out in the introduction, some shapes α that appear in the decomposition have $\|\lambda_\alpha M_\alpha\|$ too large to be charged against $\mathcal{M}_{\text{triv}}$. These are shapes with a certain substructure (actually the same structure that appears in the matrix Q used to project the pseudoexpectation operator!) whose norms cannot be handled by the preceding argument, and which we denote *spiders*. The following graph depicts one such *spider* shape (and also motivates this terminology):

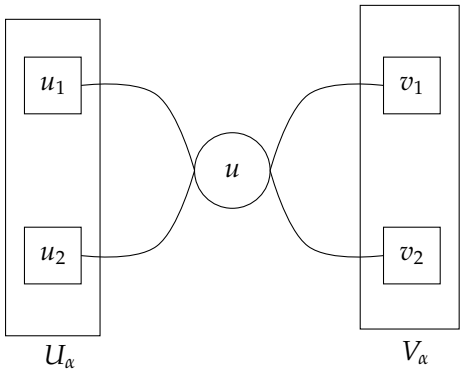

Figure 4: Picture of basic spider shape α.

The norm $\|\lambda_\alpha M_\alpha\|$ of this graph is $\widetilde{\Omega}(\frac{1}{n^2})$, as can be easily estimated through the norm bounds (the coefficient is $\lambda_\alpha = \frac{-2}{n^4}$, the minimum vertex separator is $\textcircled{u}$, and there are no isolated vertices). This is too large to bound against $\frac{1}{n^2} Id$, which is the coefficient of M_{triv} on this spider's block.

To skirt this and other spiders, we restrict ourselves to vectors $x \perp \text{Null}(M)$, and observe that this spider α satisfies $x^\intercal M_\alpha \approx 0$. To be more precise, consider the following argument. Consider the two shapes in Fig. 5, β_1 and β_2 (take note of the label 2 on the edge in β_2).

We claim that every column of the matrix $2M_{\beta_1} + \frac{1}{n} M_{\beta_2}$ is in the null space of $\mathcal{M}$. There are m nonzero columns indexed by assignments to V, which can be a single circle $\textcircled{1}, \textcircled{2}, \ldots, \textcircled{m}$. The nonzero rows are $\emptyset$ in β_2 and $\{\boxed{i}, \boxed{j}\}$ for $i \neq j$ in β_1. Fixing $I \subseteq [n]$, entry $(I, \textcircled{u})$ of the product

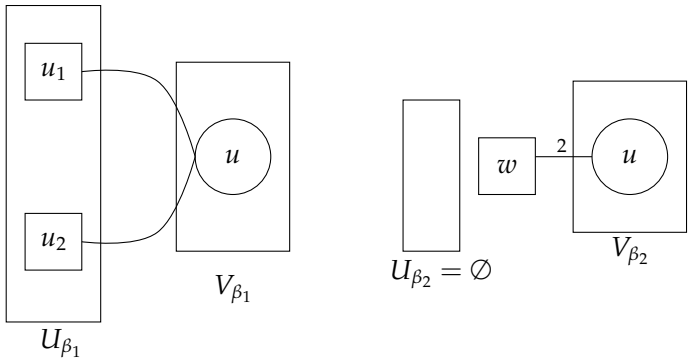

Figure 5: Picture of shapes β_1 and β_2.

matrix $\mathcal{M}(2M_{\beta_1} + \frac{1}{n}M_{\beta_2})$ is

$$\begin{aligned}
&2\sum_{i<j}\widetilde{\mathbb{E}}[v^I v_i v_j]\cdot d_{ui}d_{uj} + \frac{1}{n}\widetilde{\mathbb{E}}[v^I]\cdot\sum_i(d_{ui}^2-1)\\
&= 2\sum_{i<j}\widetilde{\mathbb{E}}[v^I v_i v_j]\cdot d_{ui}d_{uj} + \widetilde{\mathbb{E}}[v^I v_i^2]\cdot\sum_i d_{ui}^2 - \widetilde{\mathbb{E}}[v^I] && (\widetilde{\mathbb{E}}\text{ satisfies ``}v_i^2 = \frac{1}{n}\text{''})\\
&= \sum_{i,j}\widetilde{\mathbb{E}}[v^I v_i v_j]d_{ui}d_{uj} - \widetilde{\mathbb{E}}[v^I]\\
&= \widetilde{\mathbb{E}}[v^I(\langle v, d_u\rangle^2 - 1)]\\
&= 0 && (\widetilde{\mathbb{E}}\text{ satisfies ``}\langle v, d_u\rangle^2 = 1\text{''})
\end{aligned}$$

In words, the constraint "$\langle v, d_u\rangle^2 = 1$" creates a shape $2\beta_1 + \frac{1}{n}\beta_2$ that lies in the null space of the moment matrix. On the other hand, we can approximately factor the spider α across its central vertex, and when we do so, the shape β_1 appears on the left side.

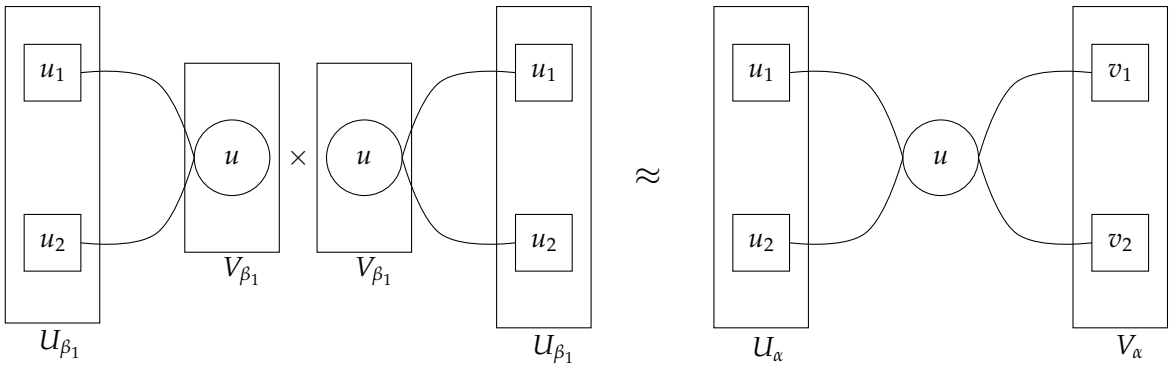

Figure 6: Approximation $\beta_1 \times \beta_1^\intercal \approx \alpha$.

Therefore $M_\alpha \approx M_{\beta_1}M_{\beta_1}^\intercal \approx (M_{\beta_1} + \frac{1}{2n}M_{\beta_2})M_{\beta_1}^\intercal$. The columns of the matrix $M_{\beta_1} + \frac{1}{2n}M_{\beta_2}$ are in the null space of $\mathcal{M}$, so for $x \perp \text{Null}(\mathcal{M})$ we have $x^\intercal M_\alpha \approx 0$.

More formally, we are able to find coefficients c_β so that all columns of the matrix

$$A = M_\alpha + \sum_\beta c_\beta M_\beta$$

are in Null$(\mathcal{M})$. We then observe the following fact:

Fact 3.1. *If $x \perp \text{Null}(\mathcal{M})$ and $\mathcal{M}A = 0$, then $x^\intercal(AB + \mathcal{M})x = x^\intercal(B^\intercal A^\intercal + \mathcal{M})x = x^\intercal \mathcal{M} x$.*

Using the fact, we can freely add multiples of A to $\mathcal{M}$ without changing the action of $\mathcal{M}$ on Null$(\mathcal{M})^\perp$. A judicious choice is to subtract $\lambda_\alpha A$ which will "kill" the spider from $\mathcal{M}$. Doing this for all spiders, we produce a matrix whose action is equivalent on Null$(\mathcal{M})^\perp$, and which has high minimum eigenvalue by virtue of the fact that it has no spiders, showing that $\mathcal{M}$ is PSD.

The catch is two-fold: first, the coefficients c_β may contribute to the coefficients on the non-spiders; second, the further intersection terms M_β may themselves be spiders (though they will always have fewer square vertices than α). Thus we must recursively kill these spiders, until there are no spiders remaining in the decomposition of $\mathcal{M}$. The resulting matrix has some new coefficients on the non-spiders

$$\mathcal{M}' = \sum_{\text{non-spiders } \beta} \lambda'_\beta M_\beta.$$

We must bound the accumulation on the coefficients λ'_β. We do this by considering the *web* of spiders and non-spiders created by each spider and using bounds on the c_β and λ_α to argue that the contributions do not blow up, via an interesting charging scheme that exploits the structure of these graphs.

4 Pseudocalibration

To be able to apply the pseudocalibration technique of [BHK+16] to an average-case feasibility problem, in our case the PAP problem, one needs to design a planted distribution supported on feasible instances. This is done in Section 4.1. In Section 4.2, we recall the precise details in applying pseudocalibration. Then we pseudocalibrate in the Gaussian (Section 4.3) and boolean (Section 4.4) settings.

4.1 PAP planted distribution

We formally define the random and the planted distributions for the Planted Affine Planes problem in the Gaussian and boolean settings. These two (families of) distributions are required by the pseudocalibration machinery in order to define a candidate pseudoexpectation operator $\widetilde{\mathbb{E}}$. For the Gaussian setting, we have the following distributions.

Definition 4.1 (Gaussian PAP distributions)**.** *The Gaussian PAP distributions are as follows.*

1. *(Random distribution) m i.i.d. vectors $d_u \sim \mathcal{N}(0, I)$.*
2. *(Planted distribution) A vector v is sampled uniformly from $\left\{\pm\frac{1}{\sqrt{n}}\right\}^n$, as well as signs $b_u \in_R \{\pm 1\}$, and m vectors d_u are drawn from $\mathcal{N}(0, I)$ conditioned on $\langle d_u, v\rangle = b_u$.*

For the boolean setting, we have the following distributions.

Definition 4.2 (Boolean PAP distributions)**.** *The boolean PAP distributions are as follows*

1. *(Random distribution) m i.i.d. vectors $d_u \in_R \{-1, +1\}^n$.*
2. *(Planted distribution) A vector v is sampled uniformly from $\left\{\pm\frac{1}{\sqrt{n}}\right\}^n$, as well as signs $b_u \in_R \{\pm 1\}$, and m vectors d_u are drawn from $\{\pm 1\}^n$ conditioned on $\langle d_u, v\rangle = b_u$.*

4.2 Pseudocalibration technique

We will use the shorthand $\mathbb{E}_{\text{ra}}$ and $\mathbb{E}_{\text{pl}}$ for the expectation under the random and planted distributions. Pseudocalibration gives a method for constructing a candidate pseudoexpectation operator $\widetilde{\mathbb{E}}$. The idea behind pseudocalibration is that $\mathbb{E}_{\text{ra}} \widetilde{\mathbb{E}} f(v)$ should match with $\mathbb{E}_{\text{pl}} f(v)$ for every low-degree test of the data $t = t(d) = t(d_1, \dots, d_m)$,

$$\mathop{\mathbb{E}}_{\text{ra}} t(d) \widetilde{\mathbb{E}} f(v) = \mathop{\mathbb{E}}_{\text{pl}} t(d) f(v).$$

When pseudocalibrating, one can freely choose the "outer" basis in which to express the polynomial $f(v)$, as well as the "inner" basis of low-degree tests which should agree with the planted distribution. Though we attempted to use alternate bases to simplify the analysis, ultimately we opted for the standard choice of bases: a Fourier basis for the inner basis in each setting (Hermite functions for the Gaussian setting, parity functions for the boolean setting), and the coordinate basis v^I for the outer basis.

When the inner basis is orthonormal under the random distribution (as a Fourier basis is), the pseudocalibration condition gives a formula for the coefficients of $\widetilde{\mathbb{E}} f(v)$ in the orthonormal basis (though it only gives the coefficients of the low-degree functions $t(d)$). Concretely, letting the inner basis be indexed by $\alpha \in \mathcal{F}$, as a function of d the pseudocalibration condition enforces

$$\widetilde{\mathbb{E}} f(v) = \sum_{\substack{\alpha \in \mathcal{F}: \\ |\alpha| \le n^\tau}} \left(\mathop{\mathbb{E}}_{\text{pl}} t_\alpha(d) f(v) \right) t_\alpha(d).$$

Here we use "$|\alpha| \le n^\tau$" to describe the set of low-degree tests. The pseudocalibration condition does not prescribe any coefficients for functions $t_\alpha(d)$ with $|\alpha| > n^\tau$ and an economical choice is to set these coefficients to zero.

When pseudocalibrating, our pseudoexpectation operator is guaranteed to be linear, as the expression above is linear in f. It is guaranteed to satisfy all constraints of the form "$f(v) = 0$". It will approximately satisfy constraints of the form "$f(v, d) = 0$", though only up to truncation error.

Fact 4.3 (Proof in Lemma 7.7)**.** *If $p(v)$ is a polynomial which is uniformly zero on the planted distribution, then $\widetilde{\mathbb{E}}[p]$ is the zero function. If $p(v, d)$ is a polynomial which is uniformly zero on the planted distribution, then the only nonzero Fourier coefficients of $\widetilde{\mathbb{E}}[p]$ are those with size between $n^\tau \pm \deg_d(p)$.*

Truncation introduces a tiny error in the constraints, which we are able to handle in a uniform way in Section 7.

For the pseudocalibration we truncate to only Fourier coefficients of size at most n^τ. The relationship between the parameters is $\delta \le c\tau \le c'\varepsilon$ where $c' < c < 1$ are absolute constants. We will assume that they are sufficiently small for all our proofs to go through.

Pseudocalibration also by default does not enforce the condition $\widetilde{\mathbb{E}}[1] = 1$. However, this is easily fixed by dividing the operator by $\widetilde{\mathbb{E}}[1]$. As will be pointed out in Remark 5.9, w.h.p. in the unnormalized pseudocalibration, $\widetilde{\mathbb{E}}[1] = 1 + o_n(1)$ and so the error introduced does not impact the statement of any lemmas.

4.3 Gaussian setting pseudocalibration

We start by computing the pseudocalibration for the Gaussian setting. Here the natural choice of Fourier basis is the Hermite polynomials. Let $\alpha \in (\mathbb{N}^n)^m$ denote a Hermite polynomial index. Define $\alpha! := \prod_{u,i} \alpha_{u,i}!$ and $|\alpha| := \sum_{u,i} \alpha_{u,i}$ and $|\alpha_u| := \sum_i \alpha_{u,i}$. We let $h_\alpha(d_1, \dots, d_m)$ denote an unnormalized Hermite polynomial, so that $h_\alpha / \sqrt{\alpha!}$ forms an orthonormal basis for polynomials in the entries of the vectors $d_1, \dots, d_m$, under the inner product $\langle p, q \rangle = \mathbb{E}_{d_1,\dots,d_m \sim \mathcal{N}(0,I)}[p \cdot q]$.

We can view α as an $m \times n$ matrix of natural numbers, and with this view we also define $\alpha^\intercal \in (\mathbb{N}^m)^n$.

Lemma 4.4. *For any $I \subseteq [n]$, the pseudocalibration value is*

$$\widetilde{\mathbb{E}} v^I = \sum_{\substack{\alpha : |\alpha| \le n^\tau, \\ |\alpha_u| \text{ even}, \\ |(\alpha^\intercal)_i| \equiv I_i \pmod 2}} \left(\prod_{u=1}^m h_{|\alpha_u|}(1) \right) \cdot \frac{1}{n^{|I|/2 + |\alpha|/2}} \cdot \frac{h_\alpha(d_1, \dots, d_m)}{\alpha!}.$$

In words, the nonzero Fourier coefficients are those which have even row sums, and whose column sums match the parity of I.

Proof. The truncated pseudocalibrated value is defined to be

$$\widetilde{\mathbb{E}} v^I = \sum_{\alpha : |\alpha| \le n^\tau} \frac{h_\alpha(d_1, \dots, d_m)}{\alpha!} \cdot \mathop{\mathbb{E}}_{\text{pl}} [h_\alpha(d_1, \dots, d_m) \cdot v^I]$$

So we set about to compute the planted moments. For this computation, the following lemma is crucial. Here, we give a short proof of this lemma using generating functions. For a combinatorial proof, see Appendix C.

Lemma 4.5. *Let $\alpha \in \mathbb{N}^n$. When v is fixed and b is fixed (not necessarily +1 or -1) and $d \sim N(0, I)$ conditioned on $\langle v, d \rangle = b \|v\|$,*

$$\mathop{\mathbb{E}}_d [h_\alpha(d)] = \frac{v^\alpha}{\|v\|^{|\alpha|}} \cdot h_{|\alpha|}(b).$$

Proof. It suffices to prove the claim when $\|v\| = 1$ since the left-hand side is independent of $\|v\|$. Express $d = bv + (I - vv^\intercal)x$ where $x \sim N(0, I)$ is a standard normal variable. Now we want

$$\mathop{\mathbb{E}}_{x \sim N(0,I)} h_\alpha \left(bv + (I - vv^\intercal)x \right).$$

The Hermite polynomial generating function is

$$\sum_{\alpha \in \mathbb{N}^n} \mathop{\mathbb{E}}_{x \sim N(0,I)} h_\alpha \left(bv + (I - vv^\intercal)x \right) \frac{t^\alpha}{\alpha!} = \mathop{\mathbb{E}}_x \exp \left(\langle bv + (I - vv^\intercal)x, t \rangle - \frac{\|t\|_2^2}{2} \right)$$

$$= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \exp \left(\langle bv + (I - vv^\intercal)x, t \rangle - \frac{\|t\|_2^2}{2} - \frac{\|x\|_2^2}{2} \right) dx.$$

Completing the square,

$$
\begin{aligned}
&= \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \exp\left(\langle bv, t\rangle - \frac{\langle v, t\rangle^2}{2} - \frac{1}{2} \cdot \|x - (t - \langle v, t\rangle v)\|_2^2\right) \, dx \\
&= \exp\left(\langle bv, t\rangle - \frac{\langle v, t\rangle^2}{2}\right) \\
&= \exp\left(b\langle v, t\rangle - \frac{1}{2} \cdot \langle v, t\rangle^2\right).
\end{aligned}
$$

How can we Taylor expand this in terms of t? The Taylor expansion of $\exp(by - \frac{y^2}{2})$ is $\sum_{i=0}^{\infty} h_i(b)\frac{y^i}{i!}$. That is, the i-th derivative in y of $\exp(by - \frac{y^2}{2})$, evaluated at 0, is $h_i(b)$. Using the chain rule with $y = \langle v, t\rangle$, the α-derivative in t of our expression, evaluated at 0, is $v^\alpha \cdot h_{|\alpha|}(b)$. This is the expression we wanted when $\|v\| = 1$, and along with the aforementioned remark about homogeneity in $\|v\|$ this completes the proof. ∎

Now we can finish the calculation. To compute $\mathbb{E}_{\text{pl}}[h_\alpha(d_1, \ldots, d_m) \cdot v^I]$, marginalize v and the b_u and factor the conditionally independent b_u and d_u.

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\text{pl}}[h_\alpha(d_1, \ldots, d_m)v^I] &= \mathop{\mathbb{E}}_{v, b_u} v^I \prod_{u=1}^{m} \mathop{\mathbb{E}}_{d} [h_{\alpha_u}(d_u) \mid v, b_u] \\
&= \mathop{\mathbb{E}}_{v, b_u} v^I \cdot \prod_{u=1}^{m} \frac{v^{\alpha_u}}{\|v\|^{|\alpha_u|}} \cdot h_{|\alpha_u|}(b_u) \qquad \text{(Lemma 4.5)} \\
&= \left(\mathop{\mathbb{E}}_{v} \frac{v^{I + \sum_{u=1}^m \alpha_u}}{\|v\|^{\sum_{u=1}^m |\alpha_u|}}\right) \cdot \left(\prod_{u=1}^{m} \mathop{\mathbb{E}}_{b_u} h_{|\alpha_u|}(b_u)\right)
\end{aligned}
$$

The Hermite polynomial expectations will be zero in expectation over b_u if the degree is odd, and otherwise b_u is raised to an even power and can be replaced by 1. This requires that $|\alpha_u|$ is even for all u. The norm $\|v\|$ is constantly 1 and can be dropped. The numerator will be $\frac{1}{n^{|I|/2 + |\alpha|/2}}$ if the parity of every $|(\alpha^\intercal)_i|$ matches I_i, and 0 otherwise. This completes the pseudocalibration calculation. ∎

We can now write $\mathcal{M}$ in terms of graph matrices.

Definition 4.6. *Let $\mathcal{L}$ be the set of all proper shapes α with the following properties*

- *U_α and V_α only contain square vertices and $|U_\alpha|, |V_\alpha| \le n^\delta$*
- *W_α has no degree 0 vertices*
- *$\deg(\boxed{i}) + U_\alpha(\boxed{i}) + V_\alpha(\boxed{i})$ is even for all $\boxed{i} \in V(\alpha)$*
- *$\deg(\textcircled{u})$ is even and $\deg(\textcircled{u}) \ge 4$ for all $\textcircled{u} \in V(\alpha)$*
- *$|E(\alpha)| \le n^\tau$*

Remark 4.7. *Note that the shapes in $\mathcal{L}$ can have isolated vertices in $U_\alpha \cap V_\alpha$.*

Remark 4.8. *$\mathcal{L}$ captures all the shapes that have nonzero coefficient when we write $\mathcal{M}$ in terms of graph matrices. The constraint $\deg(\textcircled{u}) \ge 4$ arises because pseudocalibration gives us that $\deg(\textcircled{u})$ is even, $\textcircled{u}$ cannot be isolated, and $h_2(1) = 0$.*

For a shape α, we define

$$\alpha! := \prod_{e \in E(\alpha)} l(e)!$$

Note that this equals the factorial of the corresponding index of the Hermite polynomial for this shape.

Definition 4.9. *For any shape α, if $\alpha \in \mathcal{L}$, define*

$$\lambda_\alpha := \left(\prod_{\textcircled{u} \in V(\alpha)} h_{\deg(\textcircled{u})}(1) \right) \cdot \frac{1}{n^{(|U_\alpha|+|V_\alpha|+|E(\alpha)|)/2}} \cdot \frac{1}{\alpha!}$$

Otherwise, define $\lambda_\alpha := 0$.

Corollary 4.10. *Modulo the footnote*[5]*, $\mathcal{M} = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$.*

4.4 Boolean setting pseudocalibration

We now present the pseudocalibration for the boolean setting. For the sequel, we need notation for vectors on a slice of the boolean cube.

Definition 4.11 (Slice)**.** *Let $v \in \{\pm 1\}^n$ and $\theta \in \mathbb{Z}$. The slice $\mathcal{S}_v(\theta)$ is defined as*

$$\mathcal{S}_v(\theta) := \{d \in \{\pm 1\}^n \mid \langle v, d \rangle = \theta\}.$$

We use $\mathcal{S}_v(\pm\theta)$ to denote $\mathcal{S}_v(\theta) \cup \mathcal{S}_v(-\theta)$ and $\mathcal{S}(\theta)$ to denote $\mathcal{S}_v(\theta)$ when v is the all-ones vector.

Remark 4.12. *With our notation for the slice, the planted distribution in the boolean setting can be equivalently described as*

1. *Sample $v \in \{\frac{\pm 1}{\sqrt{n}}\}^n$ uniformly, and then*
2. *Sample $d_1, \dots, d_m$ independently and uniformly from $\mathcal{S}_{\sqrt{n} \cdot v}(\pm\sqrt{n})$.*

The planted distribution doesn't actually exist for every n, but this is immaterial, as we can still define the pseudoexpectation via the same formula.

We will also need the expectation of monomials over the slice $\mathcal{S}(\sqrt{n})$ since they will appear in the description of the pseudocalibrated Fourier coefficients.

Definition 4.13. $e(k) := \mathbb{E}_{x \in_R \mathcal{S}(\sqrt{n})} [x_1 \cdots x_k]$.

We now compute the Fourier coefficients of $\widetilde{\mathbb{E}} v^\beta$, where $\beta \in \mathbb{F}_2^n$. The Fourier basis when $d_1, \dots, d_m \in_R \{\pm 1\}^n$ is the set of parity functions. Thus a character can be specified by $\alpha \in (\mathbb{F}_2^n)^m$, where α is composed of m vectors $\alpha_1, \dots, \alpha_m \in \mathbb{F}_2^n$. More precisely, the character χ_α associated to α is defined as

$$\chi_\alpha(d_1, \dots, d_m) := \prod_{u=1}^m d_u^{\alpha_u}$$

We denote by $|\alpha|$ the number of non-zero entries of α and define $|\alpha_u|$ similarly. Thinking of α as an $m \times n$ matrix with entries in $\mathbb{F}_2$, we also define $\alpha^\intercal \in (\mathbb{F}_2^n)^m$.

[5] Technically, the graph matrices M_α have rows and columns indexed by all subsets of $\mathcal{C}_m \cup \mathcal{S}_n$. The submatrix with rows and columns from $\binom{\mathcal{S}_n}{\leq D/2}$ equals the moment matrix for $\widetilde{\mathbb{E}}$ as defined in Section 2.2.2.

Lemma 4.14. *We have*

$$\widetilde{\mathbb{E}} v^\beta = \frac{1}{n^{|\beta|/2}} \sum_{\substack{\alpha:\, |\alpha| \le n^\tau, \\ |\alpha_u| \text{ even}, \\ |\alpha_i^\intercal| \equiv \beta_i \pmod 2}} \prod_{u=1}^m e(|\alpha_u|) \cdot \chi_{\alpha_u}(d_u).$$

The set of nonzero coefficients has a similar structure as in the Gaussian case: the rows of α must have an even number of entries, and the i-th column must have parity matching β_i.

Proof. Given $\alpha \in (\mathbb{F}_2^n)^m$ with $|\alpha| \le n^\tau$, the pseudocalibration equation enforces by construction that

$$\mathop{\mathbb{E}}_{d_1,\dots,d_m \in \{\pm 1\}^n} (\widetilde{\mathbb{E}} v^\beta)(d_1,\dots,d_m) \cdot \chi_\alpha(d_1,\dots,d_m) = \mathop{\mathbb{E}}_{\text{pl}} v^\beta \cdot \chi_\alpha(d_1,\dots,d_m).$$

Computing the RHS above yields

$$\begin{aligned}
\mathop{\mathbb{E}}_{v \in \{\pm 1\}^n} \mathop{\mathbb{E}}_{d_1,\dots,d_m \in_{\mathrm{R}} \mathcal{S}_v(\pm\sqrt{n})} \left[v^\beta \prod_{u=1}^m \chi_{\alpha_u}(d_u) \right] &= \mathop{\mathbb{E}}_{v \in \{\pm 1\}^n} \mathop{\mathbb{E}}_{d_1,\dots,d_m \in_{\mathrm{R}} \mathcal{S}(\pm\sqrt{n})} \left[v^\beta \prod_{u=1}^m \chi_{\alpha_u}(v) \chi_{\alpha_u}(d_u) \right] \\
&= \mathop{\mathbb{E}}_{v \in \{\pm 1\}^n} \chi_{\alpha_1+\cdots+\alpha_m+\beta}(v) \mathop{\mathbb{E}}_{d_1,\dots,d_m \in \mathcal{S}(\pm\sqrt{n})} \left[\prod_{i=1}^m \chi_{\alpha_i}(d_i) \right] \\
&= \mathbf{1}_{[\alpha_1+\cdots+\alpha_m=\beta]} \cdot \prod_{i=1}^m \mathop{\mathbb{E}}_{d_i \in \mathcal{S}(\pm\sqrt{n})} [\chi_{\alpha_i}(d_i)] \\
&= \mathbf{1}_{[\alpha_1+\cdots+\alpha_m=\beta]} \cdot \prod_{i=1}^m \mathbf{1}_{[|\alpha_i| \equiv 0 \pmod 2]} \cdot \prod_{i=1}^m e(|\alpha_i|).
\end{aligned}$$

Since we have a general expression for the Fourier coefficient of each character, applying Fourier inversion concludes the proof. ∎

We can now express the moment matrix in terms of graph matrices.

Definition 4.15. *Let $\mathcal{L}_{bool}$ be the set of shapes in $\mathcal{L}$ from Definition 4.6 in which the edge labels are all 1.*

Remark 4.16. *$\mathcal{L}_{bool}$ captures all the shapes that have nonzero coefficient when we write $\mathcal{M}$ in terms of graph matrices. Similar to Remark 4.8, since $e(2) = 0$ (see Claim B.1), we have the same condition $deg(\textcircled{u}) \ge 4$ for shapes in $\mathcal{L}_{bool}$.*

Definition 4.17. *For all shapes α, if $\alpha \in \mathcal{L}_{bool}$ define*

$$\lambda_\alpha := \frac{1}{n^{(|U_\alpha|+|V_\alpha|)/2}} \prod_{\textcircled{u} \in V(\alpha)} e(\deg(\textcircled{u}))$$

Otherwise, let $\lambda_\alpha := 0$.

Corollary 4.18. $\mathcal{M} = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$

4.4.1 Unifying the analysis

It turns out that the analysis of the boolean setting mostly follows from the analysis in the Gaussian setting. Initially, the boolean pseudocalibration is essentially equal to the Gaussian pseudocalibration in which we have removed all shapes containing at least one edge with a label $k \geq 2$. The coefficients on the graph matrices will actually be slightly different, but they both admit an upper bound that is sufficient for our purposes (see Proposition 5.13 for the precise statement).

To unify the notation in our analysis, we conveniently set the edge functions of the graphs in the boolean case to be

$$h_k(x) = \begin{cases} 1 & \text{if } k = 0 \\ x & \text{if } k = 1 \\ 0 & \text{if } k \geq 2 \end{cases}$$

This choice of $h_k(x)$ preserves the fact that $\{h_0(x) = 1, h_1(x) = x\}$ is an orthogonal polynomial basis in the boolean setting, while zeroing out graphs with larger labels.

During the course of the analysis, we may multiply two graph matrices and produce graph matrices with improper parallel edges (so-called "intersections terms"). For a fixed pair u, i of vertices, parallel edges between u and i with labels $l_1, \dots, l_s$ correspond to the product of orthogonal polynomials $\prod_{j=1}^{s} h_{l_j}(d_{u,i}) =: q(d_{u,i})$. We will re-express this product as a linear combination of polynomials in the orthogonal family, i.e., $q(d_{u,i}) = \sum_{i=0}^{\deg(q)} \lambda_i \cdot h_i(d_{u,i})$ for some coefficients $\lambda_i \in \mathbb{R}$. For the boolean case, the polynomial $q(d_{u,i})$ will be either $h_0(d_{u,i}) = 1$ or $h_1(d_{u,i}) = d_{u,i}$. However, for the Gaussian setting there may be up to $\deg(q)$ non-zero, potentially larger coefficients λ_i for the corresponding Hermite polynomials h_i. For the graphs that arise in this way, we will always bound their contributions to $\mathcal{M}$ by applying the triangle inequality and norm bounds. Since we show bounds using the larger coefficients λ_i from the Gaussian case, the same bounds apply when using the $0/1$ coefficients in the boolean case.

We will consider separate cases at any point where the analysis differs between the two settings.

5 Proving PSD-ness

Looking at the shapes that make up $\mathcal{M}$, the trivial shape with k square vertices contributes an identity matrix on the degree-$2k$ submatrix of $\mathcal{M}$. Our ultimate goal will be to bound all shapes against these identity matrices.

Definition 5.1 (Block)**.** *For $k, l \in \{0, 1, \dots, D/2\}$, the (k, l) block of $\mathcal{M}$ is the submatrix with rows from $\binom{[n]}{k}$ and columns from $\binom{[n]}{l}$. Note that when $\mathcal{M}$ is expressed as a sum of graph matrices, this exactly restricts $\mathcal{M}$ to shapes α with $|U_\alpha| = k$ and $|V_\alpha| = l$.*

We define the parameter $\eta := 1/\sqrt{n}$. The trivial shapes live in the diagonal blocks of $\mathcal{M}$, and on the (k, k) block contribute a factor of $\frac{1}{n^k} = \eta^{2k}$ on the diagonal. In principle, we could make η as small as we like[6] by considering the moments of a rescaling of v rather than v itself. Counterintuitively, it will turn out that the scaling helps us prove PSD-ness (see Appendix D for more details). It turns out that pseudocalibrating v as a unit vector (equivalently, using $\eta = 1/\sqrt{n}$) is sufficient for our analysis.

[6] Though pseudocalibration truncation errors may become nonnegligible for extremely tiny η.

Towards the goal of bounding $\mathcal{M}$ by the identity terms, we will bound the norm of matrices on each block of $\mathcal{M}$, and invoke the following lemma to conclude PSD-ness.

Lemma 5.2. *Suppose a symmetric matrix $\mathcal{A} \in \mathbb{R}^{\binom{[n]}{\leq D} \times \binom{[n]}{\leq D}}$ satisfies, for some parameter $\eta \in (0,1)$,*

1. *For each $k \in \{0,1,\ldots,D\}$, the (k,k) block has minimum singular value at least $\eta^{2k}(1-\frac{1}{D+1})$*
2. *For each $k,l \in \{0,1,\ldots,D\}$ such that $k \neq l$, the (k,l) block has norm at most $\frac{\eta^{k+l}}{D+1}$.*

Then $\mathcal{A} \succeq 0$.

Proof. We need to show that for all vectors x, $x^\intercal \mathcal{A} x \geq 0$. Given a vector x, let $x_0,\ldots,x_D$ be its components in blocks $0,\ldots,D$. Observe that

$$x^\intercal \mathcal{A} x \geq \sum_{k\in[0,D]} \eta^{2k}\left(1-\frac{1}{D+1}\right)\|x_k\|^2 - \sum_{k\neq l\in[0,D]} \frac{\eta^{k+l}}{D+1}\|x_k\|\,\|x_l\|$$

$$= \left(\|x_0\|, \eta\|x_1\|, \ldots, \eta^D\|x_D\|\right) \begin{pmatrix} 1-\frac{1}{D+1} & -\frac{1}{D+1} & \cdots & -\frac{1}{D+1} \\ -\frac{1}{D+1} & 1-\frac{1}{D+1} & \cdots & -\frac{1}{D+1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{D+1} & -\frac{1}{D+1} & \cdots & 1-\frac{1}{D+1} \end{pmatrix} \begin{pmatrix} \|x_0\| \\ \eta\|x_1\| \\ \vdots \\ \eta^D\|x_D\| \end{pmatrix} \geq 0.$$

■

We start by defining spiders, which are special shapes α that we will handle separately in the decomposition of $\mathcal{M}$. Informally, these contain special substructures which allow their norm bounds not to be negligible with respect to the identity matrix. We then show that shapes which are not spiders have bounded norms.

Definition 5.3 (Left Spider)**.** *A left spider is a proper shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ with the property that there exist two distinct square vertices $\boxed{i}, \boxed{j} \in U_\alpha$ of degree 1 and a circle vertex $\textcircled{u} \in V(\alpha)$ such that $E(\alpha)$ contains the edges $(\boxed{i}, \textcircled{u})$ and $(\boxed{j}, \textcircled{u})$ (these are necessarily the only edges incident to $\boxed{i}$ and $\boxed{j}$).*

The vertices $\boxed{i}$ and $\boxed{j}$ are called the *end vertices* of α. Because of degree parity, the end vertices must lie in $U_\alpha \setminus (U_\alpha \cap V_\alpha)$.

Definition 5.4 (Right spider)**.** *A shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is a right spider if $\alpha^\intercal = (V(\alpha), E(\alpha), V_\alpha, U_\alpha)$ is a left spider. The end vertices of $\alpha^\intercal$ are also called the end vertices of α.*

Definition 5.5 (Spider)**.** *A shape α is a spider if it is either a left spider or a right spider.*

Remark 5.6. *A spider can have many pairs of end vertices. For each possible spider shape, we single out a pair of end vertices, so that in what follows we can discuss "the" end vertices of the spider.*

5.1 Non-spiders are negligible

For non-spiders, we will now show that their norm is small. We point out that this norm bound on non-spiders critically relies on the assumption $m \leq n^{3/2-\varepsilon}$.

Lemma 5.7. *If $\alpha \in \mathcal{L}$ is not a trivial shape and not a spider, then*

$$\frac{1}{n^{|E(\alpha)|/2}} n^{\frac{w(V(\alpha))-w(S_{\min})}{2}} \leq \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

where S_{min} is the minimum vertex separator of α.

Proof. The idea behind the proof is as follows. Each square vertex which is not in the minimum vertex separator contributes $\sqrt{n}$ to the norm bound while each circle vertex which is not in the minimum vertex separator contributes $\sqrt{m}$. To compensate for this, we will try and take the factor of $\frac{1}{\sqrt{n}}$ from each edge and distribute it among its two endpoints so that each square vertex which is not in the minimum vertex separator is assigned a factor of $\frac{1}{\sqrt{n}}$ or smaller and each circle vertex which is not in the minimum vertex separator is assigned a factor of $\frac{1}{\sqrt{m}}$ or smaller.

Remark 5.8. *Instead of using the minimum vertex separator, we will actually use a set S of square vertices such that $w(S) \leq w(S_{\min})$. For details, see the actual distribution scheme below.*

To motivate the distribution scheme which we use, we first give two attempts which don't quite work. For simplicity, for these first two attempts we assume that $U_\alpha \cap V_\alpha = \emptyset$ as vertices in $U_\alpha \cap V_\alpha$ can essentially be ignored.

Attempt 1: Take each edge and assign a factor of $\frac{1}{\sqrt[4]{n}}$ to its square endpoint and a factor of $\frac{1}{\sqrt[8]{m}}$ to its circle endpoint.

With this distribution scheme, since each circle vertex has degree at least 4, each circle vertex is assigned a factor of $\frac{1}{\sqrt{m}}$ or smaller. Since each square vertex in W_α has degree at least 2, each square vertex in W_α is assigned a factor of $\frac{1}{\sqrt{n}}$ or smaller. However, square vertices in $U_\alpha \cup V_\alpha$ may only have degree 1 in which case they are assigned a factor of $\frac{1}{\sqrt[4]{n}}$ which is not small enough.

To fix this issue, we can have all of the edges which are incident to a square vertex in $U_\alpha \cup V_\alpha$ give their entire factor of $\frac{1}{\sqrt{n}}$ to the square vertex.

Remark 5.9. *For analyzing $\widetilde{\mathbb{E}}[1]$, this first attempt works as $U_\alpha = V_\alpha = \emptyset$. Thus, as long as $m \leq n^{2-\varepsilon}$, with high probability $\widetilde{\mathbb{E}}[1] = 1 \pm o_n(1)$.*

Attempt 2: For each edge which is between a square vertex in $U_\alpha \cup V_\alpha$ and a circle vertex, we assign a factor of $\frac{1}{\sqrt{n}}$ to the square vertex and nothing to the circle vertex. For all other edges, we assign a factor of $\frac{1}{\sqrt[4]{n}}$ to its square endpoint and a factor of $\frac{1}{\sqrt[6]{m}}$ to its circle endpoint (which we can do because $m \leq n^{\frac{3}{2}-\varepsilon}$).

With this distribution scheme, each square vertex is assigned a factor of $\frac{1}{\sqrt{n}}$. Since α is not a spider, no circle vertex is adjacent to two vertices in U_α or V_α. Thus, any circle vertex which is not adjacent to both a square vertex in U_α and a square vertex in V_α must be adjacent to at least 3 square vertices in W_α and is thus assigned a factor of $\frac{1}{\sqrt{m}}$ or smaller. However, we can have circle vertices which are adjacent to both a square vertex in U_α and a square vertex in V_α. These circle vertices may be assigned a factor of $\frac{1}{\sqrt[3]{m}}$, which is not small enough.

To fix this, observe that whenever we have a circle vertex which is adjacent to both a square vertex in U_α and a square vertex in V_α, this gives a path of length 2 from U_α to V_α. Any vertex separator must contain one of the vertices in this path, so we can put one of these two square vertices in S and not assign it a factor of $\frac{1}{\sqrt{n}}$.

Actual distribution scheme: Based on these observations, we use the following distribution scheme. Here we are no longer assuming that $U_\alpha \cap V_\alpha$ is empty.

1. Choose a set of square vertices $S \subseteq U_\alpha \cup V_\alpha$ as follows. Start with $S = U_\alpha \cap V_\alpha$. Whenever we have a circle vertex which is adjacent to both a square vertex in $U_\alpha \setminus V_\alpha$ and a square

vertex in $V_\alpha \setminus U_\alpha$, put one of these two square vertices in S (this choice is arbitrary). Observe that $w(S) \leq w(S_{\min})$

2. For each edge which is incident to a square vertex in S, assign a factor of $\frac{1}{\sqrt[3]{m}}$ to its circle endpoint and nothing to this square.
3. For each edge which is incident to a square vertex in $(U_\alpha \cup V_\alpha) \setminus S$, assign a factor of $\frac{1}{\sqrt{n}}$ to the square vertex and nothing to the circle vertex.
4. For all other edges, assign a factor of $\frac{1}{\sqrt[4]{n}}$ to its square endpoint and a factor of $\frac{1}{\sqrt[6]{m}}$ to its circle endpoint.

Now each square vertex which is not in S is assigned a factor of $\frac{1}{\sqrt{n}}$ and since α is not a spider, all circle vertices are assigned a factor of $\frac{1}{\sqrt{m}}$ or smaller.

We now make this argument formal.

Let $\mathcal{C}_\alpha$ and $\mathcal{S}_\alpha$ be the set of circle vertices and the set of square vertices in α respectively. We have $n^{\frac{w(V(\alpha))-w(S_{\min})}{2}} \leq n^{0.5|\mathcal{S}_\alpha \setminus S_{min}|+(0.75-\frac{\varepsilon}{2})|\mathcal{C}_\alpha \setminus S_{min}|}$. So, it suffices to prove that

$$|E(\alpha)| - |\mathcal{S}_\alpha \setminus S_{min}| - (1.5-\varepsilon)|\mathcal{C}_\alpha \setminus S_{min}| \geq \Omega(\varepsilon|E(\alpha)|)$$

Let $Q = U_\alpha \cap V_\alpha$, $P = (U_\alpha \cup V_\alpha) \setminus Q$ and let P' be the set of vertices of P that have degree 1 and are not in S_{min}. Let E_1 be the set of edges incident to P' and let $E_2 = E(\alpha) \setminus E_1$.

For each vertex $\boxed{i}$ (resp. $\textcircled{u}$), let the number of edges of E_2 incident to it be $\deg'(\boxed{i})$ (resp. $\deg'(\textcircled{u})$). Since α is bipartite, we have that $|E_2| = \sum_{\boxed{i} \in \mathcal{S}_\alpha} \deg'(\boxed{i}) = \sum_{\textcircled{u} \in \mathcal{C}_\alpha} \deg'(\textcircled{u})$. We get that

$$|E(\alpha)| = |E_1| + |E_2| = |P'| + \frac{1}{2}\Big(\sum_{\boxed{i} \in \mathcal{S}_\alpha} \deg'(\boxed{i}) + \sum_{\textcircled{u} \in \mathcal{C}_\alpha} \deg'(\textcircled{u})\Big)$$

We also have $|S_\alpha \setminus S_{min}| \leq |P'| + |\mathcal{S}_\alpha \cap W_\alpha| + |\mathcal{S}_\alpha \cap (P \setminus P')| \leq |P'| + \frac{1}{2}\sum_{\boxed{i} \in \mathcal{S}_\alpha} \deg'(\boxed{i})$ because each square vertex outside $P' \cup Q$ has degree at least 2 and is not incident to any edge in E_1. So, it suffices to prove

$$\frac{1}{2}\sum_{\textcircled{u} \in \mathcal{C}_\alpha} \deg'(\textcircled{u}) - (1.5-\varepsilon)|\mathcal{C}_\alpha \setminus S_{min}| \geq \Omega(\varepsilon|E(\alpha)|)$$

Now, observe that each $\textcircled{u} \in \mathcal{C}_\alpha$ is incident to at most two edges in E_1. This is because if it were adjacent to at least 3 edges in E_1, then either $\textcircled{u}$ is adjacent to at least two vertices of degree 1 in U_α or $\textcircled{u}$ is adjacent to at least two vertices of degree 1 in V_α. However, this cannot happen since α is not a spider. This implies that $\deg'(\textcircled{u}) \geq \deg(\textcircled{u}) - 2$.

Note moreover that if $\textcircled{u} \in \mathcal{C}_\alpha \setminus S_{min}$, we have that $\deg'(\textcircled{u}) \geq \deg(\textcircled{u}) - 1$. This is because, building on the preceding argument, $\deg'(\textcircled{u}) = \deg(\textcircled{u}) - 2$ can only happen if there exist $\boxed{i} \in U_\alpha, \boxed{j} \in V_\alpha$ such that $(\boxed{i}, \textcircled{u}), (\boxed{j}, \textcircled{u}) \in E_1$. But then, note that we have $\boxed{i}, \boxed{j} \notin S_{min}$ by definition of P' and also, $\textcircled{u} \notin S_{min}$ by assumption. This means that there is a path from U_α to V_α which does not pass through S_{min}, which is a contradiction.

Finally, we set ε small enough such that the following inequalities are true, both of which follow from the fact that $\deg(\textcircled{u}) \geq 4$ for all $\textcircled{u} \in \mathcal{C}_\alpha$.

1. For any $\textcircled{u} \in \mathcal{C}_\alpha \cap S_{min}$, we have $\frac{\deg(\textcircled{u})-2}{2} \geq \frac{\varepsilon}{10}\deg(\textcircled{u})$.
2. For any $\textcircled{u} \in \mathcal{C}_\alpha \setminus S_{min}$, we have $\frac{\deg(\textcircled{u})-1}{2} - 1.5 + \varepsilon \geq \frac{\varepsilon}{10}\deg(\textcircled{u})$.

Using this, we get

$$\begin{aligned}\frac{1}{2}\sum_{\textcircled{u}\in\mathcal{C}_\alpha}\deg'(\textcircled{u})-(1.5-\varepsilon)|\mathcal{C}_\alpha\setminus S_{min}| &\geq \sum_{\textcircled{u}\in\mathcal{C}_\alpha\cap S_{min}}\frac{\deg(\textcircled{u})-2}{2}+\sum_{\textcircled{u}\in\mathcal{C}_\alpha\setminus S_{min}}\frac{\deg(\textcircled{u})-1}{2}-(1.5-\varepsilon)|\mathcal{C}_\alpha\setminus S_{min}|\\ &\geq \sum_{\textcircled{u}\in\mathcal{C}_\alpha\cap S_{min}}\frac{\varepsilon}{10}\deg(\textcircled{u})+\sum_{\textcircled{u}\in\mathcal{C}_\alpha\setminus S_{min}}\left(\frac{\deg(\textcircled{u})-1}{2}-1.5+\varepsilon\right)\\ &\geq \sum_{\textcircled{u}\in\mathcal{C}_\alpha\cap S_{min}}\frac{\varepsilon}{10}\deg(\textcircled{u})+\sum_{\textcircled{u}\in\mathcal{C}_\alpha\setminus S_{min}}\frac{\varepsilon}{10}\deg(\textcircled{u})\\ &= \sum_{\textcircled{u}\in\mathcal{C}_\alpha}\frac{\varepsilon}{10}\deg(\textcircled{u})=\Omega(\varepsilon|E(\alpha)|)\end{aligned}$$

■

Since $\mathcal{L}_{bool}\subseteq\mathcal{L}$, the above result extends to non-trivial non spider shapes in $\mathcal{L}_{bool}$ too.

Corollary 5.10. *If $\alpha\in\mathcal{L}_{bool}$ is not a trivial shape and not a spider, then*

$$\frac{1}{n^{|E(\alpha)|/2}}n^{\frac{w(V(\alpha))-w(S_{\min})}{2}}\leq\frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

Corollary 5.11. *If $\alpha\in\mathcal{L}$ is not a trivial shape and not a spider, then w.h.p.*

$$\frac{1}{n^{|E(\alpha)|/2}}\|M_\alpha\|\leq\frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

Proof. Using the norm bounds in Lemma A.3, we have

$$\|M_\alpha\|\leq 2\cdot(|V(\alpha)|\cdot(1+|E(\alpha)|)\cdot\log(n))^{C\cdot(|V_{rel}(\alpha)|+|E(\alpha)|)}\cdot n^{q}\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}$$

We have $W_{iso}=\emptyset$. Observe that since there are no degree 0 vertices in $V_{rel}(\alpha)$, we have that $|V_{rel}(\alpha)|\leq 2|E(\alpha)|$ and since we also have $|V(\alpha)|\cdot(1+|E(\alpha)|)\cdot\log n\leq n^{O(\tau)}$, the factor $2\cdot(|V(\alpha)|\cdot(1+|E(\alpha)|)\cdot\log(n))^{C\cdot(|V_{rel}(\alpha)|+|E(\alpha)|)}$ can be absorbed into $\frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$. The result follows from Lemma 5.7. ■

This says that nontrivial non-spider shapes have $o_n(1)$ norm (ignoring the extra factor η for the moment). We now demonstrate how to use this norm bound to control the total norm of all non-spiders in a block of $\mathcal{M}$, Corollary 5.14. We will first need a couple propositions which will also be of use to us later after we kill the spiders.

Proposition 5.12. *The number of proper shapes with at most L vertices and exactly k edges is at most $L^{8(k+1)}$.*

Proof. The following process captures all shapes (though many will be constructed multiple times):

- Choose the number of square and circle variables in each of the four sets $U\cap V, U\setminus(U\cap V), V\setminus(U\cap V), W$. This contributes a factor of L^8.
- Place each edge between two of the vertices. This contributes a factor of L^{2k}.

■

Proposition 5.13. $|\lambda_\alpha| \le \eta^{|U_\alpha|+|V_\alpha|} \cdot \frac{|E(\alpha)|^{3\cdot|E(\alpha)|}}{n^{|E(\alpha)|/2}}$ *where we assume by convention that* $0^0 = 1$.

Proof. **(Gaussian setting)** Recall that the coefficients λ_α are either zero or are defined by the formula

$$\lambda_\alpha = \eta^{|U_\alpha|+|V_\alpha|} \cdot \left(\prod_{\textcircled{u} \in V(\alpha)} h_{\deg(\textcircled{u})}(1) \right) \cdot \frac{1}{n^{|E(\alpha)|/2}} \cdot \frac{1}{\alpha!}$$

The sequence $h_k(1)$ satisfies the recurrence $h_0(1) = h_1(1) = 1, h_{k+1}(1) = h_k(1) - k h_{k-1}(1)$. We can prove by induction that $|h_k(1)| \le k^k$ and hence,

$$\prod_{\textcircled{u} \in V(\alpha)} \left| h_{\deg(\textcircled{u})}(1) \right| \le \prod_{\textcircled{u} \in V(\alpha)} (\deg(\textcircled{u}))^{\deg(\textcircled{u})} \le |E(\alpha)|^{|E(\alpha)|}.$$

(Boolean setting) In the boolean setting the coefficients λ_α are defined by

$$\lambda_\alpha = \eta^{|U_\alpha|+|V_\alpha|} \cdot \left(\prod_{\textcircled{u} \in V(\alpha)} e(\deg(\textcircled{u})) \right)$$

Using Corollary B.12, we have that $|e(k)| \le k^{3k} \cdot n^{-k/2}$. Thus,

$$|\lambda_\alpha| = \eta^{|U_\alpha|+|V_\alpha|} \cdot \prod_{\textcircled{u} \in V(\alpha)} |e(\deg(\textcircled{u}))| \le \eta^{|U_\alpha|+|V_\alpha|} \cdot \frac{|E(\alpha)|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}}.$$

■

Corollary 5.14. *For* $k, l \in \{0, 1, \ldots, D/2\}$, *let* $\mathcal{B}_{k,l} \subseteq \mathcal{L}$ *denote the set of nontrivial, non-spiders* $\alpha \in \mathcal{L}$ *on the* (k, l) *block i.e.* $|U_\alpha| = k, |V_\alpha| = l$. *The total norm of the non-spiders in* $\mathcal{B}_{k,l}$ *satisfies*

$$\sum_{\alpha \in \mathcal{B}_{k,l}} |\lambda_\alpha| \, \|M_\alpha\| = \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$$

Proof.

$$\begin{aligned}
\sum_{\alpha \in \mathcal{B}_{k,l}} |\lambda_\alpha| \, \|M_\alpha\| &\le \sum_{\alpha \in \mathcal{B}_{k,l}} \eta^{k+l} \cdot \frac{|E(\alpha)|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}} \, \|M_\alpha\| && \text{(Proposition 5.13)} \\
&\le \eta^{k+l} \cdot \sum_{\alpha \in \mathcal{B}_{k,l}} \left(\frac{|E(\alpha)|^3}{n^{\Omega(\varepsilon)}} \right)^{|E(\alpha)|} && \text{(Corollary 5.11)} \\
&\le \eta^{k+l} \cdot \sum_{\alpha \in \mathcal{B}_{k,l}} \left(\frac{n^{3\tau}}{n^{\Omega(\varepsilon)}} \right)^{|E(\alpha)|} && (\alpha \in \mathcal{L}) \\
&\le \eta^{k+l} \cdot \sum_{\alpha \in \mathcal{B}_{k,l}} \frac{1}{n^{\Omega(\varepsilon |E(\alpha)|)}} \\
&\le \eta^{k+l} \cdot \sum_{i=1}^{\infty} \frac{n^{O(\tau i)}}{n^{\Omega(\varepsilon i)}} && \text{(Proposition 5.12 and } |E(\alpha)| \ge 1 \text{ for } \alpha \in \mathcal{B}_{k,l}) \\
&= \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}} \qquad \blacksquare
\end{aligned}$$

5.2 Killing a single spider

We saw in the Proof Strategy section that the shape $2\beta_1 + \frac{1}{n}\beta_2$ lies in the nullspace of a moment matrix which satisfies the constraints "$\langle v, d_u\rangle^2 = 1$". The shape β_1 is exactly the kind of substructure that appears in a spider! Therefore it is natural to hope that if α is a left spider, then $\mathcal{M}_{fix}M_\alpha = 0$. This doesn't quite hold because $\langle v, d_u\rangle^2$ is "missing" some terms: in realizations of α, the end vertices are required to be distinct from the other squares in α, which prevents terms for all pairs i, j from appearing in the product $\mathcal{M}_{fix}M_\alpha$. There are smaller "intersection terms" (which we call collapses of α) that we can add so that the end vertices are permitted to take on all pairs i, j. After adding in these terms, we will produce a matrix L with $\mathcal{M}_{fix}L = 0$.

We first define what it means to collapse a shape into another shape by merging two vertices. Here, we only define it for merging two square vertices, since these are the only kind of merges that will happen in our analysis of intersection terms.

Definition 5.15 (Improper collapse)**.** *Let α be a shape and let $\boxed{i}, \boxed{j}$ be two distinct square vertices in $V(\alpha)$. We define the improper collapse of $\boxed{i}, \boxed{j}$ by:*

- *Remove $\boxed{i}, \boxed{j}$ from $V(\alpha)$ and replace them by a single new vertex $\boxed{k}$.*
- *Replace each edge $\{\boxed{i}, \textcircled{u}\}$ and $\{\boxed{j}, \textcircled{u}\}$, if present, by $\{\boxed{k}, \textcircled{u}\}$, keeping the same labels (note that there may be multiedges and so the new shape may not be proper).*
- *Set $U(\boxed{k}) = U(\boxed{i}) + U(\boxed{j}) \pmod 2$ and $V(\boxed{k}) = V(\boxed{i}) + V(\boxed{j}) \pmod 2$.*

Improper collapses have parallel edges, but we can convert them back to a sum of proper shapes. This is done by, for each set of parallel edges, expanding the product of Fourier characters in the Fourier basis. For example, two parallel edges with label 1 should be expanded as

$$h_1(z)^2 = (z^2 - 1) + 1 = h_2(z) + h_0(z)$$

Definition 5.16 (Collapsing a shape)**.** *Let α be a shape with two distinct square vertices $\boxed{i}, \boxed{j}$. We say that β is a (proper) collapse of $\boxed{i}, \boxed{j}$ if β appears in the expansion of the improper collapse of $\boxed{i}, \boxed{j}$.*

Remark 5.17. *If $l_1, \ldots, l_k$ are the labels of a set of parallel edges, then the product $h_{l_1}(z) \cdots h_{l_k}(z)$ is even/odd depending on the parity of $l_1 + \cdots + l_k$. Thus the nonzero Fourier coefficients will be the terms of matching parity. Therefore, in both the boolean and Gaussian cases, the shapes that are proper collapses of a given improper collapse are formed by replacing each set of parallel edges by a single edge e such that $l(e) \le l_1 + \ldots + l_k$ and $l(e) \equiv l_1 + \cdots + l_k \pmod 2$.*

Remark 5.18. *Looking at the definition and in light of the previous remark, we have the following.*

1. *The number of circle vertices does not change by collapsing a shape but the number of square vertices decreases by 1.*
2. *$\alpha \in \mathcal{L}$ has the property that the vertices have odd degree if and only if they are in $(U_\alpha \cup V_\alpha) \setminus (U_\alpha \cap V_\alpha)$. When α collapses, this property is preserved.*

We now define the desired shapes L_k which lie in the null space of $\mathcal{M}_{fix}$.

Definition 5.19. *For $k \ge 2$ define the shape ℓ_k on $\{\boxed{1}, \ldots, \boxed{k}, \textcircled{1}\}$ with two edges $\{\{\boxed{1}, \textcircled{1}\}, \{\boxed{2}, \textcircled{1}\}\}$. The left side of ℓ_k consists of $U_{\ell_k} = \{\boxed{1}, \ldots, \boxed{k}\}$. The right side consists of $V_{\ell_k} = \{\boxed{3}, \ldots, \boxed{k}, \textcircled{1}\}$.*

Definition 5.20. *Define the "completed" version L_k of ℓ_k to be the matrix which is the sum of $c_\beta M_\beta$ for β being the following shapes with coefficients:*

- *($L_{k,1}$): ℓ_k, with coefficient 2.*
- *($L_{k,2}$): If $k \geq 3$, collapse $\boxed{1}$ and $\boxed{3}$ in ℓ_k with coefficient $\frac{2}{n}$*
- *($L_{k,3}$): If $k \geq 4$, collapse $\boxed{1}$ and $\boxed{3}$, and collapse $\boxed{2}$ and $\boxed{4}$ in ℓ_k with coefficient $\frac{2}{n^2}$*
- *($L_{k,4}$): Collapse $\boxed{1}$ and $\boxed{2}$, replacing the edges by an edge with label 2, with coefficient $\frac{1}{n}$*
- *($L_{k,5}$): If $k \geq 3$, collapse $\boxed{1}$, $\boxed{2}$, and $\boxed{3}$, replacing the edges by an edge with label 2, with coefficient $\frac{1}{n}$.*

For a pictorial representation of the ribbons/shapes, see Fig. 7 below.

Lemma 5.21. $\mathcal{M}_{fix} L_k = 0$

Proof. These shapes are constructed so that if we fix a partial realization of the vertices ① and $\boxed{3}, \ldots, \boxed{k}$ as $\textcircled{u} \in \mathcal{C}_m$ and $S \in \binom{\mathcal{S}_n}{k-2}$, the squares $\boxed{1}$ and $\boxed{2}$ can still be realized as any $j_1, j_2 \in [n]$. That is, exactly the following equality holds,

$$\begin{aligned}(\mathcal{M}_{fix} L_k)_I &= \sum_{\substack{\textcircled{u} \in \mathcal{C}_m, \\ S \in \binom{\mathcal{S}_n}{k-2}}} \left(\sum_{\substack{j_1, j_2 \in [n]: \\ j_1 \neq j_2}} \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2} + \sum_{j_1 \in [n]} \widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1) \right) \\ &= \sum_{\substack{\textcircled{u} \in \mathcal{C}_m, \\ S \in \binom{\mathcal{S}_n}{k-2}}} \widetilde{\mathbb{E}}[v^I v^S (\langle v, d_u \rangle^2 - 1)] \\ &= 0\end{aligned}$$

To demonstrate how the coefficients arise, we analyze the ribbons R which L_k is composed of and see how they contribute to the output. For pictures of the ribbons/shapes, see Fig. 7 below. Let the ribbon be partially realized as $\textcircled{u}$ and $S = \{\boxed{j_3}, \ldots, \boxed{j_k}\}$. Let $(M_{fix} L_k)_{I(u,S)}$ denote the terms in $(M_{fix} L_k)_I$ with this partial realization. In this notation we want to show

$$(\mathcal{M}_{fix} L_k)_{I(u,S)} = \sum_{\substack{j_1, j_2 \in [n]: \\ j_1 \neq j_2}} \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2} + \sum_{j_1 \in [n]} \widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1).$$

1. If we take a ribbon R with $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\}$, $B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ and $E(R) = \{\{\boxed{j_1}, \textcircled{u}\}, \{\boxed{j_2}, \textcircled{u}\}\}$ where $j_1 \neq j_2$ and $j_1, j_2 \notin S$ then

   $$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2}.$$

   This ribbon must "cover" both ordered pairs (j_1, j_2) and (j_2, j_1), so we want each such ribbon R to appear with a coefficient of 2 in L_k.
2. If we take a ribbon R with $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_3}\}$, $B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ and $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}, \{\boxed{j_2}, \textcircled{u}\}\}$ where $j_1 = j_3 \in S$ then

   $$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^{S \setminus \{j_3\}} v_{j_2}] d_{uj_3} d_{uj_2} = n \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2}.$$

   Taking a coefficient of $\frac{2}{n}$ in L_k covers the two pairs (j_1, j_2) and (j_2, j_1) for this case of overlap with S.

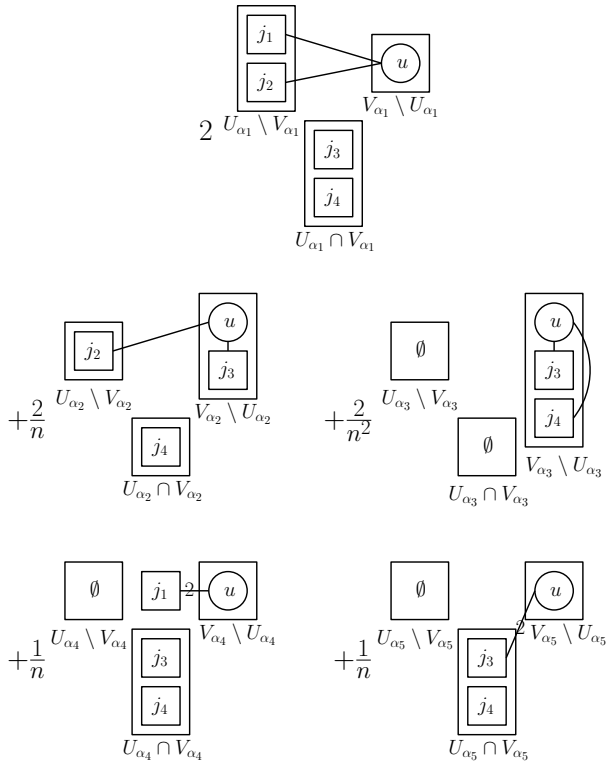

Figure 7: The five shapes that make up L_4.

3. If we take a ribbon R with $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_2}, \boxed{j_3}, \boxed{j_4}\}$, $B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ and $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}, \{\boxed{j_4}, \textcircled{u}\}\}$ where $j_1 = j_3 \in S$ and $j_2 = j_4 \in S$ then

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^{S \setminus \{j_3, j_4\}}] d_{uj_3} d_{uj_4} = n^2 \widetilde{\mathbb{E}}[v^I v^S v_{j_1} v_{j_2}] d_{uj_1} d_{uj_2}.$$

   Taking a coefficient of $\frac{2}{n^2}$ in L_k covers the two pairs (j_1, j_2) and (j_2, j_1) for this case of overlap with S.

4. If we take a ribbon R with $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_2}\}$, $B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ and $E(R) = \{\{\boxed{j_1}, \textcircled{u}\}_2\}$ where $j_1 = j_2 \notin S$ then

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^S](d_{uj_1}^2 - 1) = n \widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1).$$

   Taking a coefficient of $\frac{1}{n}$ in L_k covers these terms.

5. If we take a ribbon R with $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_2}\}$, $B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$ and $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}_2\}$ where $j_1 = j_2 = j_3 \in S$ then

$$(\mathcal{M}_{fix} M_R)_{I(u,S)} = \widetilde{\mathbb{E}}[v^I v^S](d_{uj_3}^2 - 1) = n \widetilde{\mathbb{E}}[v^I v^S v_{j_1}^2](d_{uj_1}^2 - 1).$$

   Taking a coefficient of $\frac{1}{n}$ in L_k covers these terms.

■

One of the key facts about graph matrices is that multiplication of graph matrices approximately equals a new graph matrix, $M_\alpha \cdot M_\beta \approx M_\gamma$, where γ is the result of gluing V_α with U_β (and if V_α, U_β do not have the same number of vertices of each type, the product is zero). The error terms in the approximation are intersection terms (collapses) between the variables in α and β.

Definition 5.22. *Say that shapes α and β are composable if V_α and U_β have the same number of square and circle vertices. We say a shape γ is a gluing of α and β, if the graph of γ is the disjoint union of the graphs of α and β, followed by identifying V_α and U_β under some type-preserving bijection, and if $U_\gamma = U_\alpha$ and $V_\gamma = V_\beta$.*

Proposition 5.23. *Let α, β be composable shapes. Assume that $V(\alpha) \setminus V_\alpha$ has only square vertices. Let $\{\gamma_i\}$ be the distinct gluings of α and β, and let $\widetilde{\mathcal{I}}$ be the set of improper collapses of any number of squares (possibly zero) in $V(\alpha) \setminus V_\alpha$ with distinct squares in $V(\beta) \setminus U_\beta$ in any gluing γ_i. Then there are coefficients c_γ for $\gamma \in \widetilde{\mathcal{I}}$ such that*

$$M_\alpha \cdot M_\beta = \sum_{\gamma \in \widetilde{\mathcal{I}}} c_\gamma M_\gamma.$$

Furthermore, the coefficients satisfy $|c_\gamma| \le 2^{|V(\alpha)\setminus V_\alpha|} \, |V(\gamma)|^{|V(\alpha)\setminus U_\alpha|}$.

Proof. The product $M_\alpha \cdot M_\beta$ is a matrix which is a symmetric function of the inputs $(d_1, \dots, d_m)$, the space of which is spanned by the M_γ over all possible shapes γ (not restricted to $\widetilde{\mathcal{I}}$), so there exist coefficients c_γ if we allow all shapes γ. We need to check that $M_\alpha \cdot M_\beta$ actually lies in the span of shapes in $\widetilde{\mathcal{I}}$ by showing that all ribbons in $M_\alpha \cdot M_\beta$ have shapes in $\widetilde{\mathcal{I}}$. Expanding the definition,

$$M_\alpha \cdot M_\beta = \left(\sum_{R \text{ is a ribbon of shape } \alpha} M_R \right) \left(\sum_{S \text{ is a ribbon of shape } \beta} M_S \right) = \sum_{\substack{R \text{ is a ribbon of shape } \alpha, \\ S \text{ is a ribbon of shape } \beta}} M_R M_S.$$

In order for $M_R M_S$ to be nonzero, we require $B_R = A_S$ as sets; R may assign the labels arbitrarily inside B_R, resulting in different gluings of α and β. Fix R and S, and let γ be the corresponding gluing of α and β for this R and S.

The matrix $M_R M_S$ has one nonzero entry; we claim that it is a Fourier character for a ribbon T which is a collapse of γ. The labels of R outside of B_R can possibly overlap with the labels of S outside of A_S, and naturally the shape of T is the result of collapsing vertices in γ with the same label.

To bound the coefficients c_γ that appear, it suffices to bound the coefficient on a ribbon M_T, which is bounded by the number of contributing ribbons R, S, where we say ribbons R of shape α and S of shape β contribute to T if $M_R M_S = M_T$. From T, we can completely recover the sets A_R and B_S. The labels of $V(R) \setminus A_R$ must be among the labels of T; choose them in at most $|V(\gamma)|^{|V(\alpha)\setminus U_\alpha|}$ ways. This also determines $B_R = A_S$. All that remains is to determine the graph structure of S. Since improper collapsing doesn't lose any edges, knowing the labels of R we know exactly which edges of T must come from R and S. The vertices $V(T) \setminus V(R)$ must come from S, as must B_R; pick a subset of $V(R) \setminus B_R$ to include in $2^{|V(\alpha)\setminus V_\alpha|}$ ways. ∎

Let α be a left spider with end vertices $\boxed{i}, \boxed{j}$ which are adjacent to a circle $\textcircled{u}$. Recall that our goal is to argue that $\mathcal{M} M_\alpha \approx 0$. To get there, we can try and factor M_α across the vertex separator $S = U_\alpha \cup \{\textcircled{u}\} \setminus \{\boxed{i}, \boxed{j}\}$ which separates α into

$$M_\alpha \approx L_{|U_\alpha|} \cdot M_{\text{body}(\alpha)}$$

where we have defined,

Definition 5.24. *Let α be a left spider with end vertices $\boxed{i}, \boxed{j}$. Define* $\text{body}(\alpha)$ *as the shape whose graph is α with $\boxed{i}$ and $\boxed{j}$ deleted and with $U_{\text{body}(\alpha)} = U_\alpha \cup \{\textcircled{u}\} \setminus \{\boxed{i}, \boxed{j}\}, V_{\text{body}(\alpha)} = V_\alpha$. The definition is analogous for right spiders.*

Due to Lemma 5.21, the right-hand side of the approximation is in the null space of $\mathcal{M}$. We now formalize this approximate factorization.

Definition 5.25. *Let α be a spider with end vertices $\boxed{i}, \boxed{j}$. Define $\widetilde{\mathcal{I}}_\alpha$ to be the set of shapes that can be obtained from α by performing at least one of the following steps:*

- *Improperly collapse $\boxed{i}$ with a square vertex in α*
- *Improperly collapse $\boxed{j}$ with a square vertex in α*

Let $\mathcal{I}_\alpha$ be the set of proper shapes that can be obtained via the same process but using proper collapses.

In the above definition, we allow $\boxed{i}, \boxed{j}$ to collapse with two distinct squares, or to collapse together, or to both collapse with a common third vertex. For technical reasons we need to work with a refinement of $\mathcal{I}_\alpha$ into two sets of shapes and use tighter bounds on coefficients of one set.

Definition 5.26. *Let $\mathcal{I}_\alpha^{(1)}$ be the set of shapes that can be obtained from α by performing at least one of the following steps:*

- *Collapse $\boxed{i}$ with a square vertex in $\text{body}(\alpha) \setminus U_\alpha$*
- *Collapse $\boxed{j}$ with a square vertex in $\text{body}(\alpha) \setminus U_\alpha$ (distinct from $\boxed{i}$'s collapse if it happened)*

Let $\mathcal{I}_\alpha^{(2)} := \mathcal{I}_\alpha \setminus \mathcal{I}_\alpha^{(1)}$ and define the improper versions $\widetilde{\mathcal{I}}_\alpha^{(1)}, \widetilde{\mathcal{I}}_\alpha^{(2)}$ analogously.

Lemma 5.27. *Let α be a left spider with end vertices $\boxed{i}, \boxed{j}$. There are coefficients c_β for $\beta \in \widetilde{\mathcal{I}}_\alpha$ such that*

$$L_{|U_\alpha|} \cdot M_{\text{body}(\alpha)} = 2M_\alpha + \sum_{\beta \in \widetilde{\mathcal{I}}_\alpha} c_\beta M_\beta,$$

$$|c_\beta| \leq \begin{cases} 40\,|V(\alpha)|^3 & \beta \in \widetilde{\mathcal{I}}_\alpha^{(1)} \\ \frac{40|V(\alpha)|^3}{n} & \beta \in \widetilde{\mathcal{I}}_\alpha^{(2)} \end{cases}.$$

Proof. First, we can check that the coefficient of M_α is 2. Only the ℓ_k term of L_k has the full number of squares, and it has a factor of 2 in L_k.

The shapes in $\widetilde{\mathcal{I}}_\alpha$ are definitionally the intersection terms that appear in this graph matrix product, and furthermore the shapes in $\widetilde{\mathcal{I}}_\alpha$ are definitionally the intersection terms for the ℓ_k term. Using Proposition 5.23, for each of the five shapes in $L_{|U_\alpha|}$ the coefficient it contributes is bounded by $4\,|V(\alpha)|^3$. The coefficient on ℓ_k is 2, so the coefficients for $\widetilde{\mathcal{I}}_\alpha^{(1)}$ are at most $8\,|V(\alpha)|^3$. The maximum coefficient of the other four shapes in $L_{|U_\alpha|}$ is $\frac{2}{n}$, so their total contribution to coefficients on $\widetilde{\mathcal{I}}_\alpha^{(2)}$ is at most $\frac{32|V(\alpha)|^3}{n}$. ∎

We now want to turn our improper shapes into proper ones from $\mathcal{I}_\alpha$. Unfortunately it is not quite true that to expand an improper shape, one can just expand each edge individually (though this is true for improper ribbons). There is an additional difficulty that arises due to ribbon symmetries. To see the difficulty, consider the example given in Fig. 8 below.

One would expect both coefficients on the right shapes to be 1 since $h_1(z)^2 = h_2(z) + h_0(z)$. However, in the left shape, the two circles are distinguishable, hence summing over all ribbons includes one with $w_1 = i, w_2 = j$ and a second with $w_1 = j, w_2 = i$. On the top right shape, the circles are indistinguishable, hence the graph/ribbon where the circles are assigned $\{i, j\}$ is counted twice. On the bottom right shape, the circles are distinguishable, so all ribbons are summed once. To bound the new coefficients, we use the concept of shape automorphisms.

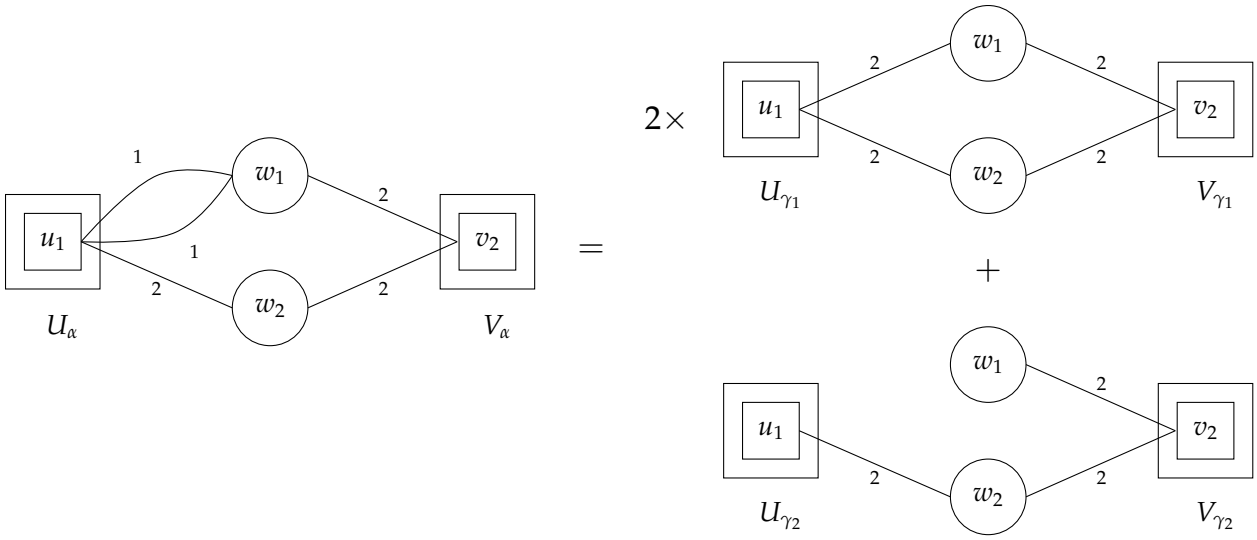

Figure 8: A surprising equality of graph matrices.

Definition 5.28. *An automorphism of a shape α is a function $\varphi : V(\alpha) \to V(\alpha)$ that preserves the sets U_α, V_α and is an automorphism of the underlying edge-labeled graph. Let* $\mathrm{Aut}(\alpha)$ *denote the automorphism group of α.*

Proposition 5.29. *Let α be an improper shape, and let $\mathcal{P}$ be the set of proper shapes that can be obtained by expanding α. Then there are coefficients $|c_\gamma| \le C_{Fourier} \cdot C_{Aut}$ such that*

$$M_\alpha = \sum_{\gamma \in \mathcal{P}} c_\gamma M_\gamma$$

where $C_{Fourier}$ is a bound on the magnitude of Fourier coefficients in the expansion and $C_{Aut} = \max_{\gamma \in \mathcal{P}} \frac{|\mathrm{Aut}(\gamma)|}{|\mathrm{Aut}(\alpha)|}$.

Proof. The number of realizations of a graph matrix giving a particular ribbon is exactly the number of automorphisms, therefore

$$M_\alpha = \frac{1}{|\mathrm{Aut}(\alpha)|} \sum_{\text{realizations } \sigma} M_{\sigma(\alpha)}$$

Expand each improper ribbon $M_{\sigma(\alpha)}$ into proper ribbons with coefficients at most $C_{Fourier}$. Because the realizations of α and any γ are the same, this exactly sums over all γ and all realizations of γ. The Fourier coefficient on each realization of γ is the same; let it be c'_γ with $|c'_\gamma| \le C_{Fourier}$. Continuing,

$$\begin{aligned} &= \frac{1}{|\mathrm{Aut}(\alpha)|} \sum_{\gamma \in \mathcal{P}} c'_\gamma \sum_{\text{realizations } \sigma} M_{\sigma(\gamma)} \\ &= \sum_{\gamma \in \mathcal{P}} c'_\gamma \frac{|\mathrm{Aut}(\gamma)|}{|\mathrm{Aut}(\alpha)|} M_\gamma \end{aligned}$$

■

Proposition 5.30. *Let $l_1 \leq \cdots \leq l_k \in \mathbb{N}$ and let $L = l_1 + \cdots + l_k$. Assume $L \geq 1$. In the Fourier expansion of $h_{l_1}(z) \cdots h_{l_k}(z)$, the maximum coefficient is bounded in magnitude by $(2L)^{L-l_k}$.*

Proof. In the boolean case, the coefficient is 1. In the Gaussian case, the "linearization coefficient" of $h_p(z)$ in this product is given by orthogonality to be

$$\frac{\mathbb{E}_{z\sim\mathcal{N}(0,1)}[h_{l_1}(z)\cdots h_{l_k}(z)\cdot h_p(z)]}{\mathbb{E}_{z\sim\mathcal{N}(0,1)}[h_p^2(z)]} = \frac{\mathbb{E}_{z\sim\mathcal{N}(0,1)}[h_{l_1}(z)\cdots h_{l_k}(z)\cdot h_p(z)]}{p!}$$

A formula from, e.g., [RW97, Example G (Continued)] shows that $\mathbb{E}[h_{l_1}\cdots h_{l_k}\cdot h_p]$ equals the number of "block perfect matchings": perfect matchings on $l_1 + \cdots + l_k + p$ elements divided into blocks of size l_i or p such that no two elements from the same block are matched. Bound the number of block perfect matchings by:

- Pick a partial function from blocks $l_1, \ldots, l_{k-1}$ to $[L]$ in at most $(L+1)^{L-l_k}$ ways.
- If this forms a valid partial matching and there are p unmatched elements remaining, match them with the elements from the block of size p in $p!$ ways.

Therefore the coefficient is bounded by $(L+1)^{L-l_k} \leq (2L)^{L-l_k}$. ∎

Proposition 5.31. *For a shape α, let $\alpha \pm e$ denote the shape with edge e added or deleted. Then*

$$\frac{|\mathrm{Aut}(\alpha \pm e)|}{|\mathrm{Aut}(\alpha)|} \leq |V(\alpha)|^2.$$

Proof. We show that the two groups have a large subgroup which are equal. Consider $\mathrm{Aut}(\alpha \pm e)$ and $\mathrm{Aut}(\alpha)$ as group actions on the set $\binom{V(\alpha)}{2}$. Letting G^e denote the stabilizer of edge e, observe that $\mathrm{Aut}(\alpha \pm e)^e = \mathrm{Aut}(\alpha)^e$. By the orbit-stabilizer lemma, the index $|G : G^e|$ is equal to the size of the orbit of e, which is at least 1 and at most $|V(\alpha)|^2$. So,

$$\frac{|\mathrm{Aut}(\alpha \pm e)|}{|\mathrm{Aut}(\alpha)|} = \frac{|\mathrm{Aut}(\alpha \pm e) : \mathrm{Aut}(\alpha \pm e)^e|}{|\mathrm{Aut}(\alpha) : \mathrm{Aut}(\alpha)^e|} \leq |V(\alpha)|^2. \qquad ∎$$

Lemma 5.32. *If α is a left spider, there are coefficients c_β for each $\beta \in \mathcal{I}_\alpha$ such that*

$$L_{|U_\alpha|} \cdot M_{\mathrm{body}(\alpha)} = 2M_\alpha + \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta,$$

$$|c_\beta| \leq \begin{cases} 160\,|V(\alpha)|^7\,|E(\alpha)|^2 & \beta \in \mathcal{I}_\alpha^{(1)} \\ \frac{160|V(\alpha)|^7|E(\alpha)|^2}{n} & \beta \in \mathcal{I}_\alpha^{(2)} \end{cases}.$$

Proof. We express each $M_\beta, \beta \in \widetilde{\mathcal{I}}_\alpha$ in Lemma 5.27 in terms of proper shapes. We apply Proposition 5.29 using the following bounds on $C_{Fourier}$ and C_{Aut}. The only improperness in β comes from collapsing (at most) the two end vertices, which have a single incident edge each. Therefore the set of labels of any parallel edges is either $\{1, k\}$ or $\{1, 1, k\}$, for some $k \leq |E(\alpha)|$. By Proposition 5.30, we have $C_{Fourier} \leq 4\,|E(\alpha)|^2$. There are at most two extra parallel edges in β, so we have $C_{Aut} \leq |V(\alpha)|^4$ using Proposition 5.31. Therefore the coefficients increase by at most $C_{Fourier} \cdot C_{Aut} \leq 4\,|E(\alpha)|^2\,|V(\alpha)|^4$. ∎

Corollary 5.33. *If α is a right spider, there are coefficients c_β with the same bounds given in Lemma 5.32 such that*

$$M_{\text{body}(\alpha)} \cdot L^{\intercal}_{|U_\alpha|} = 2M_\alpha + \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta.$$

Corollary 5.34. *If $x \perp \text{Null}(\mathcal{M}_{fix})$ and α is a spider, then for some c_β with the same bounds given in Lemma 5.32,*

$$x^\top (M_\alpha - \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta) x = 0$$

Proof. For a left spider, since

$$\mathcal{M}_{fix}(2M_\alpha + \sum_{\beta \in \mathcal{I}_\alpha} c_\beta M_\beta) = \mathcal{M}_{fix} \cdot L_{|U_\alpha|} \cdot M_{\alpha'} = 0$$

we are in position to use Fact 3.1. For a right spider, the proof is analogous. ∎

5.3 Killing all the spiders

The strategy is to start with the moment matrix $\mathcal{M}$ and apply Corollary 5.34 repeatedly until we end up with no spiders in our decomposition. For each spider, killing it via Corollary 5.34 leaves only intersection terms. Some of those intersection terms may themselves be smaller spiders, in which case we will apply the corollary again and again until only non-spiders remain. The difficulty during this procedure is to bound the total coefficient accumulated on each non-spider. To capture this process, we define the web of a spider α, which will be a directed acyclic graph that will capture the spider killing process. For the sake of distinction, we will call the vertices of this graph "nodes".

Definition 5.35 (Web of α)**.** *The web $W(\alpha)$ of a spider α is a rooted directed acyclic graph (DAG) whose nodes are shapes and whose root is α. Each spider node γ has edges to nodes β for each shape $\beta \in \mathcal{I}_\gamma$. The non-spider nodes are leaves/sinks of the DAG.*

Remark 5.36. *The DAG structure arises because each shape in $\mathcal{I}_\gamma$ has strictly fewer square vertices than γ for any spider γ. As a consequence, the height of a web $W(\alpha)$ is at most $|V(\alpha)|$.*

Each node γ of $W(\alpha)$ also has an associated value v_γ, which is defined by the following process:

- Initially, set $v_\alpha = 1$ and for all other γ, set $v_\gamma = 0$.
- Starting from the root and in topological order, each spider node γ adds $v_\gamma c_\beta$ to v_β for each child $\beta \in \mathcal{I}_\gamma$, where the c_β are the coefficients from Corollary 5.34.

Proposition 5.37. *If $x \perp \text{Null}(\mathcal{M}_{fix})$, then*

$$x^\intercal (M_\alpha - \sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma) x = 0.$$

Proof. Start with the equation $x^\intercal M_\alpha x = x^\intercal v_\alpha M_\alpha x$. In each step, we take the topologically first spider γ, which in this case means the spider closest to the root of $W(\alpha)$, that is present in the right

hand side of our equation and using Corollary 5.34, we replace $v_\gamma M_\gamma$ by $\sum_{\beta \in \text{children}(\gamma)} v_\gamma c_\beta M_\beta$. Precisely by the definition of the v_γ, this process ends with the equation

$$x^\intercal M_\alpha x = x^\intercal \Big(\sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma \Big) x$$

■

Proposition 5.38. *For any node β in $W(\alpha)$, $|\text{parents}(\beta)| \leq 4\,|V(\alpha)|^3 \cdot |E(\alpha)|^2$ where parents(β) is the set of nodes γ in $W(\alpha)$ such that $\beta \in \mathcal{I}_\gamma$.*

Proof. The following process covers all parent left spiders γ which could possibly collapse their end vertices to form β. Starting from $\gamma = \beta$,

- Pick a circle vertex $\textcircled{u} \in V(\gamma)$ to be the neighbor of the end vertices.
- Pick a square vertex $\boxed{i} \in V(\gamma)$ to be the collapse of the first end vertex. "Uncollapse" it by adding a new square to U_γ with a single edge to $\textcircled{u}$ with label 1. Flip the value of $U_\gamma(\boxed{i})$. Modify the label of $\{\boxed{i}, \textcircled{u}\}$ to any number up to $|E(\alpha)|$.
- Pick a square vertex $\boxed{j} \in V(\gamma)$ to be the second end vertex. Optionally uncollapse it by adding a new square to γ in the same way as above.

The process can be carried out in at most $|V(\alpha)|^3\,|E(\alpha)|\,(|E(\alpha)|+1) \leq 2\,|V(\alpha)|^3\,|E(\alpha)|^2$ ways. We multiply by 2 to accommodate right spiders. ■

Let us label each parent-child edge (γ, β) as either a "type 1" edge if $\beta \in \mathcal{I}_\gamma^{(1)}$ or a "type 2" edge if $\beta \in \mathcal{I}_\gamma^{(2)}$.

Proposition 5.39. *Let p be a path in $W(\alpha)$ with $\#_1(p)$ type 1 edges and $\#_2(p)$ type 2 edges. Then $\#_1(p) \leq |E(\alpha)| + 2\#_2(p)$.*

Proof. For a shape γ, let S_γ be the set of square vertices in γ. Then, $S_\gamma \cap W_\gamma$ will be the set of middle vertices of γ which are squares. We claim that the quantity $|\mathcal{S}_\gamma \cap W_\gamma| + |U_\gamma \setminus (U_\gamma \cap V_\gamma)| + |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$ decreases during a collapse.

Fix a pair of consecutive shapes (γ, β) which form a type 1 edge. Looking at the definition of $\mathcal{I}_\gamma^{(1)}$, each end vertex either collapses with (1) nothing, or (2) a vertex of W_γ, or (3) a vertex from $V_\gamma \setminus U_\gamma$ (if γ is a left spider; for a right spider, $U_\gamma \setminus V_\gamma$). Furthermore, case (2) or (3) must occur for at least one of the end vertices and also, they do not collapse together.

If case (2) occurs, then $\big|\mathcal{S}_\beta \cap W_\beta\big| < |\mathcal{S}_\gamma \cap W_\gamma|$ while $\big|U_\beta \setminus (U_\beta \cap V_\beta)\big| = |U_\gamma \setminus (U_\gamma \cap V_\gamma)|$ and $\big|V_\beta \setminus (U_\beta \cap V_\beta)\big| = |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$. If case (3) occurs, then $W_\beta = W_\gamma$ while $\big|U_\beta \setminus (U_\beta \cap V_\beta)\big| < |U_\gamma \setminus (U_\gamma \cap V_\gamma)|$ and $\big|V_\beta \setminus (U_\beta \cap V_\beta)\big| < |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$. In all cases, $\big|\mathcal{S}_\beta \cap W_\beta\big| + \big|U_\beta \setminus (U_\beta \cap V_\beta)\big| + \big|V_\beta \setminus (U_\beta \cap V_\beta)\big| < |\mathcal{S}_\gamma \cap W_\gamma| + |U_\gamma \setminus (U_\gamma \cap V_\gamma)| + |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$ as desired.

Now we bound this expression for α. From the definition of $\mathcal{L}$, Definition 4.6, for spiders appearing in the pseudocalibration, the square vertices in W_α, $U_\alpha \setminus (U_\alpha \cap V_\alpha)$ and $V_\alpha \setminus (U_\alpha \cap V_\alpha)$ have degree at least 1 and can only be connected to circle vertices. Therefore their number is bounded by $|E(\alpha)|$. Hence, initially $|\mathcal{S}_\alpha \cap W_\alpha| + |U_\alpha \setminus (U_\alpha \cap V_\alpha)| + |V_\alpha \setminus (U_\alpha \cap V_\alpha)| \leq |E(\alpha)|$.

Finally, each type 2 edge in p can only increase $|\mathcal{S}_\gamma \cap W_\gamma| + |U_\gamma \setminus (U_\gamma \cap V_\gamma)| + |V_\gamma \setminus (U_\gamma \cap V_\gamma)|$ by at most 2. Therefore, we have the desired inequality $\#_1(p) \leq |E(\alpha)| + 2\#_2(p)$. ■

Corollary 5.40. $\#_2(p) \geq \frac{|p|}{3} - \frac{|E(\alpha)|}{3}$.

Proof. Plug in $|p| = \#_1(p) + \#_2(p)$ and rearrange. ∎

Finally, we can bound the accumulation on each non-spider by a term which only depends on the parameters of the spider α.

Lemma 5.41. *There are absolute constants C_1, C_2 so that for all leaves γ of $W(\alpha)$,*

$$|v_\gamma| \leq (C_1 \cdot |V(\alpha)| \cdot |E(\alpha)|)^{C_2|E(\alpha)|}.$$

Proof. To bound $|v_\gamma|$ we will sum the contributions of all paths $p = (\beta_0 = \alpha, \ldots, \beta_r = \gamma)$ in $W(\alpha)$ starting from α and ending at γ. This path contributes a product of coefficients c_β towards v_γ.

Remark 5.42. *Here it is important that type 2 edges have stronger bounds on their coefficients $|c_\beta| \leq C \cdot (|V(\alpha)|\,|E(\alpha)|)^{O(1)}/n \ll 1$.*

Before we proceed with the proof we establish some convenient notation and recall some facts. For consecutive shapes β_{i-1}, β_i (i.e., β_i is a child of β_{i-1}), we denote by c_{β_i} the coefficient from Corollary 5.34 applied on β_{i-1}. By Proposition 5.38, the in-degree of $W(\alpha)$ can be bounded as $B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}$ for some constants B_1, B_2. Thus, the number of paths of length r ending at γ is at most $(B_1\,|V(\alpha)|\,|E(\alpha)|)^{B_2 r}$. Using Corollary 5.34, set B_1, B_2 large enough so that c_{β_i} is at most $B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}$ for a type 1 edge (resp. $B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}/n$ for a type 2 edge).

$$\begin{aligned}
|v_\gamma| &\leq \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \prod_{i=1}^{r} \left|c_{\beta_i}\right| \\
&\leq \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{\#_1(p)} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}/n\right)^{\#_2(p)} && \text{(Corollary 5.34)} \\
&\leq \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{|E(\alpha)|+2\#_2(p)} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}/n\right)^{\#_2(p)} && \text{(Proposition 5.39)} \\
&= \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{|E(\alpha)|} \left(B_1' \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2'}/n\right)^{\#_2(p)}
\end{aligned}$$

for some constants B_1', B_2'. We split the above sum into two sums, $r \leq 3|E(\alpha)|$ and $r > 3|E(\alpha)|$. For $r \leq 3\,|E(\alpha)|$, upper bounding the $\#_2(p)$ term by 1 and upper bounding the number of paths by $(B_1\,|V(\alpha)|\,|E(\alpha)|)^{B_2 r}$ gives a bound of $(B_1''\,|V(\alpha)|\,|E(\alpha)|)^{B_2''|E(\alpha)|}$ for some constants B_1'', B_2''. For larger r, we lower bound $\#_2(p) \geq r/9 = |E(\alpha)|/3$ using Corollary 5.40. Applying the same bound on the number of paths, the total contribution of the terms corresponding to larger r is bounded by 1 using the power of n in the denominator (assuming δ, τ are small enough). ∎

We define the result of all this spider killing to be a new matrix $\mathcal{M}^+$.

Definition 5.43. *Define the matrix $\mathcal{M}^+$ as the result of killing all the spiders,*

$$\mathcal{M}^+ := \mathcal{M} - \sum_{\text{spiders } \alpha} \lambda_\alpha \left(M_\alpha - \sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma \right)$$

5.4 Finishing the proof

The final step of the proof is to argue that, after the spider killing process is completed, the newly created non-spider terms in $\mathcal{M}^+$ also have small norm. Towards this, we would like to prove a statement similar to Corollary 5.11. In that proof, we used special structural properties of the non-spiders in $\mathcal{L}$ to prove that non-spiders in the pseudocalibration were negligible. But now, the non-spiders in $\mathcal{M}^+$ need not have the properties of $\mathcal{L}$ – for instance, there could be circle vertices of degree 2 or isolated vertices. To handle the potentially larger norms, we will use that the coefficients of these new non-spider terms β come with the coefficients λ_α of the spider terms α in whose web they lie. Since α has more vertices/edges than β, the power of $\frac{1}{n}$ in λ_α is larger than the "expected pseudocalibration" coefficient of $\eta^{|U_\beta|+|V_\beta|} \cdot \frac{1}{n^{|E(\beta)|/2}}$. We prove that these extra factors of $\frac{1}{n}$ are enough to overpower isolated vertices or a smaller vertex separator using a careful charging argument.

Lemma 5.44. *If β is a nontrivial non-spider and $\beta \in W(\alpha)$ for some spider $\alpha \in \mathcal{L}$, then*

$$\eta^{|U_\alpha|+|V_\alpha|} \cdot \frac{1}{n^{|E(\alpha)|/2}} \cdot n^{\frac{w(V(\beta))-w(S_{\min})+w(W_{iso})}{2}} \leq \eta^{|U_\beta|+|V_\beta|} \cdot \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$$

where S_{min} and W_{iso} are the minimum vertex separator of β and the set of isolated vertices of $V(\beta) \setminus (U_\beta \cup V_\beta)$ respectively.

Proof. We start by giving the idea of the proof. Suppose we try to use the same distribution scheme as in the proof of Lemma 5.7. It doesn't work for two reasons. Firstly, the circle vertices in β still have even degree, which follows from Remark 5.18, but now, they could have degrees 0 or 2. For the previous distribution scheme to go through, we needed them to have degree at least 4 which gave the necessary edge decay to handle the norm bounds. Secondly, the square vertices can now have degree 0 hence getting no decay from the edges.

The first issue is relatively easy to handle. Since β was obtained by collapsing α, the circle vertices of degrees 0 or 2 in β must have had degree at least 4 in α to begin with. Hence, we can fix a particular sequence of collapses from α to β and then assume for the sake of analysis that the removed edges are still present. In this case, the same charging argument as in Lemma 5.7 would go through. This is made formal by looking at the sequence of improper collapses of this chain of collapses.

To handle the second issue, let's analyze more carefully how degree 0 square vertices appear. Fix a sequence of collapses from α to β and consider a specific step where γ collapsed to γ' and a square vertex of degree 0 was formed. Let the two square vertices that collapsed in γ be $\boxed{i}, \boxed{j}$ and let the square vertex of degree 0 that formed in γ' be $\boxed{k}$. In light of Remark 5.18, since $\boxed{k}$ has degree 0, it must not be in $(U_{\gamma'} \cup V_{\gamma'}) \setminus (U_{\gamma'} \cap V_{\gamma'})$ and hence, $U_{\gamma'}(\boxed{k}) = V_{\gamma'}(\boxed{k}) = 0$ or $U_{\gamma'}(\boxed{k}) = V_{\gamma'}(\boxed{k}) = 1$. But in the latter case, this vertex does not contribute to norm bounds since it's in $U_{\gamma'} \cap V_{\gamma'}$ so it can be safely disregarded. Note that it doesn't have to stay in this set since future collapses might collapse this vertex, but this is not a problem as we can charge for this collapse if it happens.

So, assume we have $U_{\gamma'}(\boxed{k}) = V_{\gamma'}(\boxed{k}) = 0$. But by the definition of collapse, at least one of $\boxed{i}$ or $\boxed{j}$ must have been in $U_\gamma \setminus (U_\gamma \cap V_\gamma)$ or $V_\gamma \setminus (U_\gamma \cap V_\gamma)$. Also from the definition of collapse, we have $U_{\gamma'}(\boxed{k}) = U_\gamma(\boxed{i}) + U_\gamma(\boxed{j}) \pmod 2$ and $V_{\gamma'}(\boxed{k}) = V_\gamma(\boxed{i}) + V_\gamma(\boxed{j}) \pmod 2$. Putting these together, we immediately get that the only way this could have happened is if either $\boxed{i}, \boxed{j} \in U_\gamma \setminus (U_\gamma \cap V_\gamma)$ or if $\boxed{i}, \boxed{j} \in V_\gamma \setminus (U_\gamma \cap V_\gamma)$.

When such a collapse happens, observe that $|U_\gamma| + |V_\gamma| \geq |U_{\gamma'}| + |V_{\gamma'}| + 2$. This is precisely where the decay from our normalization factor $\eta = \frac{1}{\sqrt{n}}$ kicks in. This inequality means that an extra decay factor of $\eta^2 = \frac{1}{n}$ is available to us when we compare to the "expected pseudocalibration" coefficient of β. We will use this factor to charge the new square vertex of degree 0.

We now make these ideas formal.

Let $Q = U_\beta \cap V_\beta$, $P = (U_\beta \cup V_\beta) \setminus Q$ and let P' be the set of degree 1 square vertices in β that are not in S_{min}. Let s_0 be the number of degree 0 square vertices in $V(\beta) \setminus Q$. All the square vertices outside $P' \cup Q \cup S_{min}$ have degree at least 2, let there be $s_{\geq 2}$ of them.

Because of parity constraints, Remark 5.18, and because there are no circle vertices in $U_\beta \cup V_\beta$, all circle vertices have even degree in β. Let c_0 be the number of degree 0 circle vertices in β. Let $c_2, c_{\geq 4}$ be the number of degree 2 circle vertices and the number of circle vertices of degree at least 4 in $V(\beta) \setminus S_{min}$ respectively. Then, we have

$$n^{\frac{w(V(\beta)) - w(S_{\min}) + w(W_{iso})}{2}} \leq n^{\frac{|P'| + s_{\geq 2} + (1.5-\varepsilon)(c_2 + c_{\geq 4})}{2}} \cdot n^{s_0 + (1.5-\varepsilon)c_0}$$

Using $\eta = \frac{1}{\sqrt{n}}$, it suffices to show

$$|E(\alpha)| + (|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta|) \geq |P'| + s_{\geq 2} + (1.5-\varepsilon)(c_2 + c_{\geq 4}) + 2s_0 + 2(1.5-\varepsilon)c_0 + \Omega(\varepsilon\,|E(\alpha)|)$$

There can be many ways to collapse α to β, fix any one. We first use a charging argument for the degree 0 square vertices.

Lemma 5.45. $|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta| \geq 2s_0$

Proof. In the collapse process, in each step, a vertex $\boxed{i} \in U_\gamma \setminus (U_\gamma \cap V_\gamma)$ or $\boxed{i} \in V_\gamma \setminus (U_\gamma \cap V_\gamma)$ of degree 1 in an intermediate shape γ collapses with another square vertex $\boxed{k}$. We have that $|U_\gamma| + |V_\gamma|$ decreases precisely when $\boxed{i}$ collapses with $\boxed{k} \in U_\gamma$ (resp. $\boxed{k} \in V_\gamma$). In either case, the quantity decreases by exactly 2 which we allocate to this new merged vertex. Each degree 0 square vertex in $V(\beta) \setminus Q$ must have arisen from a collapse, and hence must have had at least an additive quantity of 2 allocated to it. This proves that $|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta| \geq 2s_0$. ∎

We will now prove a structural lemma.

Lemma 5.46. *Any vertex* $\textcircled{u}$ *that has degree at least* 2 *in* $V(\beta) \setminus S_{min}$ *is adjacent to at most* 1 *vertex of* P'.

Proof. Observe that $\textcircled{u}$ cannot be adjacent to 3 vertices in P' because otherwise, at least 2 of them would be in $U_\beta \setminus Q$ or in $V_\beta \setminus Q$ which means β would be a spider which is a contradiction. If $\textcircled{u}$ is adjacent to 2 vertices in P', then one of them is in $U_\beta \setminus Q$ and the other is in $V_\beta \setminus Q$ respectively. Since both of these vertices are not in S_{min}, it follows that $\textcircled{u}$ is in S_{min} since there is no path from U_β to V_β that doesn't pass through S_{min}. This is a contradiction. Therefore, $\textcircled{u}$ is adjacent to at most 1 vertex in P'. ∎

This lemma immediately implies $|P'| \leq c_2 + c_{\geq 4}$.

To account for edges of α that are not in β, we let $\widetilde{\beta}$ be the result of improperly collapsing α to β; note that $|E(\alpha)| = \left|E(\widetilde{\beta})\right|$. We call the edges that disappeared when properly collapsing "phantom" edges. Let $\deg_{\widetilde{\beta}}(\boxed{i})$ (resp. $\deg_{\widetilde{\beta}}(\textcircled{u})$) denote the degree of vertex $\boxed{i}$ (resp. $\textcircled{u}$) in $\widetilde{\beta}$. Observe that any circle vertex $\textcircled{u}$ in $V(\beta)$ has $deg_{\widetilde{\beta}}(\textcircled{u}) \geq 4$.

Lemma 5.47. $|E(\alpha)| \geq |P'| + s_{\geq 2} + (1.5-\varepsilon)(c_2 + c_{\geq 4}) + 2(1.5-\varepsilon)c_0 + \Omega(\varepsilon\,|E(\alpha)|)$

Proof. We will use the following charging scheme. Each edge of β incident on P' allocates 1 to the incident square vertex, which is in P'. Every other edge of β allocates $\frac{1}{2}$ to the incident square vertex and $\frac{1}{2} - \frac{\varepsilon}{10}$ to the incident circle vertex. Each phantom edge allocates $1 - \frac{\varepsilon}{10}$ to the incident circle vertex $\textcircled{u}$. So, a total of $\frac{\varepsilon}{10}(|E(\alpha)| - |P'|)$ has not been allocated.

All square vertices in P' have been allocated a value of 1. And observe that all square vertices of degree at least 2 in β have been allocated at least 1 from the incident edges of β, for a total value of $s_{\geq 2}$. So, the square vertices get a total allocation of at least $|P'| + s_{\geq 2}$.

Consider any degree-0 circle vertex $\textcircled{u}$ in $V(\beta)$. It must be incident to at least 4 phantom edges and hence, must be allocated at least a value of $4(1 - \frac{\varepsilon}{10}) > 2(1.5-\varepsilon)$. Hence, the degree-0 circle vertices in $V(\beta)$ have a total allocation of at least $2(1.5-\varepsilon)c_0$.

Suppose the degree of $\textcircled{u}$ in $V(\beta)$ is 2. Then, it is incident on at least 2 phantom edges. By Lemma 5.46, it is also adjacent to at most one vertex of P' and so, must have been allocated a value of at least $2(1 - \frac{\varepsilon}{10}) + (deg_{\widetilde{\beta}}(\textcircled{u}) - 3)(\frac{1}{2} - \frac{\varepsilon}{10})$. This is at least $1.5 - \varepsilon + \frac{\varepsilon}{10}$.

Suppose the degree of $\textcircled{u}$ in $V(\beta)$ is at least 4. By Lemma 5.46, it is adjacent to at most one vertex of P'. Then it must have been allocated a value of at least $(deg_{\widetilde{\beta}}(\textcircled{u}) - 1)(\frac{1}{2} - \frac{\varepsilon}{10})$. Using $deg_{\widetilde{\beta}}(\textcircled{u}) \geq 4$, this is at least $1.5 - \varepsilon + \frac{\varepsilon}{10}$.

This implies

$$|E(\alpha)| \geq |P'| + s_{\geq 2} + 2(1.5-\varepsilon)c_0 + (1.5 - \varepsilon + \frac{\varepsilon}{10})(c_2 + c_{\geq 4}) + \frac{\varepsilon}{10}(|E(\alpha)| - |P'|)$$

Using $|P'| \leq c_2 + c_{\geq 4}$ completes the proof. ∎

Adding Lemma 5.45 and Lemma 5.47, we get the result. ∎

Corollary 5.48. *If β is a nontrivial non-spider and $\beta \in W(\alpha)$ for some spider $\alpha \in \mathcal{L}$, then*

$$\eta^{|U_\alpha| + |V_\alpha|} \cdot \frac{1}{n^{|E(\alpha)|/2}} \left\|M_\beta\right\| \leq \eta^{|U_\beta| + |V_\beta|} \cdot \frac{1}{n^{\Omega(\varepsilon |E(\alpha)|)}}$$

Proof. From Lemma A.3, we have

$$\left\|M_\beta\right\| \leq 2 \cdot (|V(\beta)| \cdot (1 + |E(\beta)|) \cdot \log(n))^{C \cdot (|V_{rel}(\beta)| + |E(\beta)|)} \cdot n^{\frac{w(V(\beta)) - w(S_{\min}) + w(W_{iso})}{2}}$$

We have $|V(\beta)| \cdot (1 + |E(\beta)|) \cdot \log(n) \leq n^{O(\tau)}$. Also, $|V_{rel}(\beta)| \leq 2(|E(\alpha)| + |E(\beta)|)$ since all the degree 0 vertices in $V_{rel}(\beta)$ would have had vertices of $V_{rel}(\alpha)$ collapse into it in the chain of collapses and there are no degree 0 vertices in $V_{rel}(\alpha)$. Finally, since $|E(\alpha)| \geq |E(\beta)|$, the factor $2 \cdot (|V(\beta)| \cdot (1 + |E(\beta)|) \cdot \log(n))^{C \cdot (|V_{rel}(\beta)| + |E(\beta)|)}$ can be absorbed into $\frac{1}{n^{\Omega(\varepsilon |E(\alpha)|)}}$. The result follows from Lemma 5.44. ∎

Proposition 5.49. *If β is a trivial shape, $\lambda_\beta^+ = \lambda_\beta$.*

Proof. A trivial shape cannot appear in $W(\alpha)$ for any α, since every collapse of a spider always keeps its circle vertices around. ∎

Lemma 5.50. *For $k, l \in \{0, 1, \dots, D/2\}$, let $\mathcal{B}_{k,l}$ denote the set of nontrivial non-spiders on block (k, l). Then*

$$\sum_{\beta \in \mathcal{B}_{k,l}} \left|\lambda_\beta^+\right| \left\|M_\beta\right\| \le \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$$

Proof.

$$\sum_{\beta \in \mathcal{B}_{k,l}} \left\|\lambda_\beta^+ M_\beta\right\| \le \sum_{\beta \in \mathcal{B}_{k,l}} |\lambda_\beta| \left\|M_\beta\right\| + \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} |v_\beta| \, |\lambda_\alpha| \left\|M_\beta\right\|$$

To bound the first term, we checked previously in Corollary 5.14 that the total norm of nontrivial non-spiders appearing in the pseudocalibration (i.e. this term) is $\eta^{k+l} o_n(1)$. For the second term, via Lemma 5.41 we have a bound on the accumulations v_γ of one spider on one non-spider, so it is at most

$$\le \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} (C_1 \left|V(\alpha)\right| \cdot \left|E(\alpha)\right|)^{C_2 |E(\alpha)|} \cdot |\lambda_\alpha| \left\|M_\beta\right\|.$$

Use the bound on the coefficients $|\lambda_\alpha|$, Proposition 5.13,

$$\le \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} (C_1 \left|V(\alpha)\right| \cdot \left|E(\alpha)\right|)^{C_2 |E(\alpha)|} \cdot \eta^{|U_\alpha| + |V_\alpha|} \cdot \frac{\left|E(\alpha)\right|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}} \cdot \left\|M_\beta\right\|$$

Invoking the norm bound for non-spiders which are collapses, Corollary 5.48,

$$\le \eta^{k+l} \cdot \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left(\frac{C_1 \left|V(\alpha)\right| \cdot \left|E(\alpha)\right|}{n^{\Omega(\varepsilon)}}\right)^{C_2' |E(\alpha)|}$$

$$\le \eta^{k+l} \cdot \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{\substack{\text{spiders } \alpha: \\ \beta \in W(\alpha)}} \left(\frac{C_1 n^\tau \cdot n^\tau}{n^{\Omega(\varepsilon)}}\right)^{C_2' |E(\alpha)|}.$$

Bound the sum over all spiders by the sum over all shapes. By Proposition 5.12, the number of shapes with i edges is $n^{O(\tau(i+1))}$. Summing by the number of edges, observe that $|E(\alpha)| \ge \max(|E(\beta)|, 2)$ since spiders always have at least 2 edges.

$$\le \eta^{k+l} \sum_{\beta \in \mathcal{B}_{k,l}} \sum_{i=\max(|E(\beta)|,2)}^{\infty} n^{O(\tau(i+1))} \cdot \left(\frac{C_1 n^\tau \cdot n^\tau}{n^{\Omega(\varepsilon)}}\right)^{C_2' i}$$

$$\le \eta^{k+l} \sum_{\beta \in \mathcal{B}_{k,l}} \frac{1}{n^{\Omega(\varepsilon \max(|E(\beta)|,2))}}$$

$$\le \eta^{k+l} \sum_{i=0}^{\infty} \frac{n^{O(\delta(i+1))}}{n^{\Omega(\varepsilon \max(i,2))}}$$

$$= \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}} \qquad \blacksquare$$

Corollary 5.51. *For $k \in \{0, \dots, D/2\}$, the (k, k) block of $\mathcal{M}^+$ has minimum singular value at least $\eta^{2k}(1 - \frac{1}{n^{\Omega(\varepsilon)}})$, and for $k, l \in \{0, \dots, D/2\}, l \ne k$, the (k, l) off-diagonal block has norm at most $\eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$.*

Proof. By Proposition 5.49 the identity matrix appears on the (k,k) blocks with coefficient η^{2k}. By construction, $\mathcal{M}^+$ has no spider shapes. By Lemma 5.50, the total norm of the non-spider shapes on the (k,l) block is at most $\eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$. ∎

Theorem 5.52. *W.h.p.* $\mathcal{M}_{fix} \succeq 0$.

Proof. For any $x \in \text{Null}(\mathcal{M}_{fix})$, we of course have $x^\intercal \mathcal{M}_{fix} x = 0$. For any $x \perp \text{Null}(\mathcal{M}_{fix})$ with $\|x\|_2 = 1$,

$$\begin{aligned} x^\intercal \mathcal{M}_{fix} x &= x^\intercal (\mathcal{M} + \mathcal{E}) x \\ &= x^\intercal \mathcal{M}^+ x + x^\intercal \left(\sum_{\text{spiders } \alpha} \lambda_\alpha \left(\mathcal{M}_\alpha - \sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma \right) \right) x \\ &\quad + x^\intercal \mathcal{E} x \\ &= x^\intercal (\mathcal{M}^+ + \mathcal{E}) x \qquad \text{(Proposition 5.37)} \end{aligned}$$

Because the norm bound on $\mathcal{E}$ in Corollary 7.3 is significantly less than $\eta^D = n^{-n^\delta}$, the bound on the norm of each block of $\mathcal{M}^+$ in Corollary 5.51 also applies to the blocks of $\mathcal{M}^+ + \mathcal{E}$. Therefore, we use Lemma 5.2 to conclude $\mathcal{M}^+ + \mathcal{E} \succeq 0$ and the above expression is nonnegative. ∎

6 Sherrington-Kirkpatrick Lower Bounds

Here, we prove Theorem 1.5 and Theorem 1.2.

Recall that in the Planted Boolean Vector problem, we wish to optimize

$$\mathsf{OPT}(V) := \frac{1}{n} \max_{b \in \{\pm 1\}^n} b^\intercal \Pi_V b,$$

where V is a uniformly random p-dimensional subspace of $\mathbb{R}^n$.

Theorem 1.5. *[Main III] There exists a constant $c > 0$ such that, for all $\varepsilon > 0$ and $\delta \le c\varepsilon$, for $p \ge n^{2/3+\varepsilon}$, w.h.p. over V there is a degree-n^δ SoS solution for Planted Boolean Vector of value 1.*

Proof. We wish to produce an SoS solution $\widetilde{\mathbb{E}}$ on boolean variables $b_1, \dots, b_n$ such that $\widetilde{\mathbb{E}}[b^\intercal \Pi_V b] = n$. Instead of sampling a uniformly random p-dimensional subspace V of $\mathbb{R}^n$, we first sample $d_1, \dots, d_n$ i.i.d. p-dimensional Gaussian vectors from $\mathcal{N}(0, I)$, then form an n-by-p matrix A with rows $d_1, \dots, d_n$, and finally take V to be the span of the columns of A. Since the columns of A are isotropic i.i.d. random Gaussian vectors, we have that V is a uniform p-dimension subspace[7] of $\mathbb{R}^n$.

We will consider V as the input for the Planted Boolean Vector problem while the vectors $d_1, \dots, d_n$ will be used to construct a pseudoexpectation operator for the Planted Affine Planes problem[8]. Since $n \le p^{3/2-\Omega(\varepsilon)}$, by Theorem 1.4, for all $\delta \le c\varepsilon$ for a constant $c > 0$, w.h.p., there exists a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}'$ on formal variables $v = (v_1, \dots, v_p)$ such that $\widetilde{\mathbb{E}}'[\langle v, d_u \rangle^2] = 1$ for every $u \in [n]$.

[7] Except for a zero measure event.

[8] Note that the vectors d_u are not "given" in the Planted Boolean Vector problem, though the construction of $\widetilde{\mathbb{E}}$ is not required to be algorithmic in any sense anyway.

Define $\widetilde{\mathbb{E}}$ by $\widetilde{\mathbb{E}}[b_u] := \widetilde{\mathbb{E}}'[\langle v, d_u\rangle]$ for all $u \in [n]$ and extending it to all polynomials on $\{b_u\}$ by multilinearity. This is well defined because $\widetilde{\mathbb{E}}'[\langle v, d_u\rangle^2] = 1$. Note that $\widetilde{\mathbb{E}}$ is a valid pseudoexpectation operator of the same degree as $\widetilde{\mathbb{E}}'$. Finally, observe that

$$\frac{1}{n}\widetilde{\mathbb{E}}[b^\intercal \Pi_V b] = \frac{1}{n}\widetilde{\mathbb{E}}'[v^\intercal A^\intercal \Pi_V A v] = \frac{1}{n}\widetilde{\mathbb{E}}'[v^\intercal A^\intercal A v] = 1.$$

■

Now we prove lower bounds for the Sherrington-Kirkpatrick problem, using a reduction and proof due to [MRX19]. We include it here for completeness. Recall that the SK problem is to compute

$$\mathsf{OPT}(W) := \max_{x\in\{\pm 1\}^n} x^\intercal W x,$$

where W is sampled from $\mathrm{GOE}(n)$.

Theorem 1.2. *[Main I] There exists a constant $\delta > 0$ such that, w.h.p. for $W \sim \mathrm{GOE}(n)$, there is a degree-n^δ SoS solution for the Sherrington–Kirkpatrick problem with value at least $(2 - \mathrm{o}_n(1)) \cdot n^{3/2}$.*

We will use the following standard results from random matrix theory of $\mathrm{GOE}(n)$.

Fact 6.1. *Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of $W \sim \mathrm{GOE}(n)$ with corresponding normalized eigenvectors $w_1, \ldots, w_n$. Then,*

1. *For every $p \in [n]$, the span of $w_1, \ldots, w_p$ is a uniformly random p-dimensional subspace of $\mathbb{R}^n$ (see e.g. [OVW16, Section 2]).*
2. *W.h.p., $\lambda_{n^{0.67}} \geq (2 - \mathrm{o}(1))\sqrt{n}$ (Corollary of Wigner's semicircle law [Wig93])*

Proof of Theorem 1.2: Let $p = n^{0.67}$ and $W \sim \mathrm{GOE}(n)$. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of W with corresponding orthonormal set of eigenvectors $w_1, \ldots, w_n$. By Fact 6.1, we have that $\lambda_p \geq (2 - \mathrm{o}(1))\sqrt{n}$ and that $w_1, \ldots, w_p$ span a uniformly random p-dimensional subspace V of $\mathbb{R}^n$.

We consider V as the input of the Boolean Planted Vector problem and by Theorem 1.5, for some constant $\delta > 0$, w.h.p. there exists a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}$ such that $\widetilde{\mathbb{E}}[x_i^2] = 1$ and $\widetilde{\mathbb{E}}[\sum_{i=1}^p \langle x, w_i\rangle^2] = \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] = n$. Now,

$$\begin{aligned}\widetilde{\mathbb{E}}[x^\intercal W x] = \widetilde{\mathbb{E}}[\sum_{i=1}^n \lambda_i \langle x, w_i\rangle^2] &\geq \lambda_p \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] - |\lambda_n|\,\widetilde{\mathbb{E}}[\sum_{i=p+1}^n \langle x, w_i\rangle^2] \\ &\geq (2 - \mathrm{o}(1))n^{3/2} - |\lambda_n|\,\widetilde{\mathbb{E}}[\langle x, x\rangle - \sum_{i=1}^p \langle x, w_i\rangle^2] = (2 - \mathrm{o}(1))n^{3/2}.\end{aligned}$$

■

Remark 6.2. *Using the same proof as above, we can obtain Theorem 1.2 even if we were only able to prove SoS lower bounds for Planted Affine Planes for some $m = \omega(n)$. So, pushing the value of m up to $n^{3/2-\varepsilon}$, which is Theorem 1.4, offers only a modest improvement.*

7 Satisfying the Constraints Exactly

After pseudocalibration, the PAP constraints "$\langle v, d_u\rangle^2 = 1$" are not exactly satisfied by the pseudocalibration, but they are satisfied up to truncation error $\widetilde{\mathbb{E}}[\langle v, d_u\rangle^2 - 1] = n^{-\Omega(n^\tau)}$. This is enough to produce a Sherrington-Kirkpatrick solution that is *almost* boolean, meaning $\widetilde{\mathbb{E}}[x_i^2] = 1 \pm n^{-\Omega(n^\tau)}$ where the pseudocalibration is truncated to degree n^τ. To satisfy the constraints exactly, and produce an SK solution which is *exactly* boolean, we can project the pseudocalibration operator. The goal of this section is to prove the following lemma for the PAP problem,

Lemma 7.1. *W.h.p. for the PAP problem there is $\widetilde{\mathbb{E}}' \in \mathbb{R}^{\binom{[n]}{\leq D}}$ such that $\left\|\widetilde{\mathbb{E}} - \widetilde{\mathbb{E}}'\right\|_2 \leq \frac{1}{n^{\Omega(n^\tau)}}$ and $\widetilde{\mathbb{E}}'$ exactly satisfies the constraints "$\langle v, d_u\rangle^2 = 1$".*

Remark 7.2. *Note that $\widetilde{\mathbb{E}}'$ is syntactically guaranteed to still satisfy the constraints "$v_i^2 = \frac{1}{n}$".*

Corollary 7.3. *There is an $\binom{[n]}{\leq D/2} \times \binom{[n]}{\leq D/2}$ matrix $\mathcal{E}$ with $\|\mathcal{E}\| \leq \frac{1}{n^{\Omega(n^\tau)}}$ such that the matrix $M_{fix} := M + \mathcal{E}$ is SoS-symmetric and exactly satisfies the constraints "$\langle v, d_u\rangle^2 = 1$".*

We view the operators $\widetilde{\mathbb{E}}$ as vectors in $\mathbb{R}^{\binom{[n]}{\leq D}}$. The approach we take is to define a "check matrix" Q such that $\widetilde{\mathbb{E}}$ satisfies the necessary constraints iff $\widetilde{\mathbb{E}} \in \text{Null}(Q)$. When the constraints are functions of v only, the matrix Q would be filled with constants. Since the constraints depend on the inputs d_u, the matrix Q is also a function of the d_u. This allows us to deconstruct it as a sum of graph matrices – and in fact it is made out of graph matrices which we have seen already.

Definition 7.4. *We let Q be the matrix*

$$Q := \sum_{k=2}^{D} L_k^\intercal$$

where the matrices L_k are defined in Section 5.

Lemma 7.5. *$Q\widetilde{\mathbb{E}} = 0$ iff $\widetilde{\mathbb{E}}$ exactly satisfies the constraints "$\langle v, d_u\rangle^2 = 1$".*

Proof. One can see in the proof of Lemma 5.21 that the entries of $Q\widetilde{\mathbb{E}}$ measure exactly the error in the constraints. ∎

The natural choice of $\widetilde{\mathbb{E}}'$ is therefore the projection of $\widetilde{\mathbb{E}}$ to the nullspace. This is defined by

$$\widetilde{\mathbb{E}}' := \widetilde{\mathbb{E}} - Q^\intercal(QQ^\intercal)^+ Q\widetilde{\mathbb{E}}$$

where we take the pseudo-inverse of $QQ^\intercal$ as it will turn out not to be invertible.

To prove Lemma 7.1, we must decompose the second term in terms of graph matrices and show it has small norm.

As a warm-up, we end this outline by showing a simpler projection argument in the Planted Boolean Vector domain is sufficient if one just wants to satisfy the boolean constraints in the Planted Boolean Vector problem rather than the constraints of the PAP problem.[9]

Let $\widetilde{\mathbb{E}}_{\text{PBV}}$ be a candidate, not-yet-boolean, degree-D pseudoexpectation operator for the Planted Boolean Vector problem, $D = 2 \cdot n^\delta$. $\widetilde{\mathbb{E}}_{\text{PBV}}$ has an entry for each monomial b^α, therefore it is

[9] Using the translation between the two problems in Section 6, this would allow us to exactly satisfy "$\langle v, d_u\rangle^2 = 1$" for the PAP problem. Unfortunately, the constraints "$v_i^2 = \frac{1}{n}$" might be broken.

$\left(\!\!\binom{n}{\leq D}\!\!\right)$-dimensional. Let Q_{bool} be the "check matrix" for the boolean constraints. Q_{bool} has $n \cdot \left(\!\!\binom{n}{\leq D-2}\!\!\right)$ rows. The (i,α) row checks $\widetilde{\mathbb{E}}[b^\alpha \cdot b_i^2] = \widetilde{\mathbb{E}}[b^\alpha]$. It has entry 1 in column α and entry -1 in column $\alpha \cup \{i,i\}$.

Lemma 7.6. *Assume that $\widetilde{\mathbb{E}}_{PBV}$ approximately satisfies the boolean constraints:*

$$\widetilde{\mathbb{E}}_{PBV}[b^\alpha \cdot (b_i^2 - 1)] \leq n^{-\Omega(n^\tau)}$$

for any b^α with degree at most $D-2$. Then letting $\widetilde{\mathbb{E}}'_{PBV}$ be the projection to $\mathrm{Null}(Q_{bool})$, we have

$$\left\|\widetilde{\mathbb{E}}_{PBV} - \widetilde{\mathbb{E}}'_{PBV}\right\|_2 \leq n^{-\Omega(n^\tau)}.$$

Proof. The effect of projecting $\widetilde{\mathbb{E}}$ to $\mathrm{Null}(Q_{bool})$ is to symmetrize $\widetilde{\mathbb{E}}[b^{\alpha+2\beta}]$ across all β; average all entries $\widetilde{\mathbb{E}}[1], \widetilde{\mathbb{E}}[b_1^2], \widetilde{\mathbb{E}}[b_2^2], \widetilde{\mathbb{E}}[b_1^6 b_7^4 b_{10}^2]$ etc, average $\widetilde{\mathbb{E}}[b_1], \widetilde{\mathbb{E}}[b_1 b_3^2], \widetilde{\mathbb{E}}[b_1 b_3^4 b_4^4]$ etc, and so on. One can see this because this is a linear map which fixes $\mathrm{Null}(Q_{bool})$ and takes all vectors into $\mathrm{Null}(Q_{bool})$.

By assumption, there is additive error $n^{-\Omega(n^\tau)}$ between $\widetilde{\mathbb{E}}_{\mathrm{PBV}}[b^\alpha]$ and $\widetilde{\mathbb{E}}_{\mathrm{PBV}}[b^\alpha \cdot b_i^2]$. As the size of β is at most $D \ll n^\tau$, we still easily have $\widetilde{\mathbb{E}}_{\mathrm{PBV}}[b^{\alpha+2\beta}] = \widetilde{\mathbb{E}}_{\mathrm{PBV}}[b^\alpha] \pm n^{-\Omega(n^\tau)}$ for all β. Therefore averaging these entries changes each of them by at most $n^{-\Omega(n^\tau)}$. Thus,

$$\left\|\widetilde{\mathbb{E}}_{\mathrm{PBV}} - \widetilde{\mathbb{E}}'_{\mathrm{PBV}}\right\|_2 \leq \left(\!\!\binom{n}{\leq D}\!\!\right) \cdot \left\|\widetilde{\mathbb{E}}_{\mathrm{PBV}} - \widetilde{\mathbb{E}}'_{\mathrm{PBV}}\right\|_\infty$$

$$\leq n^{O(n^\delta)} \cdot n^{-\Omega(n^\tau)} = n^{-\Omega(n^\tau)}$$

■

7.1 Truncation error in the pseudocalibration

The constraint "$\langle v, d_u\rangle^2 = 1$" isn't exactly satisfied, but a general property of pseudocalibration is that it's satisfied up to truncation error, which is small w.h.p. We show a quantitative version of this bound.

We introduce the notation

$$\mu_{I,\alpha} := \mathop{\mathbb{E}}_{\mathrm{pl}}[v^I \chi_\alpha(d)]$$

where $\chi_\alpha(d) = h_\alpha(d)$ in the Gaussian case and $\chi_\alpha(d) = d^\alpha$ in the boolean case.

Lemma 7.7. *Let $p(d,v)$ such that p is uniformly zero on the planted distribution. Let $\deg_d(p) = D$. For any $I \subseteq [n]$, the only nonzero Fourier coefficients of $\widetilde{\mathbb{E}}[v^I p]$ are those with size between $n^\tau \pm D$.*

Furthermore, the nonzero coefficients are bounded in absolute value by

$$M \cdot L \cdot 2^D e^{mn} \cdot \max_I \max_{|\alpha| \in n^\tau \pm 2D} |\mu_{I,\alpha}|$$

where M is the number of nonzero monomials of p and L is the largest coefficient of p (in absolute value).

Proof. We divide the calculations into boolean and Gaussian cases. For each case we compute that Fourier coefficients below the truncation threshold neatly cancel and bound the coefficients at the threshold.

(Boolean case) Expand $p(d,v) = \sum_{|J|\le D} d^J p_J(v)$. By linearity,

$$\widetilde{\mathbb{E}}[v^I p] = \sum_{|J|\le D} d^J \widetilde{\mathbb{E}}[v^I p_J(v)].$$

The α-th Fourier coefficient gets a contribution from the J-th term equal to the $(\alpha \oplus J)$-th Fourier coefficient of $\widetilde{\mathbb{E}}[v^I p_J(v)]$. Expand the polynomial p_J in the J-th term,

$$\widetilde{\mathbb{E}}[v^I p_J(v)] = \sum_K c_{J,K} \widetilde{\mathbb{E}}[v^I v^K]$$

The $(\alpha \oplus J)$-th coefficient of $\widetilde{\mathbb{E}}[v^I v^K]$ is defined by pseudocalibration to be

$$\begin{cases} \mu_{I+K,\alpha\oplus J} & |\alpha \oplus J| \le n^\tau \\ 0 & |\alpha \oplus J| > n^\tau \end{cases} \tag{2}$$

For $|\alpha| \le n^\tau - D$ we are guaranteed to be in the first case. For this case the total α-th Fourier coefficient is

$$\begin{aligned} \sum_{|J|\le D} \sum_K c_{J,K} \mu_{I+K,\alpha\oplus J} &= \sum_{|J|\le D} \sum_K c_{J,K} \mathop{\mathbb{E}}_{\text{pl}}[v^I v^K d^{J\oplus\alpha}] \\ &= \sum_{|J|\le D} \sum_K c_{J,K} \mathop{\mathbb{E}}_{\text{pl}}[v^I v^K d^\alpha d^J] \\ &= \mathop{\mathbb{E}}_{\text{pl}}[v^I d^\alpha p(d,v)] \\ &= 0. \end{aligned}$$

For $|\alpha| > n^\tau + D$, we are guaranteed to be in the second case of Eq. (2), in which case the total Fourier coefficient will also be zero. For $|\alpha|$ within D of the truncation parameter, some terms J will not contribute their coefficients towards cancellation. We bound the Fourier coefficient for these α,

$$\begin{aligned} \left| \sum_{\substack{J:|J|\le D,\\ |\alpha\oplus J|\le n^\tau}} \sum_K c_{J,K} \cdot \mu_{I+K,\alpha\oplus J} \right| &\le \sum_{|J|\le D} \sum_K |c_{J,K} \cdot \mu_{I+K,\alpha\oplus J}| \\ &\le M \cdot L \cdot \max_I \max_{|\alpha|\in n^\tau \pm 2D} |\mu_{I,\alpha}|. \end{aligned}$$

(Gaussian case) Expand $p(d,v) = \sum_{|\beta|\le D} h_\beta(d) p_\beta(v) = \sum_{|\beta|\le D} h_\beta(d) \sum_K c_{\beta,K} v^K$. The pseudoexpectation is

$$\begin{aligned} \widetilde{\mathbb{E}}[v^I p(d,v)] &= \sum_{|\beta|\le D} h_\beta(d) \widetilde{\mathbb{E}}[v^I p_\beta(v)] \\ &= \sum_{|\beta|\le D} h_\beta(d) \sum_K c_{\beta,K} \widetilde{\mathbb{E}}[v^I v^K] \\ &= \sum_{|\beta|\le D} h_\beta(d) \sum_K c_{\beta,K} \sum_{|\alpha|\le n^\tau} \mu_{I+K,\alpha} \frac{h_\alpha(d)}{\alpha!}. \end{aligned}$$

Let $l_{\alpha,\beta,\gamma}$ be the coefficient of h_γ in the Hermite product $h_\alpha \cdot h_\beta$.

$$\widetilde{\mathbb{E}}[v^I p(d,v)] = \sum_{|\beta|\le D} \sum_K c_{\beta,K} \sum_{|\alpha|\le n^\tau} \mu_{I+K,\alpha} \sum_\gamma l_{\alpha,\beta,\gamma} \frac{h_\gamma(d)}{\alpha!}$$

In the case $|\gamma| > n^\tau + D$, the coefficient of $h_\gamma(d)$ is zero because the max degree of a Hermite polynomial appearing in $h_\alpha \cdot h_\beta$ is at most $|\alpha| + |\beta| \le n^\tau + D$. We show cancellations occur when $|\gamma| \le n^\tau - D$. Moving the summations around, the coefficient of h_γ is,

$$\begin{aligned}
&\sum_{|\beta|\le D} \sum_K c_{\beta,K} \sum_{|\alpha|\le n^\tau} \mu_{I+K,\alpha} \cdot l_{\alpha,\beta,\gamma} \frac{1}{\alpha!} \\
&= \sum_{|\beta|\le D} \sum_K c_{\beta,K} \sum_{|\alpha|\le n^\tau} \mathop{\mathbb{E}}_{\text{pl}}[v^I v^K h_\alpha(d)] \cdot l_{\alpha,\beta,\gamma} \frac{1}{\alpha!} \\
&= \mathop{\mathbb{E}}_{\text{pl}} v^I \sum_{|\beta|\le D} \sum_K c_{\beta,K} v^K \sum_{|\alpha|\le n^\tau} l_{\alpha,\beta,\gamma} \frac{h_\alpha(d)}{\alpha!}.
\end{aligned}$$

We need an explicit formula for $l_{\alpha,\beta,\gamma}$ from [Rom05, p. 92],

Proposition 7.8.

$$l_{\alpha,\beta,\alpha+\beta-2\delta} = \prod_{u,i} \binom{\alpha_{ui}}{\delta_{ui}} \binom{\beta_{ui}}{\delta_{ui}} \delta_{ui}!$$

Proposition 7.9.

$$\sum_\alpha l_{\alpha,\beta,\gamma} \frac{h_\alpha(d)}{\alpha!} = h_\beta(d) \cdot \frac{h_\gamma(d)}{\gamma!}$$

Proof. Compute using Proposition 7.8. ∎

In Proposition 7.9, the summation is actually finite. The largest α with $l_{\alpha,\beta,\gamma}$ nonzero has $|\alpha| \le |\beta| + |\gamma|$. Since we have $|\beta| \le D$ (the constraint only has degree D), as long as $|\gamma| \le n^\tau - D$, the above equality applies, in which case continuing the calculation for this case,

$$\begin{aligned}
\widetilde{\mathbb{E}}[v^I p] &= \mathop{\mathbb{E}}_{\text{pl}} v^I \sum_{|\beta|\le D} \sum_K c_{\beta,K} v^K \cdot h_\beta(d) \cdot \frac{h_\gamma(d)}{\gamma!} \\
&= \mathop{\mathbb{E}}_{\text{pl}} v^I \cdot \frac{h_\gamma(d)}{\gamma!} \cdot \sum_{|\beta|\le D} \sum_K c_{\beta,K} v^K \cdot h_\beta(d) \\
&= \mathop{\mathbb{E}}_{\text{pl}} v^I \cdot \frac{h_\gamma(d)}{\gamma!} \cdot p(d,v) \\
&= 0.
\end{aligned}$$

We now bound the coefficients that appear in the remaining case when $n^\tau - D < |\gamma| \le n^\tau + D$.

$$\left| \sum_{|\beta|\le D} \sum_K c_{\beta,K} \sum_{|\alpha|\le n^\tau} \mu_{I+K,\alpha} \cdot l_{\alpha,\beta,\gamma} \frac{1}{\alpha!} \right| \le \sum_{|\beta|\le D} \sum_K |c_{\beta,K}| \sum_{|\alpha|\le n^\tau} |\mu_{I+K,\alpha}| \cdot l_{\alpha,\beta,\gamma} \frac{1}{\alpha!}$$

If $l_{\alpha,\beta,\gamma} > 0$ then we must have $|\alpha| \ge |\gamma| - |\beta| \ge n^\tau - 2D$.

$$\le \sum_{|\beta|\le D} \sum_K |c_{\beta,K}| \cdot \left(\max_I \max_{|\alpha| \in n^\tau \pm 2D} |\mu_{I,\alpha}| \right) \sum_\alpha l_{\alpha,\beta,\gamma} \frac{1}{\alpha!}$$

Proposition 7.10.

$$\sum_{\alpha} l_{\alpha,\beta,\gamma} \frac{1}{\alpha!} = e^{mn} \prod_{u,i} \binom{\beta_{ui}}{\frac{\alpha_{ui}+\beta_{ui}-\gamma_{ui}}{2}}$$

Proof. Compute using Proposition 7.8. ∎

Using the proposition,

$$\leq \sum_{|\beta|\leq D} \sum_{K} \left|c_{\beta,K}\right| \cdot \left(\max_{I} \max_{|\alpha|\in n^{\tau}\pm 2D} \left|\mu_{I,\alpha}\right|\right) e^{mn} \prod_{u,i} \binom{\beta_{ui}}{\frac{\alpha_{ui}+\beta_{ui}-\gamma_{ui}}{2}}$$

We can bound

$$\prod_{u,i} \binom{\beta_{ui}}{k_{ui}} \leq \prod_{u,i} 2^{\beta_{ui}} = 2^{|\beta|} \leq 2^{D}.$$

In total, letting M be the number of nonzero coefficients in the constraint p and L be the largest coefficient, this Fourier coefficient is at most,

$$M \cdot L \cdot 2^{D} e^{mn} \cdot \max_{I} \max_{|\alpha|\in n^{\tau}\pm 2D} \left|\mu_{I+K,\alpha}\right|.$$

∎

Lemma 7.11. *W.h.p.* $\left\|Q\widetilde{\mathbb{E}}\right\| \leq \frac{1}{n^{\Omega(n^{\tau})}}$

Proof. Via Lemma 7.7 the only nonzero Fourier characters that appear in $Q\widetilde{\mathbb{E}}$ are those of size $n^{\tau} \pm 2$. Their coefficient in the lemma is at most

$$\begin{aligned}
& C \cdot e^{mn} \cdot \max_{I} \max_{|\alpha|\in n^{\tau}\pm 4} \left|\mu_{I+K,\alpha}\right| \\
\leq & C \cdot e^{mn} \cdot \frac{(n^{\tau}-4)^{3(n^{\tau}-4)}}{n^{(n^{\tau}-4)/2}} && \text{(Proposition 5.13)} \\
\leq & \frac{n^{3\tau n^{\tau}}}{n^{(\frac{1}{2}+o(1))n^{\tau}}}
\end{aligned}$$

Therefore we can express $Q\widetilde{\mathbb{E}}$ as a sum of graph matrices[10] of this size, with coefficients bounded by the above quantity. Now we bound the total norm by summing over all graphs.

The number of graph matrices of this size is at most $n^{O(\tau)\cdot n^{\tau}}$.

The norm of each term can be bounded using norm bounds by $n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}}$ w.h.p. Note that there is no minimum vertex separator since $V = \emptyset$, and there are $O(1)$ isolated vertices when multiplying graphs in Q with graphs in the pseudocalibration (which have no isolated vertices). The number of circle vertices can be bounded by $\frac{1}{4}\left|E(\alpha)\right| \leq \frac{1}{4}n^{\tau}$. The number of square vertices can be bounded by $O(n^{\delta}) + \frac{1}{2}\left|E(\alpha)\right| \leq 0.52n^{\tau}$. Therefore the norms are at most $m^{\frac{1}{8}n^{\tau}}\ n^{0.26n^{\tau}+O(1)} \leq n^{0.49n^{\tau}}$. Notably this is significantly less than the denominator of the graph matrix coefficient, which is $n^{(0.5+o(1))n^{\tau}}$. Assuming δ and τ are small enough, the denominator is enough to overpower all terms multiplied together. ∎

[10] Graph vectors, since $Q\widetilde{\mathbb{E}}$ is a vector.

7.2 Analyzing QQ^T

The main theorem of this subsection is that the minimum nonzero eigenvalue of QQ^T is large.

Theorem 7.12. *The minimum nonzero eigenvalue of QQ^T is $\frac{n^2}{2} - \tilde{O}(n\sqrt{m})$.*

Proof of Lemma 7.1 assuming Theorem 7.12.

$$\begin{aligned}\left\|\widetilde{\mathbb{E}} - \widetilde{\mathbb{E}}'\right\| &= \left\|Q^\intercal (QQ^\intercal)^+ Q\widetilde{\mathbb{E}}\right\| \leq \|Q\| \cdot \left\|(QQ^\intercal)^+\right\| \cdot \left\|Q\widetilde{\mathbb{E}}\right\| \\ &\leq n^{O(1)} \cdot \frac{1}{n^2} \cdot \frac{1}{n^{\Omega(n^\tau)}} \\ &= \frac{1}{n^{\Omega(n^\tau)}}\end{aligned}$$

■

Recall that $Q = \sum_k L_k^T$. Let us refer to the five shapes in Definition 5.20 as α_1 through α_5, and their coefficients as c_{α_i}. Observe that the dominant part of L_k is $2M_{\alpha_1}$ which has norm $\tilde{O}(n)$. The norm bounds for the other components of L_k are as follows:

1. $\|c_{\alpha_2} M_{\alpha_2}\|$ is $\tilde{O}\left(\frac{1}{n} \cdot \sqrt{mn}\right) = \tilde{O}\left(\frac{\sqrt{m}}{\sqrt{n}}\right)$
2. $\|c_{\alpha_3} M_{\alpha_3}\|$ is $\tilde{O}\left(\frac{1}{n^2} \cdot n\sqrt{m}\right) = \tilde{O}\left(\frac{\sqrt{m}}{n}\right)$
3. $\|c_{\alpha_4} M_{\alpha_4}\|$ is $\tilde{O}\left(\frac{1}{n} \cdot \sqrt{mn}\right) = \tilde{O}\left(\frac{\sqrt{m}}{\sqrt{n}}\right)$
4. $\|c_{\alpha_5} M_{\alpha_5}\|$ is $\tilde{O}\left(\frac{1}{n} \cdot \sqrt{m}\right) = \tilde{O}\left(\frac{\sqrt{m}}{n}\right)$

We start by analyzing $QQ^T = \sum_k L_k^T L_k$.

From the above, taking $\alpha = \alpha_1$, the dominant term of L_k is $2M_\alpha$ where $U_\alpha = \{j_1, \ldots, j_k\}$, $V_\alpha = \{j_3, \ldots, j_k\} \cup \{u\}$, and $E(\alpha) = \{(j_1, u), (j_2, u)\}$. Since $\|M_\alpha\|$ is $\tilde{O}(n)$ and $\|L_k - 2M_\alpha\|$ is $\tilde{O}(\frac{\sqrt{m}}{\sqrt{n}})$, this implies that for each k, $\left\|L_k^T L_k - 4M_\alpha^T M_\alpha\right\|$ is $\tilde{O}(\sqrt{mn})$. Thus, it is sufficient to analyze $M_\alpha^T M_\alpha$.

Lemma 7.13. *Taking α to be the shape such that $U_\alpha \setminus V_\alpha = \{j_1, j_2\}$, $U_\alpha \cap V_\alpha = \{j_3, j_4, \ldots, j_k\}$, $V_\alpha \setminus U_\alpha = \{i_{circ}\}$, and $E(\alpha) = \{(j_1, i), (j_2, i)\}$,*

$$M_\alpha^T M_\alpha = M_{\alpha_1} + M_{\alpha_2} + M_{\alpha_3} + M_{\alpha_4} + M_{\alpha_5} + M_{\alpha_6}$$

where $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$ are the following shapes. Note that α_1 and α_2 are improper shapes.

1. *$U_{\alpha_1} \setminus V_{\alpha_1} = \emptyset$, $U_{\alpha_1} \cap V_{\alpha_1} = \{j_3, j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{\alpha_1} \setminus U_{\alpha_1} = \emptyset$, $V(\alpha_1) \setminus (U_{\alpha_1} \cup V_{\alpha_1}) = \{j_1, j_2\}$, and $E(\alpha_1) = \{(i, j_1), (i, j_1), (i, j_2), (i, j_2)\}$.*
2. *$U_{\alpha_2} \setminus V_{\alpha_2} = \{j_1\}$, $U_{\alpha_2} \cap V_{\alpha_2} = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{\alpha_2} \setminus U_{\alpha_2} = \{j_1'\}$, $V(\alpha_2) \setminus (U_{\alpha_2} \cup V_{\alpha_2}) = \{j_2\}$, and $E(\alpha_2) = \{(i, j_1), (i, j_1'), (i, j_2), (i, j_2)\}$.*
3. *$U_{\alpha_3} \setminus V_{\alpha_3} = \{j_1, j_2\}$, $U_{\alpha_3} \cap V_{\alpha_3} = \{j_5, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{\alpha_3} \setminus U_{\alpha_3} = \{j_1', j_2'\}$, and $E(\alpha_3) = \{(i, j_1), (i, j_2), (i, j_1'), (i, j_2')\}$.*
4. *$U_{\alpha_4} \setminus V_{\alpha_4} = \{i_{circ}\}$, $U_{\alpha_4} \cap V_{\alpha_4} = \{j_3, j_4, \ldots, j_k\}$, $V_{\alpha_4} \setminus U_{\alpha_4} = \{i_{circ}'\}$, $V(\alpha_4) \setminus (U_{\alpha_4} \cup V_{\alpha_4}) = \{j_1, j_2\}$, and $E(\alpha_4) = \{(i, j_1'), (i, j_2'), (i', j_1), (i', j_2)\}$.*

5. $U_{\alpha_5} \setminus V_{\alpha_5} = \{j_1\} \cup \{i_{circ}\}$, $U_{\alpha_5} \cap V_{\alpha_5} = \{j_4, \ldots, j_k\}$, $V_{\alpha_5} \setminus U_{\alpha_5} = \{j_1'\} \cup \{i_{circ}'\}$, $V(\alpha_5) \setminus (U_{\alpha_5} \cup V_{\alpha_5}) = \{j_2\}$, *and* $E(\alpha_5) = \{(i, j_1'), (i, j_2), (i', j_1), (i', j_2)\}$.
6. $U_{\alpha_6} \setminus V_{\alpha_6} = \{j_1, j_2\} \cup \{i_{circ}\}$, $U_{\alpha_6} \cap V_{\alpha_6} = \{j_5, \ldots, j_k\}$, $V_{\alpha_6} \setminus U_{\alpha_6} = \{j_1', j_2'\} \cup \{i_{circ}'\}$, *and* $E(\alpha_6) = \{(i', j_1), (i', j_2), (i, j_1'), (i, j_2')\}$.

For pictures of these shapes, see Fig. 9 below.

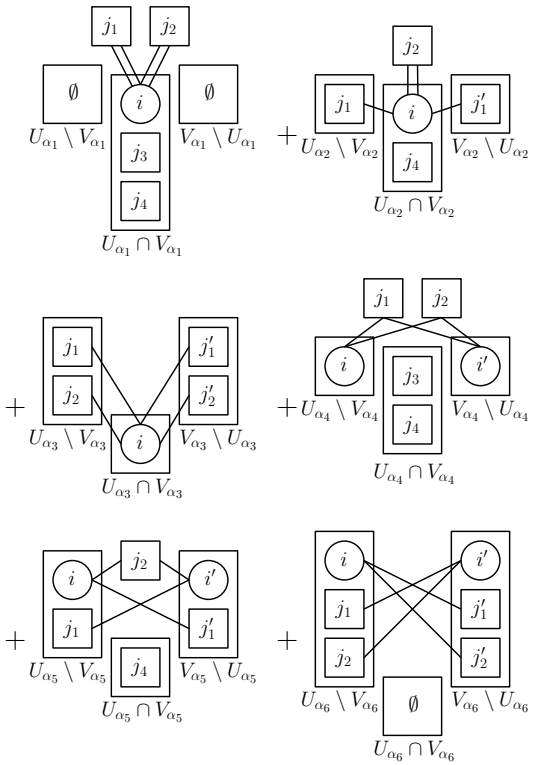

Figure 9: This figure shows the decomposition of $M_\alpha^T M_\alpha$.

Remark 7.14. *For $k = 2$, only shapes α_1 and α_4 are present and for $k = 3$, only shapes $\alpha_1, \alpha_2, \alpha_4, \alpha_5$ are present.*

Proof of Lemma 7.13. We compute $M_{\alpha^T} M_\alpha$ by considering the ribbons which appear in $M_{\alpha^T} M_\alpha$.

1. Each ribbon R with $A_R \setminus B_R = \emptyset$, $A_R \cap B_R = \{j_3, j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \emptyset$, $V(R) \setminus (A_R \cup B_R) = \{j_1, j_2\}$, and $E(R) = \{(i, j_1), (i, j_1), (i, j_2), (i, j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_3, j_4, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1, j_2\}$, and $E(R_1) = \{(i, j_1), (i, j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_3, j_4, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.
2. Each ribbon R with $A_R \setminus B_R = \{j_1\}$, $A_R \cap B_R = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \{j_1'\}$, $V(R) \setminus (A_R \cup B_R) = \{j_2\}$, and $E(R) = \{(i, j_1), (i, j_1'), (i, j_2), (i, j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_4, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1', j_2\}$, and $E(R_1) = \{(i, j_1'), (i, j_2)\}$.

(b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_4, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

3. Each ribbon R with $A_R \setminus B_R = \{j_1, j_2\}$, $A_R \cap B_R = \{j_5, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \{j_1', j_2'\}$, $V(R) \setminus (A_R \cup B_R) = \emptyset$, and $E(R) = \{(i, j_1), (i, j_1'), (i, j_2), (i, j_2')\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_2, j_5, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1', j_2'\}$, and $E(R_1) = \{(i, j_1'), (i, j_2')\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_2', j_5, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.
4. Each ribbon R with $A_R \setminus B_R = \{j_1, j_2\} \cup \{i_{circ}\}$, $A_R \cap B_R = \{j_5, \ldots, j_k\}$, $B_R \setminus A_R = \{j_1', j_2'\} \cup \{i_{circ}'\}$, $V(R) \setminus (A_R \cup B_R) = \emptyset$, and $E(R) = \{(i, j_1'), (i, j_2'), (i', j_1), (i', j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_2, j_5, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1', j_2'\}$, and $E(R_1) = \{(i, j_1'), (i, j_2')\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_2', j_5, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}'\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.
5. Each ribbon R with $A_R \setminus B_R = \{i_{circ}\}$, $A_R \cap B_R = \{j_3, j_4, \ldots, j_k\}$, $B_R \setminus A_R = \{i_{circ}'\}$, $V(R) \setminus (A_R \cup B_R) = \{j_1, j_2\}$, and $E(R) = \{(i, j_1), (i, j_2), (i', j_1), (i', j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_3, j_4, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1, j_2\}$, and $E(R_1) = \{(i, j_1), (i, j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_3, j_4, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}'\}$, and $E(R_2) = \{(i', j_1), (i', j_2)\}$.
6. Each ribbon R with $A_R \setminus B_R = \{j_1\}$, $A_R \cap B_R = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \{j_1'\}$, $V(R) \setminus (A_R \cup B_R) = \{j_2\}$, and $E(R) = \{(i, j_1), (i, j_1'), (i, j_2), (i, j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_4, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1', j_2\}$, and $E(R_1) = \{(i, j_1'), (i, j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_4, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

Based on these cases, we have that $M_{\alpha^T} M_\alpha = M_{\alpha_1} + M_{\alpha_2} + M_{\alpha_3} + M_{\alpha_4} + M_{\alpha_5} + M_{\alpha_6}$. ∎

We now analyze each of the matrices $M_{\alpha_1}, M_{\alpha_2}, M_{\alpha_3}, M_{\alpha_4}, M_{\alpha_5}, M_{\alpha_6}$.

Lemma 7.15. *Taking $Id_{k-2,1}$ to be the shape where $U_{Id_{k-2,1}} \setminus V_{Id_{k-2,1}} = \emptyset$, $U_{Id_{k-2,1}} \cap V_{Id_{k-2,1}} = \{j_3, j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{Id_{k-2,1}} \setminus U_{Id_{k-2,1}} = \emptyset$, and $E(Id_{k-2,1}) = \emptyset$, $\left\| M_{\alpha_1} - \binom{n-k+2}{2} M_{Id_{k-2,1}} \right\|$ is $\tilde{O}\left(n^{\frac{3}{2}}\right)$.*

Proof. To convert an improper shape α_1 to a sum of proper shapes, we take each ribbon of shape α_1 and decompose it into a sum of proper ribbons. Decomposing M_{α_1} in this way, each ribbon R with $A_R \setminus B_R = \emptyset$, $A_R \cap B_R = \{j_3, j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \emptyset$, and $E(R) = \emptyset$ appears $\binom{n-k+2}{2}$ times, once for each pair j_1, j_2 such that $j_1 < j_2$ and $j_1, j_2 \notin \{j_3, \ldots, j_k\}$. The other ribbons which arise all have an edge with label 2 incident with j_1 or j_2 and thus the resulting terms have norm $\tilde{O}\left(n^{\frac{3}{2}}\right)$. ∎

Definition 7.16. *Define β_2 to be the shape such that $U_{\beta_2} \setminus V_{\beta_2} = \{j_1\}$, $U_{\beta_2} \cap V_{\beta_2} = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{\beta_2} \setminus U_{\beta_2} = \emptyset$, and $E(\beta_1) = \{(i, j_1)\}$.*

Lemma 7.17. $\left\| M_{\alpha_2} - (n-k+1) M_{\beta_2} M_{\beta_2}^T \right\|$ *is* $\tilde{O}\left(n^{\frac{3}{2}}\right)$.

Proof. Again, to convert an improper shape α_2 to a sum of proper shapes, we take each ribbon of shape α_2 and decompose it into a sum of proper ribbons. Decomposing M_{α_2} in this way, each ribbon R with $A_R \setminus B_R = \{j_1\}$, $A_R \cap B_R = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \emptyset$, and $E(R) = \{(i, j_1), (i, j_1')\}$ appears $(n-k+1)$ times, once for each $j_2 \in [n] \setminus \{j_1, j_1', j_4, \ldots, j_k\}$. The other ribbons which arise have an edge with label 2 incident with j_2 and thus the resulting terms have norm $\tilde{O}\left(n^{\frac{3}{2}}\right)$. This implies that if we take α_2' to be the shape where $U_{\alpha_2} \setminus V_{\alpha_2} = \{j_1\}$, $U_{\alpha_2} \cap V_{\alpha_2} = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{\alpha_2} \setminus U_{\alpha_2} = \{j_1'\}$, and $E(\alpha_2) = \{(i, j_1), (i, j_1')\}$ then $\left\| M_{\alpha_2} - (n-k+1) M_{\alpha_2'} \right\|$ is $\tilde{O}\left(n^{\frac{3}{2}}\right)$.

$\left\| M_{\alpha_2'} \right\|$ is $\tilde{O}(n)$, so this term cannot be ignored. To handle this, we observe that $M_{\alpha_2'}$ is approximately equal to a PSD matrix. More precisely, $\left\| M_{\alpha_2'} - M_{\beta_2} M_{\beta_2}^T \right\|$ is $\tilde{O}(1)$. To see this, note that when we expand out $M_{\beta_2} M_{\beta_2}^T$, the ribbons which result when there are no collisions give $M_{\alpha_2'}$ and for each ribbon R which results from a collision, $A_R \setminus B_R = B_R \setminus A_R = \emptyset$ so the resulting terms have norm $\tilde{O}(1)$. ∎

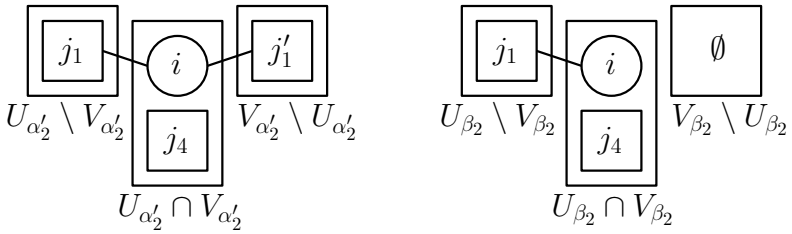

Figure 10: This figure shows α_2' and β_2 for $k = 4$.

$\|M_{\alpha_3}\|$ is $\tilde{O}(n^2)$, so this term cannot be ignored. To handle this, we observe that M_{α_3} is approximately equal to a PSD matrix. More precisely, we have the following lemma.

Definition 7.18. *Define β_3 to be the shape such that $U_{\beta_3} \setminus V_{\beta_3} = \{j_1, j_2\}$, $U_{\beta_3} \cap V_{\beta_3} = \{j_5, \ldots, j_k\} \cup \{i_{circ}\}$, $V_{\beta_3} \setminus U_{\beta_3} = \emptyset$, and $E(\beta_3) = \{(i, j_1), (i, j_2)\}$.*

Lemma 7.19. $\left\| M_{\alpha_3} - M_{\beta_3} M_{\beta_3}^T \right\|$ *is* $\tilde{O}(n)$.

Proof. To see this, note that when we expand out $M_{\beta_3} M_{\beta_3}^T$, the ribbons which result when there are no collisions give M_{α_3} and for each ribbon R which results from a collision, $|A_R \setminus B_R| = |B_R \setminus A_R| \leq 1$ so the resulting terms have norm $\tilde{O}(n)$. ∎

We now consider the norms of M_{α_4}, M_{α_5}, and M_{α_6}.

1. $\|M_{\alpha_4}\|$ is $\tilde{O}(n\sqrt{m})$.
2. $\|M_{\alpha_5}\|$ is $\tilde{O}(n\sqrt{m})$.

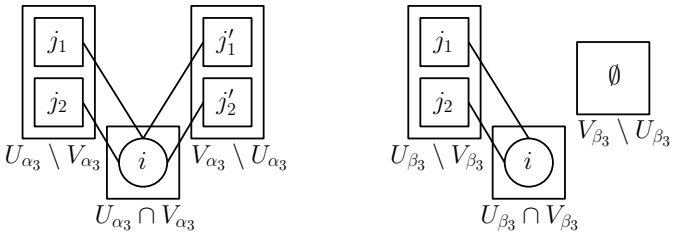

Figure 11: This figure shows α_3 and β_3 for $k = 4$.

3. $\|M_{\alpha_6}\|$ is $\tilde{O}(n^2)$.

This means that M_{α_4} and M_{α_5} can be ignored but M_{α_6} cannot be ignored. In fact, there is a very good reason for this. In particular, for $k \geq 4$, L_k has a non-trivial nullspace N_k, so we cannot show that the minimum nonzero eigenvalue of $L_k^T L_k$ is large without taking this nullspace into account. We handle this nullspace N_k in the next two subsubsections.

Putting everything together, we have the following corollary:

Corollary 7.20.

1. *For $k = 2$, $\left\| L_k^T L_k - 2n^2 M_{Id_{k-2,1}} \right\|$ is $\tilde{O}(n\sqrt{m})$*
2. *For $k = 3$, $\left\| L_k^T L_k - 2n^2 M_{Id_{k-2,1}} - 4n M_{\beta_2} M_{\beta_2}^T \right\|$ is $\tilde{O}(n\sqrt{m})$*
3. *For $k \geq 4$, $\left\| L_k^T L_k - 2n^2 M_{Id_{k-2,1}} - 4n M_{\beta_2} M_{\beta_2}^T - 4 M_{\beta_3} M_{\beta_3}^T - 4 M_{\alpha_6} \right\|$ is $\tilde{O}(n\sqrt{m})$*

Remark 7.21. *We replaced $\binom{n-k+2}{2}$ with $\frac{n^2}{2}$ as $\left\| M_{Id_{k-2,1}} \right\| = 1$ and $|\frac{n^2}{2} - \binom{n-k+2}{2}|$ is $\tilde{O}(n)$. Similarly, we replaced $(n-k+1)$ with n as $\left\| M_{\alpha'_2} \right\|$ is $\tilde{O}(n)$ and $|n - (n-k+1)|$ is $\tilde{O}(1)$*

7.2.1 The Null Space N_k

We now construct a matrix N_k for each $k \geq 4$ such that $L_k N_k = 0$ and the columns of N_k span the nullspace of L_k. To do this, we construct N_k so that the entries of each column of N_k is indexed by a subset $S = \{j_3, \ldots, j_k\} \subseteq [n]$ and an ordered tuple of circle indices (i, i') where $i < i'$. We then want that if we view $\tilde{E}$ as a vector,

$$(\tilde{E}^T L_k N_k)_{S,(i,i')} = \tilde{E}\left[v^S \left(\langle v, d_i\rangle^2 - 1\right)\left(\langle v, d_{i'}\rangle^2 - 1\right)\right] - \tilde{E}\left[v^S \left(\langle v, d_{i'}\rangle^2 - 1\right)\left(\langle v, d_i\rangle^2 - 1\right)\right] = 0$$

Lemma 7.22. *$N_k = c_{\alpha_1}(M_{\alpha_1^+} - M_{\alpha_1^-}) + c_{\alpha_2}(M_{\alpha_2^+} - M_{\alpha_2^-}) + c_{\alpha_3}(M_{\alpha_3^+} - M_{\alpha_3^-}) + c_{\alpha_4}(M_{\alpha_4^+} - M_{\alpha_4^-}) + c_{\alpha_5}(M_{\alpha_5^+} - M_{\alpha_5^-})$ for the following shapes $\alpha_1^+, \ldots, \alpha_5^+, \alpha_1^-, \ldots, \alpha_5^-$ and coefficients $c_{\alpha_1}, \ldots, c_{\alpha_5}$. Unless stated otherwise, all of these shapes have no middle vertices.*

1. *$U_{\alpha_1^+} \setminus V_{\alpha_1^+} = \{j_1, j_2\}$, $U_{\alpha_1^+} \cap V_{\alpha_1^+} = \{j_3, j_4, \ldots, j_k\} \cup \{i'_{circ}\}$, $V_{\alpha_1^+} \setminus U_{\alpha_1^+} = \{i_{circ}\}$, $E(\alpha_1^+) = \{(j_1, i), (j_2, i)\}$, and $c_{\alpha_1} = 2$.*
2. *$U_{\alpha_2^+} \setminus V_{\alpha_2^+} = \{j_2\}$, $U_{\alpha_2^+} \cap V_{\alpha_2^+} = \{j_4, \ldots, j_k\} \cup \{i'_{circ}\}$, $V_{\alpha_2^+} \setminus U_{\alpha_2^+} = \{j_3\} \cup \{i_{circ}\}$, $E(\alpha_2^+) = \{(j_3, i), (j_2, i)\}$, and $c_{\alpha_2} = \frac{2}{n}$.*

3. $U_{\alpha_3^+} \setminus V_{\alpha_3^+} = \emptyset$, $U_{\alpha_3^+} \cap V_{\alpha_3^+} = \{j_5, \ldots, j_k\} \cup \{i'_{circ}\}$, $V_{\alpha_3^+} \setminus U_{\alpha_3^+} = \{j_3, j_4\} \cup \{i_{circ}\}$, $E(\alpha_3^+) = \{(j_3, i), (j_4, i)\}$, *and* $c_{\alpha_3} = \frac{2}{n^2}$.
4. $U_{\alpha_4^+} \setminus V_{\alpha_4^+} = \emptyset$, $U_{\alpha_4^+} \cap V_{\alpha_4^+} = \{j_3, j_4, \ldots, j_k\} \cup \{i'_{circ}\}$, $V_{\alpha_4^+} \setminus U_{\alpha_4^+} = \{i_{circ}\}$, $V(\alpha_4^+) \setminus (U_{\alpha_4^+} \cup V_{\alpha_4^+}) = \{j_1\}$ $E(\alpha_4^+) = \{(j_1, i)_2\}$, *and* $c_{\alpha_4} = \frac{1}{n}$.
5. $U_{\alpha_5^+} \setminus V_{\alpha_5^+} = \emptyset$, $U_{\alpha_5^+} \cap V_{\alpha_5^+} = \{j_3, j_4, \ldots, j_k\} \cup \{i'_{circ}\}$, $V_{\alpha_5^+} \setminus U_{\alpha_5^+} = \{i_{circ}\}$, $E(\alpha_5^+) = \{(j_3, i)_2\}$, *and* $c_{\alpha_5} = \frac{1}{n}$.

where for all of these shapes, (i_{circ}, i'_{circ}) *is a tuple in the right side and* $i < i'$. $\alpha_1^-, \ldots, \alpha_5^-$ *are the same as* $\alpha_1^+, \ldots, \alpha_5^+$ *except that* i *and* i' *are swapped.*

Remark 7.23. *Note that* $\alpha_1^+, \ldots, \alpha_5^+$ *are the same shapes which appear in the decomposition of* L_k *except that the intersection of* U *and* V *now contains* i'_{circ} *and we require that* $i < i'$.

For pictures of these shapes, see Figure 12 below.

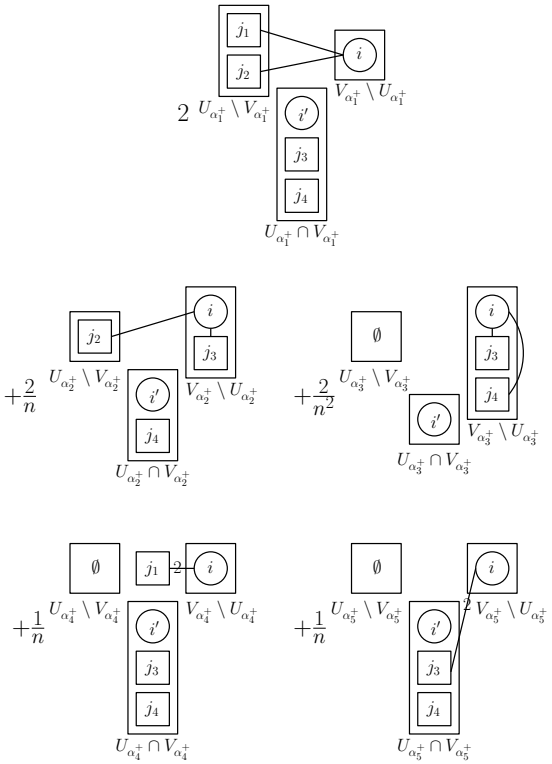

Figure 12: This figure shows the decomposition of N_k for $k = 4$. Here we always have that $i < i'$. If i and i' are swapped then this flips the signs but these parts are not shown to save space.

Proof. To determine N_k, we analyze the ribbons which N_k is composed of. Let $S = \{j_3, j_4, \ldots, j_k\}$.

1. If we take a ribbon R with $A_R = \{j_1, \ldots, j_k\} \cup \{i'_{circ}\}$, $B_R = \{j_3, \ldots, j_k\} \cup (i_{circ}, i'_{circ})$, and $E(R) = \{(j_1, i), (j_2, i)\}$ where $j_1 \neq j_2$ and $j_1, j_2 \notin S$ then

$$(\tilde{E}^T L_k M_R)_{S,i,i'} = \tilde{E}\left[v^S v_{j_1} v_{j_2} (d_i)_{j_1} (d_i)_{j_2} \left(\langle v, d_{i'} \rangle^2 - 1\right)\right]$$

Each such term appears with a coefficient of 2 in $\tilde{E}\left[v^S\left(\langle v,d_i\rangle^2-1\right)\left(\langle v,d_{i'}\rangle^2-1\right)\right]$, so we want each such ribbon R to appear with a coefficient of 2 in N_k.
Similarly, we want each ribbon R with $A_R=\{j_1,\ldots,j_k\}\cup\{i_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_1,i'),(j_2,i')\}$ where $j_1\neq j_2$ and $j_1,j_2\notin S$ to appear with a coefficient of -2 in N_k.

2. If we take a ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_3\})\cup\{i'_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_3,i),(j_2,i)\}$ where $j_1=j_3\in S$ and $j_2\notin S$ then
$$\begin{aligned}(\tilde{E}^TL_kM_R)_{S,i,i'}&=\tilde{E}\left[v^{S\setminus\{j_1,j_3\}}v_{j_2}(d_i)_{j_3}(d_i)_{j_2}\left(\langle v,d_{i'}\rangle^2-1\right)\right]\\&=n\tilde{E}\left[v^Sv_{j_1}v_{j_2}(d_i)_{j_1}(d_i)_{j_2}\left(\langle v,d_{i'}\rangle^2-1\right)\right]\end{aligned}$$
Each such term appears with a coefficient of $\frac{2}{n}$ in $\tilde{E}\left[v^S\left(\langle v,d_i\rangle^2-1\right)\left(\langle v,d_{i'}\rangle^2-1\right)\right]$, so we want each such ribbon R to appear with a coefficient of $\frac{2}{n}$ in N_k.
Similarly, we want each ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_3\})\cup\{i_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_3,i'),(j_2,i')\}$ where $j_1=j_3\in S$ and $j_2\notin S$ to appear with a coefficient of $-\frac{2}{n}$ in N_k.

3. If we take a ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_2,j_3,j_4\})\cup\{i'_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_3,i),(j_4,i)\}$ where $j_1=j_3\in S$ and $j_2=j_4\in S$ then
$$\begin{aligned}(\tilde{E}^TL_kM_R)_{S,i,i'}&=\tilde{E}\left[v^{S\setminus\{j_1,j_2,j_3,j_4\}}(d_i)_{j_3}(d_i)_{j_4}\left(\langle v,d_{i'}\rangle^2-1\right)\right]\\&=n^2\tilde{E}\left[v^Sv_{j_1}v_{j_2}(d_i)_{j_1}(d_i)_{j_2}\left(\langle v,d_{i'}\rangle^2-1\right)\right]\end{aligned}$$
Each such term appears with a coefficient of $\frac{2}{n^2}$ in $\tilde{E}\left[v^S\left(\langle v,d_i\rangle^2-1\right)\left(\langle v,d_{i'}\rangle^2-1\right)\right]$, so we want each such ribbon R to appear with a coefficient of $\frac{2}{n^2}$ in N_k.
Similarly, we want each ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_2,j_3,j_4\})\cup\{i_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_3,i'),(j_4,i')\}$ where $j_1=j_3\in S$ and $j_2=j_4\in S$ to appear with a coefficient of $-\frac{2}{n^2}$ in N_k.

4. If we take a ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_2\})\cup\{i'_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_1,i)_2\}$ where $j_1=j_2\notin S$ then
$$\begin{aligned}(\tilde{E}^TL_kM_R)_{S,i,i'}&=\tilde{E}\left[v^{S\setminus\{j_1,j_2\}}((d_i)_{j_1}^2-1)\left(\langle v,d_{i'}\rangle^2-1\right)\right]\\&=n\tilde{E}\left[v^Sv_{j_1}^2((d_i)_{j_1}^2-1)\left(\langle v,d_{i'}\rangle^2-1\right)\right]\end{aligned}$$
Each such term appears with a coefficient of $\frac{1}{n}$ in $\tilde{E}\left[v^S\left(\langle v,d_i\rangle^2-1\right)\left(\langle v,d_{i'}\rangle^2-1\right)\right]$ so we want each such ribbon R to appear with a coefficient of $\frac{1}{n}$ in N_k.
Similarly, we want each ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_2\})\cup\{i'_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_1,i)_2\}$ where $j_1=j_2\notin S$ to appear with a coefficient of $-\frac{1}{n}$ in N_k.

5. If we take a ribbon R with $A_R=(\{j_1,\ldots,j_k\}\setminus\{j_1,j_2\})\cup\{i'_{circ}\}$, $B_R=\{j_3,\ldots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_3,i)_2\}$ where $j_1=j_2=j_3\in S$ then
$$\begin{aligned}(\tilde{E}^TL_kM_R)_{S,i,i'}&=\tilde{E}\left[v^{S\setminus\{j_1,j_2\}}((d_i)_{j_3}^2-1)\left(\langle v,d_{i'}\rangle^2-1\right)\right]\\&=n\tilde{E}\left[v^Sv_{j_1}^2((d_i)_{j_1}^2-1)\left(\langle v,d_{i'}\rangle^2-1\right)\right]\end{aligned}$$

Each such term appears with a coefficient of $\frac{1}{n}$ in $\tilde{E}\left[v^S\left(\langle v,d_i\rangle^2-1\right)\left(\langle v,d_{i'}\rangle^2-1\right)\right]$ so we want each such ribbon R to appear with a coefficient of $\frac{1}{n}$ in N_k.
Similarly, we want each ribbon R with $A_R=(\{j_1,\dots,j_k\}\setminus\{j_1,j_2\})\cup\{i'_{circ}\}$, $B_R=\{j_3,\dots,j_k\}\cup(i_{circ},i'_{circ})$, and $E(R)=\{(j_3,i)_2\}$ where $j_1=j_2=j_3\in S$ to appear with a coefficient of $-\frac{1}{n}$ in N_k.

■

Observe that the dominant part of N_k is $2(M_{\alpha_1^+}-M_{\alpha_1^-})$ which has norm $\tilde{O}(n)$. The norm bounds for the other components of N_k are as follows:

1. $\left\|c_{\alpha_2}(M_{\alpha_2^+}-M_{\alpha_2^-})\right\|$ is $\tilde{O}\left(\frac{1}{n}\cdot\sqrt{mn}\right)=\tilde{O}\left(\frac{\sqrt{m}}{\sqrt{n}}\right)$
2. $\left\|c_{\alpha_3}(M_{\alpha_3^+}-M_{\alpha_3^-})\right\|$ is $\tilde{O}\left(\frac{1}{n^2}\cdot n\sqrt{m}\right)=\tilde{O}\left(\frac{\sqrt{m}}{n}\right)$
3. $\left\|c_{\alpha_4}(M_{\alpha_4^+}-M_{\alpha_4^-})\right\|$ is $\tilde{O}\left(\frac{1}{n}\cdot\sqrt{mn}\right)=\tilde{O}\left(\frac{\sqrt{m}}{\sqrt{n}}\right)$
4. $\left\|c_{\alpha_5}(M_{\alpha_5^+}-M_{\alpha_5^-})\right\|$ is $\tilde{O}\left(\frac{1}{n}\cdot\sqrt{m}\right)=\tilde{O}\left(\frac{\sqrt{m}}{n}\right)$

7.2.2 Analyzing $N_kN_k^T$

The dominant terms of N_k are $2M_{\alpha^+}-2M_{\alpha^-}$ where

1. $U_{\alpha^+}\setminus V_{\alpha^+}=\{j_1,j_2\}$, $U_{\alpha^+}\cap V_{\alpha^+}=\{j_3,\dots,j_k\}\cup\{i'_{circ}\}$, $V_{\alpha^+}\setminus U_{\alpha^+}=\{i_{circ}\}$, and $E(\alpha^+)=\{(j_1,i),(j_2,i)\}$. Note that here $i<i'$ and (i_{circ},i'_{circ}) appears as a tuple in V_{α^+}.
2. $U_{\alpha^-}\setminus V_{\alpha^-}=\{j_1,j_2\}$, $U_{\alpha^-}\cap V_{\alpha^-}=\{j_3,\dots,j_k\}\cup\{i_{circ}\}$, $V_{\alpha^-}\setminus U_{\alpha^-}=\{i'_{circ}\}$, and $E(\alpha^-)=\{(j_1,i'),(j_2,i')\}$. Note that here $i<i'$ and (i'_{circ},i_{circ}) appears as a tuple in V_{α^-}.

Lemma 7.24. *$M_{\alpha^+}M_{\alpha^+}^T+M_{\alpha^-}M_{\alpha^-}^T=M_{\alpha_1}+M_{\alpha_2}+M_{\alpha_3}$ where $\alpha_1,\alpha_2,\alpha_3$ are the following shapes.*

1. *$U_{\alpha_1}\setminus V_{\alpha_1}=\emptyset$, $U_{\alpha_1}\cap V_{\alpha_1}=\{j_1,j_2,j_3,j_4,\dots,j_k\}\cup\{i_{circ}\}$, $V_{\alpha_1}\setminus U_{\alpha_1}=\emptyset$, $V(\alpha_1)\setminus(U_{\alpha_1}\cup V_{\alpha_1})=\{i'_{circ}\}$, and $E(\alpha_1)=\{(i',j_1),(i',j_1),(i',j_2),(i',j_2)\}$.*
2. *$U_{\alpha_2}\setminus V_{\alpha_2}=\{j_1\}$, $U_{\alpha_2}\cap V_{\alpha_2}=\{j_2,j_4,\dots,j_k\}\cup\{i_{circ}\}$, $V_{\alpha_2}\setminus U_{\alpha_2}=\{j'_1\}$, $V(\alpha_2)\setminus(U_{\alpha_2}\cup V_{\alpha_2})=\{i'_{circ}\}$, and $E(\alpha_2)=\{(i',j_1),(i',j'_1),(i',j_2),(i',j_2)\}$.*
3. *$U_{\alpha_3}\setminus V_{\alpha_3}=\{j_1,j_2\}$, $U_{\alpha_3}\cap V_{\alpha_3}=\{j_5,\dots,j_k\}\cup\{i_{circ}\}$, $V_{\alpha_3}\setminus U_{\alpha_3}=\{j'_1,j'_2\}$, $V(\alpha_3)\setminus(U_{\alpha_3}\cup V_{\alpha_3})=\{i'_{circ}\}$, and $E(\alpha_3)=\{(i',j_1),(i',j_2),(i',j'_1),(i',j'_2)\}$.*

Note that for these shapes, we do not assume that $i<i'$. Also note that α_1 and α_2 are improper shapes, though this does not matter for us.

Remark 7.25. *Actually, we do not need to do this computation as we will just use that $M_{\alpha^+}M_{\alpha^+}^T+M_{\alpha^-}M_{\alpha^-}^T\succeq 0$, but we include it anyways to show the similarity with the decomposition of $L_k^TL_k$.*

For pictures of these shapes, see Figure 7.2.2 below.

Proof of Lemma 7.24. We compute $M_{\alpha^+}M_{\alpha^+}^T+M_{\alpha^-}M_{\alpha^-}^T$ by considering the ribbons which appear in $M_{\alpha^+}M_{\alpha^+}^T+M_{\alpha^-}M_{\alpha^-}^T$. For these ribbons, we do not assume that $i<i'$.

1. Each ribbon R with $A_R\setminus B_R=\emptyset$, $A_R\cap B_R=\{j_1,j_2,j_3,j_4,\dots,j_k\}\cup\{i_{circ}\}$, $B_R\setminus A_R=\emptyset$, $V(R)\setminus(A_R\cup B_R)=\{i'_{circ}\}$, and $E(R)=\{(i',j_1),(i',j_1),(i',j_2),(i',j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where

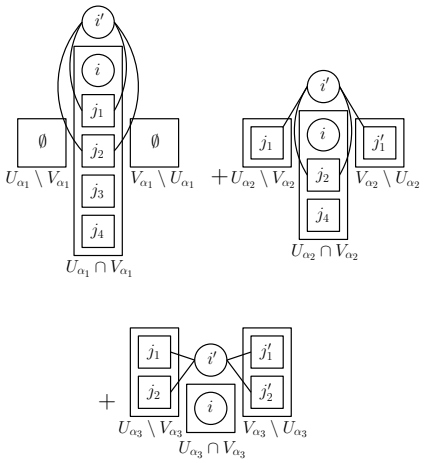

Figure 13: This figure shows the decomposition of $M_{\alpha^+}M_{\alpha^+}^T + M_{\alpha^-}M_{\alpha^-}^T$.

   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_3, j_4, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1, j_2\}$, and $E(R_1) = \{(i, j_1), (i, j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_3, j_4, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

   Whether $i < i'$ or $i' < i$ only affects whether this ribbon appears in $M_{\alpha^+}M_{\alpha^+}^T$ or $M_{\alpha^-}M_{\alpha^-}^T$.

2. Each ribbon R with $A_R \setminus B_R = \{j_1\}$, $A_R \cap B_R = \{j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \{j_1'\}$, $V(R) \setminus (A_R \cup B_R) = \{i'_{circ}\}$, and $E(R) = \{(i, j_1), (i, j_1'), (i, j_2), (i, j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_4, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1', j_2\}$, and $E(R_1) = \{(i, j_1'), (i, j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_4, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

   Whether $i < i'$ or $i' < i$ only affects whether this ribbon appears in $M_{\alpha^+}M_{\alpha^+}^T$ or $M_{\alpha^-}M_{\alpha^-}^T$.

3. Each ribbon R with $A_R \setminus B_R = \{j_1, j_2\}$, $A_R \cap B_R = \{j_5, \ldots, j_k\} \cup \{i_{circ}\}$, $B_R \setminus A_R = \{j_1', j_2'\}$, $V(R) \setminus (A_R \cup B_R) = \{i'_{circ}\}$, and $E(R) = \{(i, j_1), (i, j_1'), (i, j_2), (i, j_2')\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{i_{circ}\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_2, j_5, \ldots, j_k\}$, $B_{R_1} \setminus A_{R_1} = \{j_1', j_2'\}$, and $E(R_1) = \{(i, j_1'), (i, j_2')\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{j_1, j_2\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_2', j_5, \ldots, j_k\}$, $B_{R_2} \setminus A_{R_2} = \{i_{circ}\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

   Whether $i < i'$ or $i' < i$ only affects whether this ribbon appears in $M_{\alpha^+}M_{\alpha^+}^T$ or $M_{\alpha^-}M_{\alpha^-}^T$.

■

Lemma 7.26. *$M_{\alpha^+}M_{\alpha^-}^T + M_{\alpha^-}M_{\alpha^+}^T = M_{\alpha_4} + M_{\alpha_5} + M_{\alpha_6}$ where $\alpha_4, \alpha_5, \alpha_6$ are the following shapes.*

1. *$U_{\alpha_4} \setminus V_{\alpha_4} = \{i_{circ}\}$, $U_{\alpha_4} \cap V_{\alpha_4} = \{j_1, j_2, j_3, j_4, \ldots, j_k\}$, $V_{\alpha_4} \setminus U_{\alpha_4} = \{i'_{circ}\}$, and $E(\alpha_4) = \{(i, j_1), (i, j_2), (i', j_1), (i', j_2)\}$.*
2. *$U_{\alpha_5} \setminus V_{\alpha_5} = \{j_1\} \cup \{i_{circ}\}$, $U_{\alpha_5} \cap V_{\alpha_5} = \{j_2, j_4, \ldots, j_k\}$, $V_{\alpha_5} \setminus U_{\alpha_5} = \{j_1'\} \cup \{i'_{circ}\}$, and $E(\alpha_5) = \{(i, j_1'), (i, j_2), (i', j_1), (i', j_2)\}$.*

3. $U_{\alpha_6} \setminus V_{\alpha_6} = \{j_1, j_2\} \cup \{i_{circ}\}$, $U_{\alpha_6} \cap V_{\alpha_6} = \{j_5, \ldots, j_k\}$, $V_{\alpha_6} \setminus U_{\alpha_6} = \{j'_1, j'_2\} \cup \{i'_{circ}\}$, *and* $E(\alpha_6) = \{(i', j_1), (i', j_2), (i, j'_1), (i, j'_2)\}$.

For pictures of these shapes, see Figure 7.2.2 below.

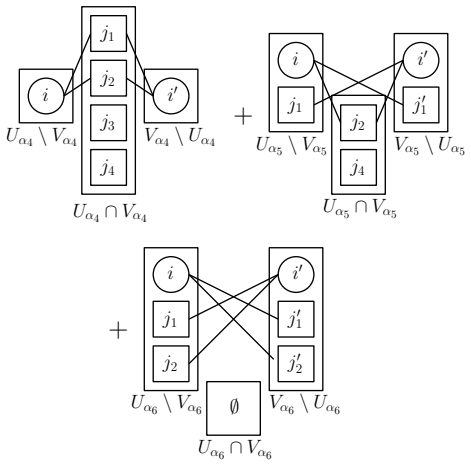

Figure 14: This figure shows the decomposition of $M_{\alpha^+} M_{\alpha^+}^T + M_{\alpha^-} M_{\alpha^-}^T$.

Proof of Lemma 7.26. We compute $M_{\alpha^+} M_{\alpha^-}^T + M_{\alpha^-} M_{\alpha^+}^T$ by considering the ribbons which appear in $M_{\alpha^+} M_{\alpha^-}^T + M_{\alpha^-} M_{\alpha^+}^T$. For these ribbons, we do not assume that $i < i'$.

1. Each ribbon R with $A_R \setminus B_R = \{i_{circ}\}$, $A_R \cap B_R = \{j_1, j_2, j_3, j_4, \ldots, j_k\}$, $B_R \setminus A_R = \{i'_{circ}\}$, $V(R) \setminus (A_R \cup B_R) = \emptyset$, and $E(R) = \{(i, j_1), (i, j_2), (i', j_1), (i', j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{j_1, j_2\}$, $A_{R_1} \cap B_{R_1} = \{j_3, j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_{R_1} \setminus A_{R_1} = \{i'_{circ}\}$, and $E(R_1) = \{(i', j_1), (i', j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{i_{circ}\}$, $A_{R_2} \cap B_{R_2} = \{j_3, j_4, \ldots, j_k\} \cup \{i'_{circ}\}$, $B_{R_2} \setminus A_{R_2} = \{j_1, j_2\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

   Whether $i < i'$ or $i' < i$ only affects whether this ribbon appears in $M_{\alpha^+} M_{\alpha^-}^T$ or $M_{\alpha^-} M_{\alpha^+}^T$.
2. Each ribbon R with $A_R \setminus B_R = \{j_1\} \cup \{i_{circ}\}$, $A_R \cap B_R = \{j_2, j_4, \ldots, j_k\}$, $B_R \setminus A_R = \{j'_1\} \cup \{i'_{circ}\}$, $V(R) \setminus (A_R \cup B_R) = \emptyset$, and $E(R) = \{(i, j_1), (i, j_2), (i', j'_1), (i', j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{j_1, j_2\}$, $A_{R_1} \cap B_{R_1} = \{j'_1, j_4, \ldots, j_k\} \cup \{i_{circ}\}$, $B_{R_1} \setminus A_{R_1} = \{i'_{circ}\}$, and $E(R_1) = \{(i', j_1), (i', j_2)\}$.
   (b) $A_{R_2} \setminus B_{R_2} = \{i_{circ}\}$, $A_{R_2} \cap B_{R_2} = \{j_1, j_4, \ldots, j_k\} \cup \{i'_{circ}\}$, $B_{R_2} \setminus A_{R_2} = \{j'_1, j_2\}$, and $E(R_2) = \{(i, j'_1), (i, j_2)\}$.

   Whether $i < i'$ or $i' < i$ only affects whether this ribbon appears in $M_{\alpha^+} M_{\alpha^-}^T$ or $M_{\alpha^-} M_{\alpha^+}^T$.
3. Each ribbon R with $A_R \setminus B_R = \{j_1, j_2\} \cup \{i_{circ}\}$, $A_R \cap B_R = \{j_5, \ldots, j_k\}$, $B_R \setminus A_R = \{j'_1, j'_2\} \cup \{i'_{circ}\}$, $V(R) \setminus (A_R \cup B_R) = \emptyset$, and $E(R) = \{(i, j'_1), (i, j'_2), (i', j_1), (i', j_2)\}$ appears in exactly one way as the composition of the ribbons R_1 and R_2 where
   (a) $A_{R_1} \setminus B_{R_1} = \{j_1, j_2\}$, $A_{R_1} \cap B_{R_1} = \{j_1, j_2, j_5, \ldots, j_k\} \cup \{i_{circ}\}$, $B_{R_1} \setminus A_{R_1} = \{i'_{circ}\}$, and $E(R_1) = \{(i, j'_1), (i, j'_2)\}$.

(b) $A_{R_2} \setminus B_{R_2} = \{i_{circ}\}$, $A_{R_2} \cap B_{R_2} = \{j_1', j_2', j_5, \ldots, j_k\} \cup \{i_{circ}'\}$, $B_{R_2} \setminus A_{R_2} = \{j_1', j_2'\}$, and $E(R_2) = \{(i, j_1), (i, j_2)\}$.

Whether $i < i'$ or $i' < i$ only affects whether this ribbon appears in $M_{\alpha^+} M_{\alpha^-}^T$ or $M_{\alpha^-} M_{\alpha^+}^T$.

■

We now consider the norm bounds for these terms

1. $\|M_{\alpha_4}\|$ is $\tilde{O}(m)$.
2. $\|M_{\alpha_5}\|$ is $\tilde{O}(m)$.
3. $\|M_{\alpha_6}\|$ is $\tilde{O}(n^2)$.

This means that M_{α_6} cannot be ignored, but this is fine. In fact, the M_{α_6} here will cancel with the M_{α_6} that appears in the decomposition of $L_k^T L_k$

Corollary 7.27. *For all $k \geq 4$, $\left\| N_k {N_k}^T - 4M_{\alpha^+} M_{\alpha^+}^T - 4M_{\alpha^-} M_{\alpha^-}^T + 4M_{\alpha_6} \right\|$ is $\tilde{O}(m)$.*

7.2.3 Putting Everything Together

We now put everything together to prove that for all k, the minimum nonzero eigenvalue of $L_k^T L_k$ is $2n^2 - \tilde{O}(n\sqrt{m})$.

1. For $k = 2$, by Corollary 7.20, $\left\| L_k^T L_k - 2n^2 M_{Id_{k-2,1}} \right\|$ is $\tilde{O}(n\sqrt{m})$ so the minimum eigenvalue of $L_k^T L_k$ is $2n^2 - \tilde{O}(n\sqrt{m})$.
2. For $k = 3$, by Corollary 7.20, $\left\| L_k^T L_k - 2n^2 M_{Id_{k-2,1}} - 4nM_{\beta_2} M_{\beta_2}^T \right\|$ is $\tilde{O}(n\sqrt{m})$ so the minimum eigenvalue of $L_k^T L_k$ is $2n^2 - \tilde{O}(n\sqrt{m})$.
3. For $k \geq 4$, by Corollary 7.20, $\left\| L_k^T L_k - 2n^2 M_{Id_{k-2,1}} - 4nM_{\beta_2} M_{\beta_2}^T - 4M_{\beta_3} M_{\beta_3}^T - 4M_{\alpha_6} \right\|$ is $\tilde{O}(n\sqrt{m})$. By Corollary 7.27, $\left\| N_k {N_k}^T - 4M_{\alpha^+} M_{\alpha^+}^T - 4M_{\alpha^-} M_{\alpha^-}^T + 4M_{\alpha_6} \right\|$ is $\tilde{O}(m)$. Combining these equations,

$$\left\| L_k^T L_k + N_k {N_k}^T - 2n^2 M_{Id_{k-2,1}} - 4nM_{\beta_2} M_{\beta_2}^T - 4M_{\beta_3} M_{\beta_3}^T - 4M_{\alpha^+} M_{\alpha^+}^T - 4M_{\alpha^-} M_{\alpha^-}^T \right\|$$

is $\tilde{O}(n\sqrt{m})$ so the minimum eigenvalue of $L_k^T L_k + N_k {N_k}^T$ is $2n^2 - \tilde{O}(n\sqrt{m})$. Since the minimum nonzero eigenvalue of $L_k^T L_k$ is at least as large as the minimum eigenvalue of $L_k^T L_k + N_k {N_k}^T$, the nonzero eigenvalue of $L_k^T L_k$ is $2n^2 - \tilde{O}(n\sqrt{m})$.

This implies that the minimum nonzero eigenvaue of QQ^T is $2n^2 - \tilde{O}(n\sqrt{m})$, as needed.

8 Open Problems

We conjecture that for the Planted Affine Planes problem, the problem remains difficult even with the number of vectors increased to $m = n^{2-\varepsilon}$.

Conjecture 8.1. *Theorem 1.4 holds with the bound on the number of sampled vectors m loosened to $m \leq n^{2-\varepsilon}$.*

The reason for the upper bound comes from Remark 5.9. Analyzing $\widetilde{\mathbb{E}}[1]$ is an established way to hypothesize about the power of SoS in hypothesis testing problems (see [HKP+17], [Hop18]).

Dual to the Planted Affine Planes problem, we conjecture a similar bound for Planted Boolean Vector problem whenever $d \geq n^{1/2+\varepsilon}$.

Conjecture 8.2. *Theorem 1.5 holds with the bound on the dimension p of a random subspace loosened to $p \geq n^{1/2+\varepsilon}$.*

We conjecture that the Planted Boolean Vector problem/Planted Affine Planes problem is still hard for SoS if the input is no longer i.i.d. Gaussian or boolean entries, but is drawn from a "random enough" distribution. For example, if in the random instance of PAP the vectors d_u are i.i.d. samples from S^n, or a random orthonormal system, degree n^δ SoS should still believe the instance is satisfiable (after appropriate normalization of v). Or, taking the view of Planted Boolean Vector, if the subspace is the eigenspace of the bottom eigenvectors of a random adjacency matrix, the instance should still be difficult. This last setting arises in MaxCut, for which we conjecture the following.

Conjecture 8.3. *Let $d \geq 3$, and let G be a random d-regular graph on n vertices. For some $\delta > 0$, w.h.p. there is a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}$ on boolean variables x_i with MaxCut value at least*

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d}(1 - o_{d,n}(1))$$

The above expression is w.h.p. the value of the spectral relaxation for MaxCut, therefore qualitatively this conjecture expresses that degree n^δ SoS cannot significantly tighten the basic spectral relaxation.

We should remark that, with respect to the goal of showing SoS cannot significantly outperform the Goemans-Williamson relaxation, random instances are not integrality gap instances. The main difficulty in comparing (even degree 4) SoS to the Goemans-Williamson algorithm seems to be the lack of a candidate hard input distribution.

Evidence for this conjecture comes from the fact that the only property required of the random inputs $d_1, \ldots, d_m$ was that norm bounds hold for the graph matrix with Hermite polynomial entries. When the variables $\{d_{u,i}\}$ are i.i.d from some other distribution, if we use graph matrices for the orthonormal polynomials under the distribution and assuming suitable bounds on the moments of the distribution, the same norm bounds hold [AMP20]. When $d_u \in_R S^n$ or another distribution for which the coordinates are not i.i.d, it seems likely that if we use e.g. the spherical harmonics then similar norm bounds hold, but this is not proven.

Acknowledgements

We thank Madhur Tulsiani and Pravesh K. Kothari for several enlightening discussions in the initial phases of this work. G.R. thanks Sidhanth Mohanty for useful discussions regarding the Sherrington-Kirkpatrick problem. We also thank the anonymous reviewers for their useful suggestions on improving the text.

References

[AMP20] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. abs/1604.03423, 2020. URL: https://arxiv.org/abs/1604.03423, arXiv:1604.03423. 4, 7, 57, 59, 60

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, 2004. 1

[BHK+16] B. Barak, S. B. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem, 2016. 1, 4, 13

[BKW19] Afonso S. Bandeira, Dmitriy Kunisky, and Alexander S. Wein. Computational hardness of certifying bounds on constrained pca problems, 2019. arXiv:1902.07324. 2

[FKP19] N. Fleming, P. Kothari, and T. Pitassi. *Semialgebraic Proofs and Efficient Algorithm Design*. 2019. 1

[FS02] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for max cut. *Random Structures & Algorithms*, 20(3):403–440, 2002. 2

[GW95] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995. Preliminary version in *Proc. of STOC'94*. 1

[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 *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, pages 720–731. IEEE, 2017. 1, 57

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006. 1

[Hop18] Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method. 2018. 1, 57

[HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-squares proofs. In *Conference on Learning Theory*, pages 956–1006, 2015. 1

[Kar72] R.M. Karp. Reducibility among combinatorial problems. In R.E. Miller and J.W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972. 1

[KB19] Dmitriy Kunisky and Afonso S. Bandeira. A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian. abs/1907.11686, 2019. URL: https://arxiv.org/abs/1907.11686, arXiv:1907.11686. 2, 5

[Las15] Jean Bernard Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2015. doi:10.1017/CBO9781107447226. 1

[Mon19] A. Montanari. Optimization of the sherrington-kirkpatrick hamiltonian. In *Proceedings of the 60th IEEE Symposium on Foundations of Computer Science*, pages 1417–1433, 2019. 2, 5

[MRX19] Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: Degree-2 to degree-4. abs/1911.01411, 2019. URL: https://arxiv.org/abs/1911.01411, arXiv:1911.01411. 2, 3, 4, 5, 6, 40

[MS16] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the 48th ACM Symposium on Theory of Computing*, pages 814–827, 2016. 2, 5

[OVW16] Sean O'Rourke, Van Vu, and Ke Wang. Eigenvectors of random matrices. *J. Comb. Theory Ser. A*, 144(C):361–442, November 2016. 40

[Pan14] Dmitry Panchenko. The parisi formula for mixed p-spin models. *Ann. Probab.*, 42(3):946–958, 05 2014. 2

[Par79] G. Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43:1754–1756, Dec 1979. 2

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 245–254, 2008. 1

[Rom05] Steven Roman. *The umbral calculus*. Springer, 2005. 44

[RW97] Gian-Carlo Rota and Timothy C. Wallstrom. Stochastic integrals: a combinatorial approach. *Ann. Probab.*, 25(3):1257–1283, 1997. doi:10.1214/aop/1024404513. 31

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35:1792–1796, Dec 1975. 2, 6

[Tal06] Michel Talagrand. The parisi formula. *Annals of mathematics*, pages 221–263, 2006. 2

[Wig93] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions i. In *The Collected Works of Eugene Paul Wigner*, pages 524–540. Springer, 1993. 40

A Norm Bounds

The norm bounds we use come from applying the trace power method in [AMP20]. The paper [AMP20] uses a slightly different definition of matrix index. They define a *matrix index piece* as a tuple of distinct elements from either $\mathcal{C}_m$ or $\mathcal{S}_n$ along with a fixed integer denoting multiplicity. A matrix index is then a set of matrix index pieces. Our graph matrix M_α appears as a submatrix of those matrices: for a given set of square vertices, order the squares in increasing order in a tuple, and assign it multiplicity 1. Hence the same norm bounds apply.

Boolean norm bounds:

Lemma A.1. *Let $V_{rel}(\alpha) := V(\alpha) \setminus (U_\alpha \cap V_\alpha)$. There is a universal constant C such that the following norm bound holds for all proper shapes α w.h.p.:*

$$\|M_\alpha\| \leq 2 \cdot (|V(\alpha)| \cdot \log(n))^{C \cdot |V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}$$

Proof. From Corollary 8.13 of [AMP20], with probability at least $1 - \varepsilon$ for a fixed shape α,

$$\|M_\alpha\| \leq 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(6e \left\lceil \frac{\log\left(\frac{n^{w(S_{\min})}}{\varepsilon}\right)}{6\,|V_{rel}(\alpha)|} \right\rceil\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha)) - w(S_{\min}) + w(W_{iso})}{2}}$$

Letting N_k be the number of distinct shapes on k vertices (either circles or squares), we apply the corollary with $\varepsilon = 1/(mnN_{|V(\alpha)|})$. Union bounding, the failure probability across all shapes of size k is at most $1/mn$, and since the number of vertices in a shape is at most $m+n \le 2m$, we have a bound that holds with high probability for all shapes. It remains to simplify the exact bound.

Proposition A.2. $N_k \le 8^k 2^{k^2}$

Proof. The following process forms all shapes on k vertices: starting from k formal variables, assign each variable to be either a circle or a square, decide whether each variable is in U_α and/or V_α, then among the k^2 variable pairs put any number of edges. ∎

We also bound $n^{w(S_{\min})} \le (mn)^{|V(\alpha)|}$.

$$\begin{aligned}
\|M_\alpha\| &\le 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(6e \left\lceil \frac{\log\left(n^{w(S_{\min})} \cdot mnN_{|V(\alpha)|}\right)}{6\,|V_{rel}(\alpha)|} \right\rceil\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\le 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(12e \log\left(n^{w(S_{\min})} \cdot mnN_{|V(\alpha)|}\right)\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\le 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(12e \log\left((mn)^{|V(\alpha)|} \cdot mn \cdot 8^{|V(\alpha)|} 2^{|V(\alpha)|^2}\right)\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\le 2\,|V(\alpha)|^{|V_{rel}(\alpha)|} \cdot \left(100e\,|V(\alpha)|^2 \log(mn)\right)^{|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}} \\
&\le 2 \cdot (|V(\alpha)| \cdot \log(mn))^{3\cdot|V_{rel}(\alpha)|} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}}
\end{aligned}$$

Note that we now assume $m \le n^2$. ∎

We have the following norm bound for Hermite shapes. For a Hermite shape α, define the *total size* to be $|U_\alpha| + |V_\alpha| + |W_\alpha| + |E(\alpha)|$.

Lemma A.3. *Let $V_{rel}(\alpha) := V(\alpha) \setminus (U_\alpha \cap V_\alpha)$ as sets. There is a universal constant C such that the following norm bound holds for all proper shapes α with total size at most n w.h.p.:*

$$\|M_\alpha\| \le 2 \cdot (|V(\alpha)| \cdot (1 + |E(\alpha)|) \cdot \log(n))^{C\cdot(|V_{rel}(\alpha)|+|E(\alpha)|)} \cdot n^{\frac{w(V(\alpha))-w(S_{\min})+w(W_{iso})}{2}}$$

The proof performs the same calculation starting from [AMP20, Corollary 8.15]. Note that in our notation, $l(\alpha) = |E(\alpha)|$. There is a further difference which is that [AMP20] uses normalized Hermite polynomials whereas we use unnormalized Hermite polynomials; this contributes the additional term $\prod_{e\in E(\alpha)} l(e)! \le (1+|E(\alpha)|)^{|E(\alpha)|}$. We must replace Proposition A.2 with the following:

Proposition A.4. *The number of Hermite shapes with total size k is at most $k2^k(k+1)^{2k+k^2}$.*

Proof. Such a shape has at most k distinct variable vertices. Each of these is either a circle or a square. Each variable can be in U_α with multiplicity between 0 and (at most) k, and also in V_α with multiplicity between 0 and k. The k^2 possible pairs of vertices can have edge multiplicity in $E(\alpha)$ between 0 and k. ∎

B Properties of $e(k)$

We establish some properties of the $e(k)$ used in the analysis. Recall that $e(k) = \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}[x_1 \dots x_k]$ where $\mathcal{S}(\sqrt{n}) := \{x \in \{\pm 1\}^n \mid \sum_{i=1}^n x_i = \sqrt{n}\}$.

Claim B.1. $e(2) = 0$.

Proof. Fix $y \in \mathcal{S}(\sqrt{n})$. Note that $(\sum_{i=1}^n y_i)^2 = n$ implying $\sum_{i<j} y_i y_j = 0$. Using this fact, we get

$$\mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})}[x_1 x_2] = \mathop{\mathbb{E}}_{\sigma \in S_n} y_{\sigma(1)} y_{\sigma(2)} = 0,$$

concluding the proof. ∎

Definition B.2. *We say that a tuple $\lambda = (\lambda_1, \dots, \lambda_k)$ of non-negative integers is a partition of k provided $\sum_{i=1}^k \lambda_i = k$ and $\lambda_1 \geq \cdots \geq \lambda_k$. We use the notation $\lambda \vdash k$ to denote a partition of k. We refer to λ_i as a row/part of λ.*

In the following, we will dealing polynomials that can be indexed by integer partitions. For this reason, we now fix a notation for partitions and some associated objects.

Definition B.3. *The transpose of partition $\lambda = (\lambda_1, \dots, \lambda_k)$ is denoted λ^t and defined as $\lambda_i^t = \big|\{j \in [k] \mid \lambda_j \geq i\}\big|$.*

Remark B.4. *For a partition $\lambda \vdash k$, λ_1^t is the number of rows/parts of λ.*

Definition B.5. *The automorphism group of a partition $\mathrm{Aut}(\lambda) \leq S_{\lambda_1^t}$ is the group generated by transpositions (i, j) of rows $\lambda_i = \lambda_j$.*

Remark B.6. *Let $\lambda \vdash k$ and $p_1(\lambda), \dots, p_k(\lambda)$ be such that $p_i(\lambda) = \big|\{j \in [\lambda_1^t] \mid \lambda_j = i\}\big|$. Then $\mathrm{Aut}(\lambda) \simeq S_{p_1} \times \cdots \times S_{p_k}$.*

Lemma B.7. *We have*

$$\sum_{\lambda \vdash k} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1} \dots x_k^{\lambda_k}\right] = n^{k/2}.$$

Proof. For $x \in \mathcal{S}(\sqrt{n})$, we have $(\sum_{i=1}^n x_i)^k = n^{k/2}$. Then expanding $(\sum_{i=1}^n x_i)^k$ in the previous equations and taking the expectation over $\mathcal{S}(\sqrt{n})$ on both sides yields the result of the lemma (after appropriately collecting terms). ∎

Claim B.8. *Let $\lambda \vdash k$. We have*

$$(n)_{\lambda_1^t} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1} \dots x_k^{\lambda_k}\right] \right| \leq 3^{k^3} \cdot n^{k/2}.$$

Proof. We induct on k. For $k = 1$, we have $n \cdot \left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}[x_1]\right| = \sqrt{n} \leq 3 \cdot n^{1/2}$. Now, suppose $k \geq 2$. We consider three cases:

1. Case $\lambda_1 \geq 3$: Let λ' be the partition obtained from λ by removing two boxes from λ_1. Note that $\lambda_1^t = (\lambda')_1^t \leq k - 2$ and $\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1'} \dots x_{k-2}^{\lambda_{k-2}'}\right] = \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1} \dots x_{k-2}^{\lambda_{k-2}}\right]$. By the induction hypothesis, we have $(n)_{(\lambda')_1^t} \cdot \left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}\left[x_1^{\lambda_1'} \dots x_{k-2}^{\lambda_{k-2}'}\right]\right| \leq 3^{(k-2)^2} \cdot n^{(k-2)/2}$.

2. Case $\lambda_1 = 2$: Let λ' be the partition obtained from λ by removing λ_1. Note that $\lambda_1^t = (\lambda')_1^t + 1 \le k-2$. By the induction hypothesis, we have

$$(n)_{\lambda_1^t} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_{k-2}^{\lambda_{k-2}} \right] \right| \le n \cdot (n)_{(\lambda')_1^t} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda'_1} \dots x_{k-2}^{\lambda'_{k-2}} \right] \right| \le 3^{(k-2)^3} \cdot n^{k/2}.$$

3. Case $\lambda_1 = 1$: To bound $(n)_k \cdot \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right]$, we use Lemma B.7 and the two preceding cases. Let $p(k)$ be the partition function, i.e., $p(k) = |\{\lambda \vdash k\}|$. We deduce that

$$\begin{aligned}
(n)_k \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| &\le n^{k/2} + \sum_{\lambda \vdash k:\ \lambda_1 \ge 2} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\
&\le n^{k/2} + k! \sum_{\lambda \vdash k:\ \lambda_1 \ge 2} (n)_{\lambda_1^t} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\
&\le n^{k/2} + k! \sum_{\lambda \vdash k:\ \lambda_1 \ge 3} (n)_{\lambda_1^t} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| + \\
&\qquad k! \sum_{\lambda \vdash k:\ \lambda_1 = 2} (n)_{\lambda_1^t} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\
&\le 3^{(k-2)^3} \cdot k! \cdot (1 + p(k) + k) \cdot n^{k/2} \le 3^{k^3} \cdot n^{k/2},
\end{aligned}$$

as desired.

■

Claim B.9. *Suppose $k < \sqrt{n}/2$. We have*

$$\left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} [x_1 \dots x_k] \right| \le 2 \cdot 3^{k^3} \cdot n^{-k/2}.$$

Proof. Follows from Claim B.8 and the bound on k. ■

Remark B.10. *In Claim B.9, the factor 3^{k^3} is too lossy to allow a meaningful bound with $k = n^\varepsilon$, where $\varepsilon > 0$ is a constant.*

Refining the ideas of Claim B.8, we prove a stronger lemma below which will imply a tighter bound on $e(k)$ sufficient for our application.

Lemma B.11. *There exists an universal constant $C \ge 1$ such that*

$$\sum_{\lambda \vdash k} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \le k^{C \cdot k} \cdot n^{k/2}. \tag{3}$$

In particular, for $n \ge 6$, Eq. (3) holds with $C = 2$.

Proof. We induct on k. For $k = 1$, we have $n \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} x_1 \right| \le \sqrt{n}$ as desired. Using $e(2) = 0$ from Claim B.1 and the case $k = 1$ of Eq. (3), we get that Lemma B.11 also holds for $k = 2$. Now,

consider $k \geq 3$. Let $\Lambda_1 = \{\lambda \vdash k \mid \lambda_1 = 1\}$, $\Lambda_2 = \{\lambda \vdash k \mid \lambda_1 = 2\}$ and $\Lambda_{\geq 3} = \{\lambda \vdash k \mid \lambda_1 = 3\}$. Note that $\Lambda_1 \sqcup \Lambda_2 \sqcup \Lambda_{\geq 3} = \{\lambda \vdash k\}$ and $|\Lambda_1| = 1$.

For convenience define a_λ to be the term associated to $\lambda \vdash k$ on the LHS of Eq. (3), i.e.,

$$a_\lambda = \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right|.$$

First we bound the contribution of the terms associated to partitions from $\Lambda_{\geq 3}$ in the LHS of Eq. (3). Let λ' be the partition obtained from λ by removing two boxes from λ_1. Note that $\lambda_1^t = (\lambda')_1^t \leq k-2$ and $\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda'_1} \dots x_{k-2}^{\lambda'_{k-2}} \right] = \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_{k-2}^{\lambda_{k-2}} \right]$. Thus,

$$\begin{aligned}
a_\lambda &= \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\
&= \frac{k(k-1)}{\lambda_1(\lambda_1 - 1)} \cdot \frac{|\mathrm{Aut}(\lambda')|}{|\mathrm{Aut}(\lambda)|} \frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')_1^t}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda'_1} \dots x_{k-2}^{\lambda'_{k-2}} \right] \right| \\
&= k^2 \cdot \frac{|\mathrm{Aut}(\lambda')|}{|\mathrm{Aut}(\lambda)|} \cdot a_{\lambda'} \leq k^3 \cdot a_{\lambda'},
\end{aligned}$$

since $|\mathrm{Aut}(\lambda')| \,/\, |\mathrm{Aut}(\lambda)| \leq k-2 \leq k$. For each $\lambda' \vdash k-2$, we can form a partition $\lambda \vdash k$ in $k-2 \leq k$ ways by adding two blocks to a single row of λ'. Hence, we have

$$\sum_{\lambda \in \Lambda_{\geq 3}} a_\lambda \leq k \cdot \sum_{\lambda' \vdash k-2} k^3 \cdot a_{\lambda'} \leq k^4 \cdot k^{C \cdot (k-2)} \cdot n^{(k-2)/2}, \tag{4}$$

where the last equality follows from the induction hypothesis.

Now we bound the contribution of the terms a_λ associated to partitions λ from Λ_2 in the LHS of Eq. (3). Let $i \geq 1$ be the number of parts of size two of λ and let λ' be the partition obtained from λ by removing these i parts of size two. Note that $\lambda_1^t = (\lambda')_1^t + i \leq k-1$. We have

$$\begin{aligned}
a_\lambda &= \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \\
&\leq n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{|\mathrm{Aut}(\lambda')|}{|\mathrm{Aut}(\lambda)|} \cdot \frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')_1^t}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right| \\
&= n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot \frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')_1^t}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right|,
\end{aligned}$$

where in the last equality we used $|\mathrm{Aut}(\lambda')| \,/\, |\mathrm{Aut}(\lambda)| = 1/(i!)$. Since $\lambda \in \Lambda_2$ is uniquely specified by its number of parts of size two, applying the induction hypothesis we have

$$\begin{aligned}
\sum_{\lambda \in \Lambda_2} a_\lambda &\leq \sum_{i=1}^{\lfloor k/2 \rfloor} n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot \left(\frac{\lambda'!}{\lambda'_1! \cdots \lambda'_k!} \cdot \frac{(n)_{(\lambda')_1^t}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right| \right) \\
&\leq \sum_{i=1}^{\lfloor k/2 \rfloor} n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot k^{C \cdot (k-2i)} \cdot n^{(k-2i)/2} \\
&\leq k^{C \cdot (k-1)} \cdot n^{k/2} \cdot \sum_{i=0}^{\infty} k^{-C \cdot i} \leq \frac{3}{2} \cdot k^{C \cdot (k-1)} \cdot n^{k/2},
\end{aligned}$$

where in the last inequality we used $k \geq 3$ and $C \geq 1$.

Finally, we consider the case $\lambda_1 = 1$. To bound a_λ, we use Lemma B.7 and the two preceding cases. We deduce that

$$\begin{aligned} a_\lambda &\leq n^{k/2} + \sum_{\mu \in \Lambda_2} a_\mu + \sum_{\mu \in \Lambda_{\geq 3}} a_\mu \leq n^{k/2} + k^4 \cdot k^{C\cdot(k-2)} \cdot n^{(k-2)/2} + \frac{3}{2} \cdot k^{C\cdot(k-1)} \cdot n^{k/2} \\ &= k^{C\cdot k} \cdot n^{k/2} \left(\frac{1}{k^{C\cdot k}} + \frac{k^4}{n \cdot k^{2\cdot C}} + \frac{3}{2 \cdot k^C} \right). \end{aligned}$$

We can bound the LHS of Eq. (3) as

$$\begin{aligned} \sum_{\mu \in \Lambda_1} a_\mu + \sum_{\mu \in \Lambda_2} a_\mu + \sum_{\mu \in \Lambda_{\geq 3}} a_\mu &\leq k^{C\cdot k} \cdot n^{k/2} \left(\frac{1}{k^{C\cdot k}} + \frac{2 \cdot k^4}{n \cdot k^{2\cdot C}} + \frac{3}{k^C} \right) \\ &\leq k^{C\cdot k} \cdot n^{k/2}, \end{aligned}$$

provided $C > 0$ is a sufficiently large constant. In particular, the constant C can be taken to be 2 for $n \geq 6$. ∎

Corollary B.12. *We have*

$$\left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} [x_1 \dots x_k] \right| \leq k^{3\cdot k} \cdot n^{-k/2}.$$

Proof. Suppose $k \leq \sqrt{n}$. Note that Lemma B.11 implies that for $\lambda \vdash k$ with λ_1 there exists a constant $C > 0$ such that

$$\begin{aligned} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| &= (n)_k \cdot \left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} [x_1 \dots x_k] \right| \\ &\leq k^{C\cdot k} \cdot n^{k/2}. \end{aligned}$$

Simplifying and using the assumption $k \leq \sqrt{n}$, we obtain

$$\left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} [x_1 \dots x_k] \right| \leq \frac{k^{C\cdot k} \cdot n^{-k/2}}{\prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right)} \leq 2 \cdot k^{C\cdot k} \cdot n^{-k/2}.$$

Furthermore, for $n \geq 6$, Lemma B.11 allows us to choose $C = 2$. Since $\left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} [x_1] \right| = 1/\sqrt{n}$, the simpler bound applies for all values of k

$$\left| \mathop{\mathbb{E}}_{x \in \mathcal{S}(\sqrt{n})} [x_1 \dots x_k] \right| \leq k^{3\cdot k} \cdot n^{-k/2},$$

Now the assumption $n \geq 6$ can be removed since, for $k \geq 2$, we have $(k^3/\sqrt{n})^k \geq 1$, where 1 is the trivial bound. Similarly, our initial assumption of $k \leq \sqrt{n}$ can also be removed as the bound also becomes trivial in the regime $k > \sqrt{n}$. ∎

C Combinatorial Proof of Lemma 4.5

In this appendix, we give a combinatorial proof of Lemma 4.5. We recall the statement of Lemma 4.5 here.

Lemma C.1. *Let $\alpha \in \mathbb{N}^n$. When v is fixed and b is fixed (not necessarily +1 or -1) and $d \sim N(0, I)$ conditioned on $\langle v, d\rangle = b\,\|v\|$,*

$$\mathop{\mathbb{E}}_{d}[h_\alpha(d)] = \frac{v^\alpha}{\|v\|^{|\alpha|}} \cdot h_{|\alpha|}(b).$$

Proof. Again, it is sufficient to prove this lemma when $\|v\| = 1$. For this proof, we need the following description of Hermite polynomials in terms of matchings and Isserlis' Theorem/Wick's Theorem.

Fact C.2.

$$h_k(x) = \sum_{M:M \text{ is a matching on } [k]} (-1)^{|M|} x^{k-2|M|}$$

Theorem C.3 (Isserlis' Theorem/Wick's Theorem)**.** *For any vectors $u_1, \ldots, u_k$,*

$$E_{x\sim N(0,I)}\left[\prod_{j=1}^{k} \langle x, u_j\rangle\right] = \sum_{M:M \text{ is a perfect matching on } [k]} \prod_{(i,j)\in M} \langle u_i, u_j\rangle$$

The idea behind this proof is to break up each coordinate vector e_i into a component which is parallel to v and a component which is perpendicular to v.

Definition C.4. *For each coordinate i, define $e_i^\perp = e_i - v_i v$*

Proposition C.5. *For any coordinate i, $\langle e_i^\perp, e_i^\perp\rangle = 1 - {v_i}^2$. For any pair of distinct coordintes i and i', $\langle e_i^\perp, e_{i'}^\perp\rangle = -v_i v_{i'}$*

Proof. Observe that for all i,

$$\langle e_i^\perp, e_i^\perp\rangle = \langle e_i - v_i v, e_i - v_i v\rangle = \langle e_i, e_i\rangle - 2v_i\langle v, e_i\rangle + {v_i}^2\langle v, v\rangle = 1 - {v_i}^2$$

and if i and i' are distinct then

$$\langle e_i^\perp, e_{i'}^\perp\rangle = \langle e_i - v_i v, e_{i'} - v_{i'} v\rangle = \langle e_i, e_{i'}\rangle - v_i\langle v, e_{i'}\rangle - v_{i'}\langle e_i, v\rangle + v_i v_{i'}\langle v, v\rangle = -v_i v_{i'}$$

■

To evaluate $\mathbb{E}_d[h_\alpha(d)]$, we proceed as follows:

1. Break up each $d_i = \langle d, e_i\rangle$ as $d_i = \langle bv, e_i\rangle + \langle d^\perp, e_i\rangle = bv_i + \langle d^\perp, e_i^\perp\rangle$ where $d^\perp$ is the component of d which is orthogonal to v.
2. Observe that since each $e_i^\perp$ is orthogonal to v, we can replace $d^\perp$ by a random vector $d' \sim N(0, I)$.
3. Apply Isserlis' Theorem/Wick's Theorem to evaluate these terms.

For this calculation, it is convenient to think of α as a tuple of $|\alpha|$ elements where each $i \in [n]$ appears α_i times.

Definition C.6. *For each $j \in [|\alpha|]$, we define $\alpha(j)$ to be the index i such that $\sum_{i'=1}^{i-1} \alpha_{i'} < j$ and $\sum_{i'=1}^{i} \alpha_{i'} \geq j$. For example, if $\alpha = (2,1,0,3)$ then $\alpha(1) = \alpha(2) = 1$, $\alpha(3) = 2$, and $\alpha(4) = \alpha(5) = \alpha(6) = 4$.*

In the special case when $\alpha(1), \dots, \alpha(|\alpha|)$ are all distinct,

$$\mathop{\mathbb{E}}_{d}[h_\alpha(d)] = \mathop{\mathbb{E}}_{d}\left[\prod_{j=1}^{|\alpha|} \langle d, e_{\alpha(j)}\rangle\right] = \mathop{\mathbb{E}}_{d' \sim N(0,I)}\left[\prod_{j=1}^{|\alpha|}\left(bv_{\alpha(j)} + \langle d', e_{\alpha(j)}^\perp\rangle\right)\right]$$

In this case, we can associate a matching M to each term we get after applying Isserlis' Theorem/Wick's Theorem as follows:

1. For each $j \in |\alpha|$ where we have the $bv_{\alpha(j)}$ term, we take j to be isolated.
2. For each pair of distinct $j, j' \in |\alpha|$ such that we have the term $\langle e_{\alpha(j)}^\perp, e_{\alpha(j')}^\perp\rangle$ (which only happens if we start with the $\langle d', e_{\alpha(j)}^\perp\rangle$ and $\langle d', e_{\alpha(j')}^\perp\rangle$ terms and $e_{\alpha(j)}^\perp$ and $e_{\alpha(j')}^\perp$ are paired together after applying Isserlis' Theorem/Wick's Theorem), we add an edge between j and j' in M.

We now have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{d}[h_\alpha(d)] &= \sum_{M : M \text{ is a matching on } [|\alpha|]} \left(\prod_{(j,j') \in M} -v_{\alpha(j)}v_{\alpha(j')}\right)\left(\prod_{j : j \text{ is unmatched by } M} bv_{\alpha(j)}\right) \\
&= \left(\sum_{M : M \text{ is a matching on } [|\alpha|]} (-1)^{|M|} b^{|\alpha| - 2|M|}\right)\prod_{j=1}^{|\alpha|} v_{\alpha(j)} \\
&= h_{|\alpha|}(b)v^\alpha
\end{aligned}$$

For the general case, we use a similar idea although it is somewhat more complicated. In particular, we associate a multi-colored matching $M = M_{blue} \cup M_{red} \cup M_{purple}$ to each term. The idea is that whenever we have a blue edge, we could have had a red edge instead and vice versa, so we can combine terms with red and blue edges to make purple edges which gives us an ordinary matching as before. More precisely, the idea is as follows.

1. When we expand out $h_\alpha(d)$ in terms of matchings, we take M_{blue} to be the union of these matchings.
2. For each $j \in |\alpha|$ where we have the $bv_{\alpha(j)}$ term, we take j to be isolated.
3. For each pair of distinct $j, j' \in |\alpha|$ such that we have the term $\langle e_{\alpha(j)}^\perp, e_{\alpha(j')}^\perp\rangle$ (which only happens if we start with the $\langle d', e_{\alpha(j)}^\perp\rangle$ and $\langle d', e_{\alpha(j')}^\perp\rangle$ terms and $e_{\alpha(j)}^\perp$ and $e_{\alpha(j')}^\perp$ are paired together after applying Isserlis' Theorem/Wick's Theorem), we add an edge between j and j'. If $\alpha(j') = \alpha(j)$ then we take this edge to be red and add it to M_{red}. If $\alpha(j') \neq \alpha(j)$ then we take this edge to be purple and add it to M_{purple}.

We now implement this idea. We have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{d}[h_\alpha(d)] &= \sum_{\substack{M_{blue} : M_{blue} \text{ is a matching on } [|\alpha|], \\ \forall (j,j') \in M_{blue}, \alpha(j) = \alpha(j')}} (-1)^{|M_{blue}|} \mathop{\mathbb{E}}_{d}\left[\prod_{j \in |\alpha| : j \text{ is unmatched by } M_{blue}} \langle d, e_{\alpha(j)}\rangle\right] \\
&= \sum_{\substack{M_{blue} : M_{blue} \text{ is a matching on } [|\alpha|], \\ \forall (j,j') \in M_{blue}, \alpha(j) = \alpha(j')}} (-1)^{|M_{blue}|} \mathop{\mathbb{E}}_{d' \sim N(0,I)}\left[\prod_{j \in |\alpha| : j \text{ is unmatched by } M_{blue}} \left(bv_{\alpha(j)} + \langle d', e_{\alpha(j)}^\perp\rangle\right)\right]
\end{aligned}$$

Expanding out $\mathbb{E}_{d' \sim N(0,I)} \left[\prod_{j \in |\alpha| : j \text{ is unmatched by } M_{blue}} \left(b v_{\alpha(j)} + \langle d', e^{\perp}_{\alpha(j)} \rangle \right) \right]$ and applying Isserlis' Theorem/Wick's Theorem, we have that

$$\mathop{\mathbb{E}}_{d}[h_\alpha(d)] = \sum_{M_{blue}, M_{red}, M_{purple}} (-1)^{|M_{blue}|} \prod_{(j,j') \in M_{red}} (1 - v^2_{\alpha(j)}) \prod_{(j,j') \in M_{purple}} (-v_{\alpha(j)} v_{\alpha(j')}) \prod_{j : j \text{ is unmatched by } M = M_{blue} \cup M_{red} \cup M_{purple}} b v_{\alpha(j)}$$

where the sum is taken over all $M_{blue}, M_{red}, M_{purple}$ such that

1. $M = M_{blue} \cup M_{red} \cup M_{purple}$ is a matching on $[|\alpha|]$ and $M_{blue}, M_{red}, M_{purple}$ are disjoint.
2. $\forall (j, j') \in M_{blue}, \alpha(j) = \alpha(j')$.
3. $\forall (j, j') \in M_{red}, \alpha(j) = \alpha(j')$.
4. $\forall (j, j') \in M_{purple}, \alpha(j) \neq \alpha(j')$.

Since whenever we have a blue edge, we could have instead had a red edge and vice versa, for each distinct j, j' such that $\alpha(j') = \alpha(j)$, we can combine terms which have a blue edge between j and j' with terms which have a red edge between j and j'. A blue edge between j and j' has a coefficient of -1 and a red edge between j and j' has a coefficient of $1 - v^2_{\alpha(j)}$, so this effectively gives a purple edge with coefficient $-v^2_{\alpha(j)} = v_{\alpha(j)} v_{\alpha(j')}$. Thus,

$$\begin{aligned} \mathop{\mathbb{E}}_{d}[h_\alpha(d)] &= \sum_{M : M \text{ is a matching on } [|\alpha|]} \left(\prod_{(j,j') \in M} -v_{\alpha(j)} v_{\alpha(j')} \right) \left(\prod_{j : j \text{ is unmatched by } M} b v_{\alpha(j)} \right) \\ &= h_{|\alpha|}(b) v^\alpha \end{aligned}$$

■

D Importance of Scaling

We remark that somewhat surprisingly, the scaling of the problem is important for our arguments. The reason this is somewhat surprising is that for the purpose of determining whether or not a matrix M is PSD, the scaling of the rows and columns of M doesn't matter. More precisely, we have the following proposition.

Proposition D.1. *For any symmetric $N \times N$ matrix M and any $N \times N$ diagonal matrix D such that $\forall i \in [N], D_{ii} \neq 0$, $M \succeq 0$ if and only if $DMD \succeq 0$.*

However, for our techniques, we also use the fact that if x is in the nullspace of M then for the purposes of determining whether M is PSD, we can freely add a non-negative multiple of xx^T to M.

Proposition D.2. *For any symmetric $N \times N$ symmetric matrix M, any vector x such that $Mx = 0$, and any constant c, $M \succeq 0$ if and only if $M + cxx^T \succeq 0$.*

As shown by the following example, the set of matrices that can be obtained using Proposition D.2 depends on the scaling of M.

If $M = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \end{pmatrix}$, $x = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$, and $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix}$ then $DMD = \begin{pmatrix} 1 & 1 & 2\lambda \\ 1 & 2 & 3\lambda \\ 2\lambda & 3\lambda & 5\lambda^2 \end{pmatrix}$ and

$$DMD + cD^{-1}xx^TD^{-1} = \begin{pmatrix} 1+c & 1+c & 2\lambda - \frac{c}{\lambda} \\ 1+c & 2+c & 3\lambda - \frac{c}{\lambda} \\ 2\lambda - \frac{c}{\lambda} & 3\lambda - \frac{c}{\lambda} & 5\lambda^2 + \frac{c}{\lambda^2} \end{pmatrix}$$

Scaling this so that the diagonal entries are 1 gives the matrix

$$\begin{pmatrix} 1 & \frac{\sqrt{1+c}}{\sqrt{2+c}} & \frac{2\lambda - \frac{c}{\lambda}}{\sqrt{(1+c)(5\lambda^2 + \frac{c}{\lambda^2})}} \\ \frac{\sqrt{1+c}}{\sqrt{2+c}} & 1 & \frac{3\lambda - \frac{c}{\lambda}}{\sqrt{(2+c)(5\lambda^2 + \frac{c}{\lambda^2})}} \\ \frac{2\lambda - \frac{c}{\lambda}}{\sqrt{(1+c)(5\lambda^2 + \frac{c}{\lambda^2})}} & \frac{3\lambda - \frac{c}{\lambda}}{\sqrt{(2+c)(5\lambda^2 + \frac{c}{\lambda^2})}} & 1 \end{pmatrix}$$

Note that the entries in the upper left 2×2 block only depend on c and are different for each c while the other off-diagonal entries also depend on λ. Thus, different λ give different sets of matrices.