Machinery for Proving Sum-of-Squares Lower Bounds on Certification Problems

Aaron Potechin* Goutham Rajendran†

November 10, 2020

Abstract

In this paper, we construct general machinery for proving Sum-of-Squares lower bounds on certification problems by generalizing the techniques used by [BHK+16] to prove Sum-of-Squares lower bounds for planted clique. Using this machinery, we prove degree $n^\epsilon$ Sum-of-Squares lower bounds for tensor PCA, the Wishart model of sparse PCA, and a variant of planted clique which we call planted slightly denser subgraph.

*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.
1 Introduction

The Sum-of-Squares (SoS) hierarchy is an optimization technique that harnesses the power of semidefinite programming to solve optimization tasks. For polynomial optimization problems, the SoS hierarchy, first independently investigated by Shor [Sho87], Nesterov [Nes00], Parillo [Par00], Lasserre [Las01] and Grigoriev [Gri01a, Gri01b], offers a sequence of convex relaxations parameterized by an integer called the degree of the SoS hierarchy. As we increase the degree $d$ of the hierarchy, we get progressively stronger convex relaxations, while being solvable in $n^{O(d)}$ time. This has paved the way for the SoS hierarchy to be a powerful tool in algorithm design both in the worst case and the average case settings. Indeed, there has been tremendous success in using the SoS hierarchy to obtain efficient algorithms for combinatorial optimization problems (e.g., [GW95, ARV04, GS11, RRS17]) as well as problems stemming from Statistics and Machine Learning (e.g., [BBH12, BKS15, HSS15, PS17, KS17]).

On the flip side, some problems have remained intractable beyond a certain threshold even by considering higher degrees of the SoS hierarchy [BHK+16, KMOW17, MRX20, GJJ+20]. For example, consider the Planted Clique problem where we have to distinguish a random graph sampled from the Erdős-Rényi model $G(n, \frac{1}{2})$ from a random graph which is obtained by first sampling a graph from $G(n, \frac{1}{2})$ and then planting a clique of size $n^{2\epsilon}$ for a small constant $\epsilon > 0$. It was shown in [BHK+16] that with high probability degree $o(\log n)$ SoS fails to solve this distinguishing problem.

There are many reasons for why studying lower bounds against the SoS hierarchy is important. Firstly, since SoS is a generic proof system that captures a broad class of algorithmic reasoning, SoS lower bounds indicate to the algorithm designer the intrinsic hardness of the problem and that if they want to break the algorithmic barrier, they need to search for algorithms that are not captured by SoS. Secondly, in average case problem settings, standard complexity theoretic assumptions such as $P \neq NP$ have not been shown to give insight into the limits of efficient algorithms. Instead, lower bounds against powerful proof systems such as SoS have served as strong evidence of hardness [Hop18]. Moreover, for a large class of problems, it’s been shown that SoS relaxations are the most efficient among all semidefinite programming relaxations [LRS15]. Thus, understanding the power of the SoS hierarchy on these problems is an important step towards understanding the approximability of these problems.

1.1 Our Contributions

In this paper, we consider the following general category of problems. Given a random input, can we certify that it does not contain a given structure?

Some important examples of this kind of problem are as follows.

1. Planted clique: Can we certify that a random graph does not have a large clique?

2. Tensor PCA: Given an order $k$ tensor $T$ with random Gaussian entries, can we certify that there is no unit vector $x$ such that $\langle T, x \otimes \ldots \otimes x \rangle$ is large?

3. Wishart model of sparse PCA: Given an $m \times d$ matrix $S$ with random Gaussian entries (which corresponds to taking $m$ samples from $\mathcal{N}(0, I_d)$), can we certify that there is no $k$-sparse unit vector $x$ such that $\|Sx\|$ is large?

These kinds of problems, known as certification problems, are closely related to their optimization or estimation variants. A certification algorithm is required to produce a proof/certificate of a bound that
holds for all inputs, as opposed to most inputs. The Sum-of-Squares hierarchy provides such certificates, so analyzing SoS paves the way towards understanding the certification complexity of these problems. We investigate the following question.

For certification problems, what are the best bounds that SoS can certify?

In this work, we build general machinery for proving probabilistic Sum of Squares lower bounds on certification problems. To build our machinery, we generalize the techniques pioneered by [BHK+16] for proving Sum of Squares lower bounds for planted clique. We start with the standard framework for proving probabilistic Sum of Squares lower bounds:

1. Construct candidate pseudo-expectation values \( \tilde{E} \) and the corresponding moment matrix \( \Lambda \) (see Section 2.1).
2. Show that with high probability, \( \Lambda \succeq 0 \).

For planted clique, [BHK+16] constructed \( \tilde{E} \) and the corresponding moment matrix \( \Lambda \) by introducing the pseudo-calibration technique (see Section 2.2). They then showed through a careful and highly technical analysis that with high probability \( \Lambda \succeq 0 \).

In this paper, by generalizing the techniques used in this analysis, we give general conditions which are sufficient to show that a candidate moment matrix \( \Lambda \) is positive-semidefinite (PSD) with high probability. These conditions, which are our main result, are stated informally in Theorem 3.32 and stated formally in Theorem 7.95 and Theorem 7.101.

1.1.1 Pseudo-calibration and Our Machinery

A natural way to prove lower bounds on a certification problem is as follows.

1. Construct a planted distribution of inputs which has the given structure.
2. Show that we cannot distinguish between the random and planted distributions and thus cannot certify that a random input does not have the given structure.

Based on this idea, the pseudo-calibration technique introduced by [BHK+16] constructs candidate pseudo-expectation values \( \tilde{E} \) so that as far as low degree tests are concerned, \( \tilde{E} \) for the random distribution mimics the behavior of the given structure for the planted distribution (for details, see Section 2.2). This gives a candidate moment matrix \( \Lambda \) which we can then analyze with our machinery. Indeed, this is how we prove our SoS lower bounds for tensor PCA, the Wishart model of sparse PCA, and a variant of planted clique which we call planted slightly denser subgraph. That said, our machinery does not require that the candidate moment matrix \( \Lambda \) be obtained via pseudo-calibration.

1.1.2 Results on Tensor PCA, Sparse PCA, and Planted Slightly Denser Subgraph

We now describe the planted distributions we use to show our SoS lower bounds for planted slightly denser subgraph, tensor PCA, and the the Wishart model of sparse PCA. We also state the random distributions for completeness and for contrast. We then state our results.
**Planted slightly denser subgraph**  We use the following distributions.

- Random distribution: Sample $G$ from $G(n, \frac{1}{2})$
- Planted distribution: Let $k$ be an integer and let $p > \frac{1}{2}$. Sample a graph $G'$ from $G(n, \frac{1}{2})$. Choose a random subset $S$ of the vertices, where each vertex is picked independently with probability $\frac{k}{n}$. For all pairs $i, j$ of vertices in $S$, rerandomize the edge $(i, j)$ where the probability of $(i, j)$ being in the graph is now $p$. Set $G$ to be the resulting graph.

In Section 4, we compute the candidate moment matrix $\Lambda$ obtained by using pseudo-calibration on this planted distribution.

**Theorem 1.1.** Let $C_p > 0$. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $k \leq n^{1-\varepsilon}$ and $p = \frac{1}{2} + \frac{\varepsilon}{n}$, then with high probability, the candidate moment matrix $\Lambda$ given by pseudo-calibration for degree $n^{C_\varepsilon}$ sum-of-squares is PSD.

**Corollary 1.2.** Let $C_p > 0$. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $k \leq n^{1-\varepsilon}$ and $p = \frac{1}{2} + \frac{\varepsilon}{n}$, then with high probability, degree $n^{C_\varepsilon}$ sum-of-squares cannot certify that a random graph $G$ from $G(n, \frac{1}{2})$ does not have a subgraph of size $\approx k$ with edge density $\approx p$.

**Tensor PCA**  Let $k \geq 2$ be an integer. We use the following distributions.

- Random distribution: Sample $A$ from $\mathcal{N}(0, I_{[m]})$.
- Planted distribution: Let $\lambda, \Delta > 0$. Sample $u$ from $\left\{ -\frac{1}{\sqrt{\Delta n}}, 0, \frac{1}{\sqrt{\Delta n}} \right\}^n$ where the values are taken with probabilities $\frac{\lambda}{\Delta}, 1 - \Delta, \frac{\lambda}{\Delta}$ respectively. Then sample $B$ from $\mathcal{N}(0, I_{[m]})$. Set $A = B + \lambda u \otimes k$.

In Section 5, we compute the candidate moment matrix $\Lambda$ obtained by using pseudo-calibration on this planted distribution.

**Theorem 1.3.** Let $k \geq 2$ be an integer. There exist constants $C, C_\Delta > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $\lambda \leq n^{\frac{1}{2} - \varepsilon}$ and $\Delta = n^{-C_\Delta \varepsilon}$ then with high probability, the candidate moment matrix $\Lambda$ given by pseudo-calibration for degree $n^{C_\varepsilon}$ sum-of-squares is PSD.

**Corollary 1.4.** Let $k \geq 2$ be an integer. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $\lambda \leq n^{\frac{1}{2} - \varepsilon}$, then with high probability, degree $n^{C_\varepsilon}$ sum-of-squares cannot certify that for a random tensor $A$ from $\mathcal{N}(0, I_{[m]})$, there is no vector $u$ such that $\|u\| \approx 1$ and $\langle A, x \otimes \cdots \otimes x \rangle \approx \lambda$.

**Wishart model of Sparse PCA**  We use the following distributions.

- Random distribution: $v_1, \ldots, v_m$ are sampled from $\mathcal{N}(0, I_d)$ and we take $S$ to be the $m \times d$ matrix with rows $v_1, \ldots, v_m$.
- Planted distribution: Sample $u$ from $\left\{ -\frac{1}{\sqrt{k}}, 0, \frac{1}{\sqrt{k}} \right\}^d$ where the values are taken with probabilities $\frac{k}{2d}, 1 - \frac{k}{d}, \frac{k}{2d}$ respectively. Then sample $v_1, \ldots, v_m$ as follows. For each $i \in [m]$, with probability $\Delta$, sample $v_i$ from $\mathcal{N}(0, I_d + \lambda uu^T)$ and with probability $1 - \Delta$, sample $v_i$ from $\mathcal{N}(0, I_d)$. Finally, take $S$ to be the $m \times d$ matrix with rows $v_1, \ldots, v_m$. 

3
In Section 6, we compute the candidate moment matrix Λ obtained by using pseudo-calibration on this planted distribution.

**Theorem 1.5.** There exists a constant $C > 0$ such that for all sufficiently small constants $ε > 0$, if $m ≤ \frac{d^{1−ε}}{A^2}, m ≤ \frac{k^2−ε}{A^2}$, and there exists a constant $A$ such that $0 < A < \frac{1}{ε}, d^{AA} ≤ k ≤ d^{1−ε},$ and $\frac{\sqrt{A}}{\sqrt{k}} ≤ d^{−ε}$, then with high probability, the candidate moment matrix Λ given by pseudo-calibration for degree $d^{Cε}$ Sum-of-Squares is PSD.

**Corollary 1.6.** There exists a constant $C > 0$ such that for all sufficiently small constants $ε > 0$, if $m ≤ \frac{d^{1−ε}}{A^2}, m ≤ \frac{k^2−ε}{A^2}$, and there exists a constant $A$ such that $0 < A < \frac{1}{ε}, d^{AA} ≤ k ≤ d^{1−ε},$ and $\frac{\sqrt{A}}{\sqrt{k}} ≤ d^{−ε}$, then with high probability, the degree $d^{Cε}$ degree Sum-of-Squares cannot certify that for a random $m \times d$ matrix $S$ with Gaussian entries, there is no vector $u$ such that $u$ has $≈ k$ nonzero entries, $\|u\| ≤ 1$, and $\|Su\|^2 ≈ m + mΔλ.$

**Remark 1.7.** Note that our planted distributions only approximately satisfy constraints such as having a subgraph of size $k$, having a unit vector $u$, and having $u$ be $k$-sparse. While we would like to use planted distributions which satisfy such constraints exactly, these distributions don’t quite satisfy the conditions of our machinery. This same issue appeared in the SoS lower bounds for planted clique [BHK+16]. Resolving this issue is a subtle but important open problem.

## 1.2 Relation to previous work on Planted Clique/Dense Subgraph, Tensor PCA, and Sparse PCA

### 1.2.1 Planted Dense Subgraph

In the planted dense subgraph problem, we are given a random graph $G$ where a dense subgraph of size $k$ has been planted and we are asked to find this planted dense subgraph. This is a natural generalization of the $k$-clique problem [Kar72] and has been subject to a long line of work over the years (e.g. [FS+97, FPK01, Kho06, BCC+10, BCG+12, BKRW17, Man17]). In this work, we consider the following certification variant of planted dense subgraph.

*Given a random graph $G$ sampled from the Erdős-Rényi model $G(n, \frac{1}{k}),$ certify an upper bound on the edge density of the densest subgraph on $k$ vertices.*

In [BHK+16], they show that for $k ≤ n^{\frac{1}{2}−ε}$ for a constant $ε > 0$, the degree $o(\log n)$ Sum-of-Squares cannot distinguish between a fully random graph sampled from $G(n, \frac{1}{k})$ from a random graph which has a planted $k$-clique. This implies that degree $o(\log n)$ SoS cannot certify an edge density better than 1 for the densest $k$-subgraph if $k ≤ n^{\frac{1}{2}−ε}$.

In Corollary 1.2, we show that for $k ≤ n^{\frac{1}{4}−ε}$ for a constant $ε > 0$, degree $n^{\Omega(ε)}$ SoS cannot certify an edge density better than $\frac{1}{2} + \frac{\sqrt{\log(n/k)}}{\sqrt{k}} + o(\frac{1}{\sqrt{k}})$. To the best of our knowledge, this is the first result that proves such a lower bound for SoS degrees as high as $n^{\Omega(ε)}$. When the SoS degree is only $o(\log n)$, our result is not as strong as the work of [BHK+16].

We remark that when we take $k = n^{\frac{1}{4}−ε}$, the true edge density of the densest $k$-subgraph is $\frac{1}{2} + \frac{\sqrt{\log(n/k)}}{\sqrt{k}} + o(\frac{1}{\sqrt{k}}) ≈ \frac{1}{2} + \frac{1}{n^{1/4-ε}}$ as was shown in [GZ19, Corollary 2] whereas, by Corollary 1.2, the SoS optimum is as large as $\frac{1}{2} + \frac{1}{n}$. This highlights a significant additive difference in the optimum value.

4
1.2.2 Tensor PCA

The Tensor Principal Component Analysis problem, originally introduced by [RM14], is a generalization of the PCA problem from machine learning to higher order tensors. Given an order $k$ tensor of the form $\lambda u^\otimes k + B$ where $u \in \mathbb{R}^n$ is a unit vector and $B \in \mathbb{R}^{[n]^k}$ has independent Gaussian entries, we would like to recover $u$. Here, $\lambda$ is known as the signal-to-noise ratio.

This can be equivalently considered to be the problem of optimizing a homogenous degree $k$ polynomial $f(x)$, with random Gaussian coefficients over the unit sphere $\|x\| = 1$. In general, polynomial optimization over the unit sphere is a fundamental primitive with a lot of connections to other areas of optimization (e.g. [FK08, BV09, BH17, BKS14, BKS15, BGG]). Tensor PCA is an average case version of the above problem and has been studied before in the literature [RM14, HSS15, BGL16, HKP+17]. In this work, we consider the certification version of this average case problem.

For an integer $k \geq 2$, given a random tensor $A \in \mathbb{R}^{[n]^k}$ with entries sampled independently from $\mathcal{N}(0,1)$, certify an upper bound on $\langle A, x^\otimes k \rangle$ over unit vectors $x$.

In [BGL16], it was shown that $q \leq n$ levels of SoS certifies an upper bound of $2^{O(k)(\epsilon \cdot \text{polylog}(n))^{k/4}}$ for the Tensor PCA problem. When $q = n^\varepsilon$ for sufficiently small $\varepsilon$, this gives an upper bound of $n^{\frac{k}{4} - O(\varepsilon)}$. Corollary 1.4 shows that this is tight.

In [HKP+17], they state a theorem similar to Corollary 1.4 and observe that it can be proved by applying the techniques used to prove the SoS lower bounds for planted clique. However, they do not give an explicit proof. Also, while they consider the setting where the random distribution has entries from $\{-1,1\}$, we work with $\mathcal{N}(0,1)$. We remark that our machinery can recover their result.

When $k = 2$, the maximum value of $\langle x^\otimes 2, A \rangle$ over the unit sphere $\|x\|^2 = 1$ is precisely the largest eigenvalue of $(A + A^T)/2$ which is $\Theta(\sqrt{n})$ with high probability. For any integer $k \geq 2$, the true maximum of $\langle x^\otimes k, A \rangle$ over $\|x\|^2 = 1$ is $O(\sqrt{n})$ with high probability [TS14]. In contrast, by Corollary 1.4, the optimum value of the degree $n^\varepsilon$ SoS is as large as $n^{\frac{k}{4} - O(\varepsilon)}$. This exhibits an integrality gap of $n^\frac{k}{4} - \frac{k}{2} - O(\varepsilon)$.

1.2.3 Wishart model of Sparse PCA

The Wishart model of Sparse PCA, also known as the Spiked Covariance model, was originally proposed by [JL09]. In this problem, we observe $m$ vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^d$ from the distribution $\mathcal{N}(0, I_d + \lambda uu^T)$ where $u$ is a $k$-sparse unit vector, and we would like to recover $u$. Here, the sparsity of a vector is the number of nonzero entries and $\lambda$ is known as the signal-to-noise ratio.

Sparse PCA is a fundamental problem that has applications in a diverse range of fields (e.g. [WLY12, NYS11, Maj09, TPW14, CK09, AMS11]). It’s known that vanilla PCA does not yield good estimators in high dimensional settings [BAP+05, Pau07, JL09]. A large volume of work has gone into studying Sparse PCA and its variants, both from an algorithmic perspective (e.g. [AW08, Ma13, KNV+15, DM16, WBS+16]) as well as from an inapproximability perspective (e.g. [BR13a, MW15, DKS17, HKP+17, BB19]).

Between this work and prior works, we completely understand the parameter regimes where sparse PCA is easy or conjectured to be hard up to polylogarithmic factors.
- If \( m \gg k^2 \) then the sparse vector can be recovered by diagonal thresholding [JL09, AW08], covariance thresholding [KNV15, DM16], or SoS [dKNS20].

- If \( m \geq d \) and \( m \gg \frac{d}{\lambda} \) or \( m \leq d \) and \( m \gg \frac{d}{\lambda} \) then vanilla PCA can recover the sparse vector (i.e. we do not need to use the fact that the vector is sparse) (e.g. [BR+13b, dKNS20]).

- If \( m \leq d, \frac{d}{\lambda^2} \ll m \ll \frac{d}{\lambda} \), and \( m \gg \frac{k}{\lambda} \) then there is an efficient spectral algorithm to recover the sparse vector (e.g. [dKNS20]).

- If \( m \leq d, \frac{d}{\lambda^2} \ll m \ll \frac{d}{\lambda} \), and \( m \ll \frac{k}{\lambda} \) then there is a simple spectral algorithm which distinguishes the planted distribution from the random distribution but it is information theoretically impossible to recover the sparse vector [dKNS20, Appendix E].

- If \( m \ll \frac{k^2}{\lambda^2} \) and \( m \ll \frac{d}{\lambda^2} \) then it is conjectured to be hard to distinguish between the random and the planted distributions. We discuss the evidence for this below.

For the parameter regime where \( m \ll \frac{k^2}{\lambda^2} \) and \( m \ll \frac{d}{\lambda^2} \), several SoS lower bounds have been shown. The works [KNV15, BR+13b] obtain degree 2 SoS lower bounds. [MW15] obtain degree 4 SoS lower bounds using different techniques. The bounds they obtain are tight up to polylogarithmic factors when \( \lambda \) is a constant. The work [HKP17] considers the related Wigner model of Sparse PCA and they state degree \( d^\varepsilon \) SoS lower bounds, without explicitly proving these bounds. They ask for similar SoS lower bounds in the Wishart model. We almost fully resolve their questions in this work with Corollary 1.6. Our machinery can also recover their results on the Wigner model.

In [dKNS20], they prove that if \( m \leq \frac{d}{\lambda^2} \) and \( m \leq \left( \frac{k^2}{\lambda^2} \right)^{1-\Omega(\varepsilon)} \), then degree \( n^\varepsilon \) polynomials cannot distinguish the random and planted distributions. Corollary 1.6 says that under slightly stronger assumptions, degree \( n^\varepsilon \) Sum-of-Squares cannot distinguish the random and planted distributions, so we confirm that SoS is no more powerful than low degree polynomials in this setting.

### 1.3 Comparison to Other Sum-of-Squares Lower Bounds on Certification Problems

In a seminal work, [BHK+16] proved sum of squares lower bounds for the planted clique problem. Our machinery greatly generalizes the techniques of this paper. That said, for a technical reason, our machinery actually doesn’t quite handle planted clique (See Remark 3.35).

Similar to this paper, [HKP+17] also observed that the techniques used in [BHK+16] can be used to give Sum-of-Squares lower bounds for \( \pm 1 \) variants of tensor PCA and sparse PCA, though this is not made explicit. In this paper, we use our machinery to make these lower bounds explicit. We also handle the Wishart model of sparse PCA, which is significantly harder to prove lower bounds for.

[KMOW17] proved that for random constraint satisfaction problems (CSPs) where the predicate has a balanced pairwise independent distribution of solutions, with high probability, degree \( \Omega(n) \) SoS is required to certify that these CSPs do not have a solution. While the pseudo-expectation values used by [KMOW17] can also be derived using pseudo-calibration, the analysis for showing that the moment matrix is PSD is very different. It is an interesting question whether or not it is possible to unify these analyses.

[MRX20] showed that it’s possible to lift degree 2 SoS solutions to degree 4 SoS solutions under suitable conditions, and used it to obtain degree 4 SoS lower bounds for average case \( d \)-regular Max-Cut and the Sherrington Kirkpatrick problem. Their construction is inspired by pseudo-calibration and their analysis also goes via graph matrices.
Recently, [GJJ+20] proved degree $n^c$ SoS lower bounds for the Sherrington-Kirkpatrick problem via an intermediate problem known as Planted Affine Planes. Their construction and analysis also goes via pseudo-calibration and graph matrices, but since the constructed moment matrix had a nontrivial nullspace, they had to use different techniques to handle them. However, once this nullspace is taken into account, the moment matrix is dominated by its expected value, so using our machinery would be overkill.

[Kun20] recently proposed a technique to lift degree 2 SoS lower bounds to higher levels and applied it to construct degree 6 lower bounds for the Sherrington-Kirkpatrick problem. Interestingly, their construction does not go via pseudo-calibration.

1.4 Related Algorithmic Techniques

**Low degree polynomials** Consider a problem where the input is sampled from one of two distributions and we would like to identify which distribution it was sampled from. Usually, one distribution is the random distribution while the other is a planted distribution that contains a given structure not present in the random distribution. In this setting, a closely related method is to use low degree polynomials to try and distinguish the two distributions. More precisely, if there is a low degree polynomial such that its expected value on the random distribution is very different than its expected value on the planted distribution, this distinguishes the two distributions. Recently, this method has been shown to be an excellent heuristic, as it recovers the conjectured hardness thresholds for several problems and is considerably easier to analyze [HKP+17, Hop18, KWB19].

That said, it is an open question whether low degree polynomials generally have the same power as the SoS hierarchy or if there are situations where the SoS hierarchy is more powerful. In this paper, we confirm that for tensor PCA and the Wishart model of sparse PCA with slightly adjusted planted distributions, the SoS hierarchy is no more powerful than low-degree polynomials.

**The Statistical Query Model** The statistical query model is another popular restricted class of algorithms introduced by [Kea98]. In this model, for an underlying distribution, we can access it by querying expected value of functions of the distribution. Concretely, for a distribution $D$ on $\mathbb{R}^n$, we have access to it via an oracle that given as query a function $f : \mathbb{R}^n \to [-1,1]$ returns $E_{x \sim D} f(x)$ up to some additive adversarial error. SQ algorithms capture a broad class of algorithms in statistics and machine learning and has also been used to study information-computation tradeoffs. There has also been significant work trying to understand the limits of SQ algorithms (e.g. [FGR+17, FPV18, DKS17]). The recent work [BBH+20] showed that low degree polynomials and statistical query algorithms have equivalent power under mild conditions. It’s an interesting open question whether or not SQ algorithms have the same power as Sum-of-Squares algorithms.

1.5 Organization of the paper

The remainder of this paper is organized as follows. In Section 2, we give some preliminaries. In particular, we describe the Sum-of-Squares hierarchy and present a brief overview of the machinery and some proof techniques that we use. In Section 3, we present the informal statement of the main theorem. In Section 4, Section 5 and Section 6, we qualitatively verify the conditions of the machinery for planted slightly denser subgraph, tensor PCA, and sparse PCA respectively. While these sections only verify the qualitative conditions, the results in these sections are precise and will be reused in Section 11, Section 12 and Section 13 to fully verify the conditions of the machinery. In Section 7, we introduce all the formal definitions and
state the main theorem in full generality. In Section 8, we prove the main theorem while abstracting out the choice of several functions. In Section 9, we choose these functions so that that they satisfy the conditions needed for our main theorem. In Section 10, we give tools for verifying a technical condition of our machinery which is related to truncation error. Finally, in Section 11, Section 12 and Section 13, we verify all the conditions necessary to prove Theorem 1.1, Theorem 1.3 and Theorem 1.5 respectively.

2 Preliminaries and proof overview

2.1 The Sum of Squares Hierarchy

We will first introduce the notion of a pseudoexpectation operator for a set of polynomial constraints and then describe the Sum-of-Squares relaxation for a polynomial optimization problem.

For an integer $d$, let $\mathbb{R}^{\leq d}[x_1, \ldots, x_n]$ be the set of polynomials on $x_1, \ldots, x_n$ of degree at most $d$. We denote the degree of a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ by $\deg(f)$.

**Definition 2.1** (Pseudoexpectation operator). Given polynomial constraints $g_1(x) = 0, \ldots, g_m(x) = 0$ on variables $x_1, \ldots, x_n$ such that $\deg(g_i) \leq D$ for an integer $D \geq 0$. For an even integer $d \geq D$, a degree-$d$ pseudoexpectation operator $\mathbb{E}$ satisfying these constraints is an operator $\mathbb{E} : \mathbb{R}^{\leq d}[x_1, \ldots, x_n] \to \mathbb{R}$ satisfying:

1. $\mathbb{E}[1] = 1$,
2. $\mathbb{E}$ is an $\mathbb{R}$-linear operator; i.e., $\mathbb{E}[f + cg] = \mathbb{E}[f] + c\mathbb{E}[g]$ for every $f, g \in \mathbb{R}^{\leq d}[x_1, \ldots, x_n], c \in \mathbb{R}$,
3. $\mathbb{E}[g_i \cdot f] = 0$ for every $i = 1, \ldots, m$ and $f \in \mathbb{R}^{\leq d}[x_1, \ldots, x_n]$ with $\deg(f \cdot g_i) \leq d$.
4. $\mathbb{E}[f^2] \geq 0$ for every $f \in \mathbb{R}^{\leq d}[x_1, \ldots, x_n]$ with $\deg(f^2) \leq d$.

The notion of $\mathbb{E}$ generalizes the standard expectation operator. The idea is that optimizing over this larger space of pseudoexpectation operators can be formulated as a semidefinite programming problem and hence, will serve as a relaxation of our program that can be solved efficiently.

Formally, consider an optimization task on $n$ variables $x_1, \ldots, x_n \in \mathbb{R}$ formulated as maximizing a polynomial $f(x)$ subject to polynomial constraints $g_1(x) = 0, \ldots, g_m(x) = 0$. Suppose all the polynomials $f, g_1, \ldots, g_m$ have degree at most $D$. Then, for an even integer $d \geq D$, the degree $d$ Sum-of-Squares relaxation of this program is as follows: Over all pseudoexpectation operators $\mathbb{E}$ satisfying $\mathbb{E}[g_1(x)] = 0, \ldots, \mathbb{E}[g_m(x)] = 0$, output the maximum value of $\mathbb{E}[f(x)]$.

To prove an SoS lower bound, we need to exhibit an operator $\mathbb{E}$ that satisfies these constraints with optimum value of $\mathbb{E}[f(x)]$ being far away from the true optimum.

In most cases, when constructing $\mathbb{E}$, the condition Item 4 in Definition 2.1 is the most technically challenging condition to satisfy. It can be equivalently stated as a positive semidefiniteness condition of an associated matrix called the moment matrix.

To define the moment matrix, we need to set up some more notation.

For an integer $d \geq 0$, let $\mathcal{I}_d$ denote the set of all tuples $(t_1, \ldots, t_n)$ such that $t_i \geq 0$ for all $i$ and $\sum t_i \leq d$. For $I = (t_1, \ldots, t_n) \in \mathcal{I}_d$, denote by $x^I := x_1^{t_1} x_2^{t_2} \cdots x_n^{t_n}$.
**Definition 2.2 (Moment Matrix of \( \mathbb{E} \)).** For a degree \( 2d \) pseudoexpectation operator \( \mathbb{E} \) on variables \( x_1, \ldots, x_n \), define the associated moment matrix \( \Lambda \) to be a matrix with rows and columns indexed by \( \mathbb{T}_d \) such that the entry corresponding to row \( I \) and column \( J \) is

\[
\Lambda[I, J] := \mathbb{E} \left[ x^I \cdot x^J \right].
\]

It is easy to verify that Item 4 in Definition 2.1 equivalent to \( \Lambda \succeq 0 \).

The machinery in this paper provides general conditions under which we can show that with high probability, \( \Lambda \succeq 0 \).

### 2.2 Pseudo-calibration

Pseudo-calibration is a heuristic developed in [BHK+16] to construct a candidate pseudoexpectation operator \( \mathbb{E} \) and a corresponding moment matrix \( \Lambda \) for random vs planted problems. This will be the starting point for all our applications. Here, we will describe the heuristic and show an example of how to use it.

Let \( \nu \) denote the random distribution and \( \mu \) denote the planted distribution. Let \( \nu \) denote the input and \( x \) denote the variables for our SoS relaxation. The idea is that, for an input \( \nu \) sampled from \( \nu \) and any polynomial \( f(x) \) of degree at most the SoS degree, pseudo-calibration proposes that for any low-degree test \( g(\nu) \), the correlation of \( \mathbb{E}[f] \) should match in the planted and random distributions. That is,

\[
\mathbb{E}_{\nu \sim \nu} \left[ \mathbb{E}[f(x)] g(\nu) \right] = \mathbb{E}_{(x,\nu) \sim \mu} \left[ f(x) g(\nu) \right]
\]

Let \( \mathcal{F} \) denote the Fourier basis of polynomials for the input \( \nu \). By choosing different basis functions from \( \mathcal{F} \) as choices for \( g \) such that the degree is at most \( n^d \) (hence the term low-degree test), we get all lower order Fourier coefficients for \( \mathbb{E}[f] \) when considered as a function of \( \nu \). Furthermore, the higher order coefficients are set to be 0 so that the candidate pseudoexpectation operator can be written as

\[
\mathbb{E} f(x) = \sum_{g \in \mathcal{F}} \mathbb{E}_{\nu \sim \nu} \left[ \mathbb{E}[f(x)] g(\nu) \right] g(\nu) = \sum_{g \in \mathcal{F}} \mathbb{E}_{(x,\nu) \sim \mu} \left[ [f(x)] g(\nu) \right] g(\nu)
\]

The coefficients \( \mathbb{E}_{(x,\nu) \sim \mu} \left[ [f(x)] g(\nu) \right] g(\nu) \) can be explicitly computed in many settings, which therefore gives an explicit pseudoexpectation operator \( \mathbb{E} \).

An advantage of pseudo-calibration is that this construction automatically satisfies some nice properties that the pseudoexpectation \( \mathbb{E} \) should satisfy. It’s linear in \( \nu \) by construction. For all polynomial equalities of the form \( f(x) = 0 \) that is satisfied in the planted distribution, it’s true that \( \mathbb{E}[f(x)] = 0 \). For other polynomial equalities of the form \( f(x, \nu) = 0 \) that are satisfied in the planted distribution, the equality \( \mathbb{E}[f(x, \nu)] = 0 \) is approximately satisfied. In most cases, \( \mathbb{E} \) can be mildly adjusted to satisfy these exactly.

The condition \( \mathbb{E}[1] = 1 \) is not automatically satisfied but in most applications, we usually require that \( \mathbb{E}[1] = 1 \pm o(1) \). Indeed, this has been the case for all known successful applications of pseudo-calibration. Once we have this, we simply set our final pseudoexpectation operator to be \( \mathbb{E}' \) defined as \( \mathbb{E}'[f(x)] = \mathbb{E}[f(x)] / \mathbb{E}[1] \).

We remark that the condition \( \mathbb{E}[1] = 1 \pm o(1) \) has been quite successful in predicting the right thresholds between approximability and inapproximability [HKP+17, Hop18, KWB19].
**Example: Planted Clique**  As an warmup, we review the pseudo-calibration calculation for planted clique. Here, the random distribution $\nu$ is $G(n, \frac{1}{2})$.

The planted distribution $\mu$ is as follows. For a given integer $k$, first sample $G'$ from $G(n, \frac{1}{2})$, then choose a random subset $S$ of the vertices where each vertex is picked independently with probability $\frac{k}{n}$. For all pairs $i, j$ of distinct vertices in $S$, add the edge $(i, j)$ to the graph if not already present. Set $G$ to be the resulting graph.

The input is given by $G \in \{-1, 1\}^{[n] \times [n]}$ where $G_{i,j}$ is 1 if the edge $(i, j)$ is present and $-1$ otherwise. Let $x_1, \ldots, x_n$ be the boolean variables for our SoS program such that $x_i$ indicates if $i$ is in the clique.

**Definition 2.3.** Given a set of vertices $V \subseteq [n]$, define $x_V = \prod_{v \in V} x_v$.

**Definition 2.4.** Given a set of possible edges $E \subseteq \binom{[n]}{2}$, define $\chi_E = (-1)^{|E \setminus E(G)|} = \prod_{(i,j) \in E} G_{i,j}$.

Pseudo-calibration says that for all small $V$ and $E$,

$$\mathbb{E}_{G \sim \nu} [\tilde{E}[x_V] \chi_E] = \mathbb{E}_\mu [x_V \chi_E]$$

Using standard Fourier analysis, this implies that if we take

$$c_E = \mathbb{E}_\mu [x_V \chi_E] = \left( \frac{k}{n} \right)^{|V \cup V(E)|}$$

where $V(E)$ is the set of the endpoints of the edges in $E$, then for all small $V$,

$$\tilde{E}[x_V] = \sum_{E: E \text{ is small}} c_E \chi_E = \sum_{E: E \text{ is small}} \left( \frac{k}{n} \right)^{|V \cup V(E)|} \chi_E$$

Since the values of $\tilde{E}[x_V]$ are known, by multi-linearity, this can be naturally extended to obtain values $\tilde{E}[f(x)]$ for any polynomial $f$ of degree at most the SoS degree.

### 2.3 Proof overview and techniques

In this section, we describe some ideas behind our machinery.

As explained above, pseudo-calibration will give us a candidate pseudoexpectation operator $\tilde{E}$ and we can consider a corresponding moment matrix $\Lambda$. For example, in the case of planted clique, $\Lambda$ has rows and columns indexed by sub-tuples of $[n]$ of size at most $d$, such that

$$\Lambda[I, J] = \sum_{E: E \text{ is small}} \left( \frac{k}{n} \right)^{|I \cup J \cup V(E)|} \chi_E$$

for all sub-tuples $I, J$ of $[n]$ of size at most $d$. We would like to prove that $\Lambda \succeq 0$ with high probability.

The machinery gives a set of conditions under which we can show that $\Lambda \succeq 0$ with high probability. Our first step will be decompose $\Lambda$ into a linear combination of graph matrices.
**Graph matrices** Graph matrices were originally introduced by [BHK+16, MP16] and later generalized in [AMP20]. We use the generalized graph matrices in our analysis.

Each graph matrix is a matrix valued function of the input, that can be identified by a graph with labeled edges that we call a shape. Informally, graph matrices will form a basis for all matrix valued functions of the input that have a certain symmetry. In particular, $\Lambda$ is one such matrix valued function and can thus be decomposed into graph matrices. For a shape $\alpha$, the graph matrix associated to $\alpha$ is denoted by $M_\alpha$.

Graph matrices have several useful properties. Firstly, $\|M_\alpha\|$ can be bounded with high probability in terms of simple combinatorial properties of the shape $\alpha$. Secondly, when we multiply two graph matrices $M_\alpha, M_\beta$ corresponding to shapes $\alpha, \beta$, it approximately equals the graph matrix $M_{\alpha \circ \beta}$ where the shape $\alpha \circ \beta$, called the composition of the two shapes $\alpha \circ \beta$, is easy to describe combinatorially.

These makes graph matrices a convenient tool to analyze the moment matrix. In our setting, the moment matrix decomposes as $\Lambda = \sum \lambda_\alpha M_\alpha$ where the sum is over all shapes $\alpha$ and $\lambda_\alpha \in \mathbb{R}$ are the coefficients that arise from pseudo-calibration.

**Decomposing Shapes** For graph matrices $\alpha, \beta$, $M_\alpha M_\beta \approx M_{\alpha \circ \beta}$ where we define the composition of two shapes $\alpha \circ \beta$ to be a larger shape that is obtained by concatenating the shapes $\alpha, \beta$. This equality is only approximate and handling it precisely is a significant source of difficulty in our analysis. Shape composition is also associative, hence we can define composition of three shapes.

A crucial idea for our machinery is that for any shape $\alpha$, there exists a canonical and unique decomposition of $\alpha$ as $\sigma \circ \tau \circ \sigma'^T$ satisfying some nice structural properties, for shapes $\sigma, \tau$ and $\sigma'^T$. Here, $\sigma, \tau, \sigma'^T$ are called the left part, the middle part and the right part of $\alpha$ respectively. Using this, our moment matrix can be written as

$$\Lambda = \sum \lambda_\alpha M_\alpha = \sum_{\sigma, \tau, \sigma'} \lambda_{\sigma \circ \tau \circ \sigma'^T} M_{\sigma \circ \tau \circ \sigma'^T}$$

**Giving a PSD factorization** We first consider the terms $\sum_{\sigma, \sigma'} \lambda_{\sigma \circ \tau \circ \sigma'^T} M_{\sigma \circ \tau \circ \sigma'^T} \approx \sum_{\sigma, \sigma'} \lambda_{\sigma \circ \tau \circ \sigma'^T} M_\sigma M_{\sigma'^T}$ where $\tau$ corresponds to an identity matrix and can be ignored.

If there existed real numbers $v_\sigma$ for all left shapes $\sigma$ such that $\lambda_{\sigma \circ \tau \circ \sigma'^T} = v_\sigma v_{\sigma'^T}$, then we would have

$$\sum_{\sigma, \sigma'} \lambda_{\sigma \circ \tau \circ \sigma'^T} M_\sigma M_{\sigma'^T} = \sum_{\sigma, \sigma'} v_\sigma v_{\sigma'^T} M_\sigma M_{\sigma'^T} = (\sum_{\sigma} v_\sigma M_\sigma)(\sum_{\sigma'} v_{\sigma'} M_{\sigma'})^T \succeq 0$$

which shows that the contribution from these terms is positive semidefinite. Note that the existence of $v_\sigma$ can be relaxed as follows. Let $H$ be the matrix with rows and columns indexed by left shapes $\sigma$ such that $H(\sigma, \sigma') = \lambda_{\sigma \circ \tau \circ \sigma'^T}$. If $H$ is positive semidefinite then the contribution from these terms will also be positive semidefinite. In fact, this will be the first condition of our main theorem, Theorem 7.101.

**Handling terms with a non-trivial middle part** Unfortunately, we also have terms $\lambda_{\sigma \circ \tau \circ \sigma'^T} M_{\sigma \circ \tau \circ \sigma'^T}$ where $\tau$ is non-trivial. Our strategy will be to charge these terms to other terms.

A starting point for our argument is the following inequality. For left shape $\sigma$, middle shape $\tau$ and right shape $\sigma'^T$, and real numbers $a, b$,

$$(a M_\sigma - b M_{\sigma'}) (a M_\sigma - b M_{\sigma'})^T \succeq 0$$
which rearranges to

$$ab(M_\sigma M_\tau M_{\sigma^T} + (M_\sigma M_\tau M_{\sigma^T})^T) \preceq a^2 M_\sigma M_{\sigma^T} + b^2 M_{\sigma^T} M_\tau M_{\sigma^T} M_{\sigma^T} \preceq a^2 M_\sigma M_{\sigma^T} + b^2 \|M_\tau\|^2 M_{\sigma^T} M_{\sigma^T}$$

If it is true that

$$\lambda^2_{\sigma \sigma \tau \sigma^T} \|M_\tau\|^2 \leq \lambda_{\sigma' \sigma^T} \lambda_{\sigma' \sigma^T},$$

then we can choose $a, b$ such that $a^2 \leq \lambda_{\sigma' \sigma^T}, b^2 \|M_\tau\|^2 \leq \lambda_{\sigma' \sigma^T}$ and $ab = \lambda_{\sigma \sigma \tau \sigma^T}$. This will approximately imply

$$\lambda_{\sigma \sigma \tau \sigma^T} (M_{\sigma \sigma \tau \sigma^T} + M_{\sigma \sigma \tau \sigma^T}^T) \preceq \lambda_{\sigma' \sigma^T} M_{\sigma' \sigma^T} + \lambda_{\sigma' \sigma^T} M_{\sigma' \sigma^T}$$

which will give us a way to charge terms with a nontrivial middle part against terms with a trivial middle part.

While we could try to apply this inequality term by term, it is not strong enough to give us our results. Instead, we generalize this inequality. This will lead us to the second condition of our main theorem, Theorem 7.101.

**Handling intersection terms** There’s one important technicality in the above heuristic calculations. Whenever we multiply two graph matrices $M_\alpha, M_\beta$, it is only approximately equal to $M_{\alpha \beta}$. All the other error terms have to be carefully handled in our analysis. We call these terms intersection terms.

These intersection terms themselves turn out to be graph matrices and our strategy is to now recursively decompose them into $\sigma_2 \circ \tau_2 \circ \sigma_2^T$ and apply the previous ideas. A similar approach was undertaken in [BHK+16] but this work generalizes it significantly. To do this methodically, we employ several ideas such as the notion of intersection patterns and the generalized intersection tradeoff lemma (see Section 8). Properly handling the intersection terms is one of the most technically intensive parts of our work.

This analysis leads us to condition 3 of Theorem 7.101.

**Applying the machinery** To apply the machinery to our problems of interest, we verify the spectral conditions that our coefficients should satisfy so that we can use Theorem 7.101. The Planted slightly denser subgraph application is straightforward and will serve as a good warmup to understand our machinery. In the applications to Tensor PCA and Sparse PCA, the shapes corresponding to the graph matrices with nonzero coefficients have nice structural properties that will be crucial for our analysis. We exploit this structure and use novel charging arguments to verify the conditions of our machinery.

# 3 Informal Description of our Machinery

In this section, we informally describe our machinery for proving sum of squares lower bounds on planted problems. Our goal for this section is to qualitatively state the conditions under which we can show that the moment matrix $\Lambda$ is PSD with high probability (see Theorem 3.32). For simplicity, in this section we restrict ourselves to the setting where the input is $\{ -1, 1 \}$ (i.e. a random graph on $n$ vertices). We also defer the proofs of several important facts until Section 7. In Section 7, we give the general definitions, fill in the missing proofs, and give the full, quantitative statement of our main result (see Theorem 7.101).
3.1 Fourier analysis for matrix-valued functions: ribbons, shapes, and graph matrices

For our machinery, we need the definitions of ribbons, shapes, and graph matrices from [AMP20].

3.1.1 Ribbons

Ribbons lift the usual Fourier basis for functions \( \{ f : \{ \pm 1 \}^\binom{n}{2} \to \mathbb{R} \} \) to matrix-valued functions.

**Definition 3.1** (Simplified ribbons – see Definition 7.22). Let \( n \in \mathbb{N} \). A ribbon \( R \) is a tuple \((E_R, A_R, B_R)\) where \( E_R \subseteq \binom{n}{2} \) and \( A_R, B_R \) are tuples of elements in \([n]\). \( R \) thus specifies:

1. A Fourier character \( \chi_{E_R} \).
2. Row and column indices \( A_R \) and \( B_R \).

We think of \( R \) as a graph with vertices \( V(R) = \{ \text{endpoints of } (i, j) \in E_R \} \cup A_R \cup B_R \) and edges \( E(R) = E_R \), where \( A_R, B_R \) are distinguished tuples of vertices.

**Definition 3.2** (Matrix-valued function for a ribbon \( R \)). Given a ribbon \( R \), we define the matrix valued function \( M_R : \{ \pm 1 \}^\binom{n}{2} \to \mathbb{R}^{n! \times n!} \) to have entries \( M(A_R, B_R) = \chi_{E_R} \) and \( M(A', B') = 0 \) whenever \( A' \neq A_R \) or \( B' \neq B_R \).

The following proposition captures the main property of the matrix-valued functions \( M_R \) – they are an orthonormal basis. We leave the proof to the reader.

**Proposition 3.3.** The matrix-valued functions \( M_R \) form an orthonormal basis for the vector space of matrix valued functions with respect to the inner product

\[
\langle M, M' \rangle = \mathbb{E}_{G \sim \{ \pm 1 \}^\binom{n}{2}} \left[ \operatorname{Tr} \left( M(G)(M'(G))^\top \right) \right].
\]

3.1.2 Shapes and Graph Matrices

As described above, ribbons are an orthonormal basis for matrix-valued functions. However, we will need an orthogonal basis for the subset of those functions which are symmetric with respect to the action of \( S_n \). For this, we use graph matrices, which are described by shapes. The idea is that each ribbon \( R \) has a shape \( \alpha \) which is obtained by replacing the vertices of \( R \) with unspecified indices. Up to scaling, the graph matrix \( M_\alpha \) is the average of \( M_{\pi(R)} \) over all permutations \( \pi \in S_n \).

**Definition 3.4** (Simplified shapes – see Definition 7.34). Informally, a shape \( \alpha \) is just a ribbon \( R \) where the vertices are specified by variables rather than having specific values in \([n]\). More precisely, a shape \( \alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha) \) is a graph on vertices \( V(\alpha) \), with

1. Edges \( E(\alpha) \subseteq \binom{V(\alpha)}{2} \)
2. Distinguished tuples of vertices \( U_\alpha = (u_1, u_2, \ldots) \) and \( V_\alpha = (v_1, v_2, \ldots) \), where \( u_i, v_i \in V(\alpha) \).
(Note that $V(\alpha)$ and $V_\alpha$ are not the same object!)

**Definition 3.5 (Shape transposes).** Given a shape $\alpha$, we define $\alpha^\top$ to be the shape $\alpha$ with $U_\alpha$ and $V_\alpha$ swapped i.e. $U_{\alpha^\top} = V_\alpha$ and $V_{\alpha^\top} = U_\alpha$. Note that $M_{\alpha^\top} = M_\alpha^\top$, where $M_\alpha^\top$ is the usual transpose of the matrix-valued function $M_\alpha$.

**Definition 3.6 (Graph matrices).** Let $\alpha$ be a shape. The graph matrix $M_\alpha : \{\pm 1\}^{(|V(\alpha)|)} \rightarrow \mathbb{R}^{(|U(\alpha)| \times (|U(\alpha)| - |U(\alpha)|) \times (|V(\alpha)| - |U(\alpha)|) \times (|V(\alpha)| - |V(\alpha)|)}}$ is defined to be the matrix-valued function with $A, B$-th entry $M_\alpha(A, B) = \sum_{R \text{ s.t. } A_R = A, B_R = B} \chi_{E_R}$ where the sum is over ribbons $R$ which can be obtained by assigning each vertex in $V(\alpha)$ a label from $[n]$.

In other words, $M_\alpha = \sum_R M_R$ where the sum is over ribbons $R$ which can be obtained by assigning each vertex in $V(\alpha)$ a label from $[n]$.

For examples of graph matrices, see [AMP20].

**Remark 3.7.** As noted in [AMP20], we index graph matrices by tuples rather than sets so that they are symmetric (as a function of the input) under permutations of $[n]$.

### 3.2 Factoring Graph Matrices and Decomposing Shapes into Left, Middle, and Right Parts

A crucial idea in our analysis is the idea from [BHK + 16] of decomposing each shape $\alpha$ into left, middle, and right parts. This will allow us to give an approximate factorization of each graph matrix $M_\alpha$.

#### 3.2.1 Leftmost and Rightmost Minimum Vertex Separators and Decomposition of Shapes into Left, Middle, and Right Parts

For each shape $\alpha$ we will identify three other shapes, which we denote by $\sigma, \tau, \sigma^T$ and call (for reasons we will see soon) the left, middle, and right parts of $\alpha$, respectively. The idea is that $M_\alpha \approx M_\sigma M_\tau M_{\sigma^T}$. We obtain $\sigma, \tau$, and $\sigma^T$ by splitting the shape $\alpha$ along the leftmost and rightmost minimum vertex separators.

**Definition 3.8 (Vertex Separators).** We say that a set of vertices $S$ is a vertex separator of $\alpha$ if every path from $U_\alpha$ to $V_\alpha$ in $\alpha$ (including paths of length 0) intersects $S$. Note that for any vertex separator $S$, $U_\alpha \cap V_\alpha \subseteq S$.

**Definition 3.9 (Minimum Vertex Separators).** We say that $S$ is a minimum vertex separator of $\alpha$ if $S$ is a vertex separator of $\alpha$ and for any other vertex separator $S'$ of $\alpha$, $|S| \leq |S'|$.

**Definition 3.10 (Leftmost and Rightmost Minimum Vertex Separators).**

1. We say that $S$ is the leftmost minimum vertex separator of $\alpha$ if $S$ is a minimum vertex separator of $\alpha$ and for every other minimum vertex separator $S'$ of $\alpha$, every path from $U_\alpha$ to $S'$ intersects $S$.

2. We say that $T$ is the rightmost minimum vertex separator of $\alpha$ if $T$ is a minimum vertex separator of $\alpha$ and for every other minimum vertex separator $S'$ of $\alpha$, every path from $S'$ to $V_\alpha$ intersects $T$. 

14
It is not immediately obvious that leftmost and rightmost minimum vertex separators are well-defined. For the simplified setting we are considering here, this was shown by [BHK+16]. We give a more general proof in Appendix A.

We now describe how to split \( \alpha \) into left, middle, and right parts \( \sigma, \tau, \) and \( \sigma^{JT} \).

**Definition 3.11 (Decomposition Into Left, Middle, and Right Parts).** Let \( \alpha \) be a shape and let \( S \) and \( T \) be the leftmost and rightmost minimum vertex separators of \( \alpha \). Given orderings \( O_S \) and \( O_T \) for \( S \) and \( T \), we decompose \( \alpha \) into left, middle, and right parts \( \sigma, \tau, \) and \( \sigma^{JT} \) as follows.

1. The left part \( \sigma \) of \( \alpha \) is the part of \( \alpha \) reachable from \( U_\alpha \) without passing through \( S \). It includes \( S \) but excludes all edges which are entirely within \( S \). More formally,
   - (a) \( V(\sigma) = \{ u \in V(\alpha) : \text{there is a path } P \text{ from } U_\alpha \text{ to } u \text{ in } \alpha \text{ such that } (V(P) \setminus \{ u \}) \cap S = \emptyset \} \)
   - (b) \( U_\sigma = U_\alpha \) and \( V_\sigma = S \) with the ordering \( O_S \)
   - (c) \( E(\sigma) = \{ \{ u, v \} \in E(\alpha) : u, v \in V(\sigma), u \notin S \text{ or } v \notin S \} \)

2. The right part \( \sigma^{JT} \) of \( \alpha \) is the part of \( \alpha \) reachable from \( V_\alpha \) without intersecting \( T \) more than once. It includes \( T \) but excludes all edges which are entirely within \( T \). More formally,
   - (a) \( V(\sigma^{JT}) = \{ u \in V(\alpha) : \text{there is a path } P \text{ from } V_\alpha \text{ to } u \text{ in } \alpha \text{ such that } (V(P) \setminus \{ u \}) \cap T = \emptyset \} \)
   - (b) \( U_{\sigma^{JT}} = T \) with the ordering \( O_T \) and \( V_{\sigma^{JT}} = V_\alpha \)
   - (c) \( E(\sigma^{JT}) = \{ \{ u, v \} \in E(\alpha) : u, v \in V(\sigma^{JT}), u \notin T \text{ or } v \notin T \} \)

3. The middle part \( \tau \) of \( \alpha \) is, informally, the part of \( \alpha \) between \( S \) and \( T \) (including \( S \) and \( T \) and all edges which are entirely within \( S \) or within \( T \)). More formally, let \( U_\tau = S \) with the ordering \( O_S \), let \( V_\tau = T \) with the ordering \( O_T \), and let \( E(\tau) = E(\alpha) \setminus (E(\sigma) \cup E(\sigma')) \) be all of the edges of \( E(\alpha) \) which do not appear in \( E(\sigma') \) or \( E(\sigma') \). Then \( V(\tau) \) is all of the vertices incident to edges in \( E(\tau) \) together with \( S, T \).

**Remark 3.12.** Note that the decomposition into left, middle, and right parts depends on the ordering for the vertices in \( S \) and \( T \). As we will discuss later (see Section 7.8), we will use all possible orderings simultaneously and then scale things by an appropriate constant.

Because of the minimality and leftmost/rightmost-ness of the vertex separators \( S, T \) used to define \( \sigma, \tau, \sigma' \), the shapes \( \sigma, \tau, \sigma' \) have some special combinatorial structure, which we capture in the following proposition. We defer the proof until Section 7 where we state a generalized version.

**Proposition 3.13.** \( \sigma, \tau, \) and \( \sigma^{JT} \) have the following properties:

1. \( V_\sigma = S \) is the unique minimum vertex separator of \( \sigma \).
2. \( S \) and \( T \) are the leftmost and rightmost minimum vertex separators of \( \tau \).
3. \( T = U_{\sigma^{JT}} \) is the unique minimum vertex separator of \( \sigma^{JT} \).

Based on this, we define sets of shapes which can appear as left, middle, or right parts.

**Definition 3.14 (Left, Middle, and Right Parts).** Let \( \alpha \) be a shape.

1. We say that \( \alpha \) is a left part if \( V_\alpha \) is the unique minimum vertex separator of \( \alpha \) and \( E(\alpha) \) has no edges which are entirely contained in \( V_\alpha \).
2. We say that \( \alpha \) is a proper middle part if \( U_\alpha \) is the leftmost minimum vertex separator of \( \alpha \) and \( V_\alpha \) is the rightmost minimum vertex separator of \( \alpha \).

3. We say that \( \alpha \) is a right part if \( U_\alpha \) is the unique minimum vertex separator of \( \alpha \) and \( E(\alpha) \) has no edges which are entirely contained in \( U_\alpha \).

**Remark 3.15.** For technical reasons, later on we will need to consider improper middle parts \( \tau \) where \( U_\tau \) and \( V_\tau \) are not the leftmost and rightmost minimum vertex separators of \( \tau \), which is why we make this distinction here.

The following proposition is also straightforward from the definitions.

**Proposition 3.16.** A shape \( \sigma \) is a left part if and only if \( \sigma^T \) is a right part.

### 3.2.2 Products of Graph Matrices

We now analyze what happens when we take the products of graph matrices. Roughly speaking, we will have that if \( \alpha \) can be decomposed into left, middle, and right parts \( \sigma, \tau, \) and \( \sigma^T \) then \( M_\alpha \approx M_\sigma M_\tau M_{\sigma^T} \).

However, this is only an approximation rather than an equality, and this will be the source of considerable technical difficulties.

We begin with a concatenation operation on ribbons.

**Definition 3.17** (Ribbon Concatenation). If \( R_1 \) and \( R_2 \) are two ribbons such that \( V(R_1) \cap V(R_2) = B_{R_1} = A_{R_2} \) and either \( R_1 \) or \( R_2 \) contains no edges entirely within \( B_{R_1} = A_{R_2} \) then we define \( R_1 \circ R_2 \) to be the ribbon formed by gluing together \( R_1 \) and \( R_2 \) along \( B_{R_1} = A_{R_2} \). In other words,

1. \( V(R_1 \circ R_2) = V(R_1) \cup V(R_2) \)
2. \( E(R_1 \circ R_2) = E(R_1) \cup E(R_2) \)
3. \( A_{R_1 \circ R_2} = A_{R_1} \) and \( B_{R_1 \circ R_2} = B_{R_2} \).

The following proposition is easy to check.

**Proposition 3.18.** Whenever \( R_1, R_2 \) are ribbons such that \( R_1 \circ R_2 \) is defined, \( M_{R_1} M_{R_2} = M_{R_1 \circ R_2} \)

We have an analogous definition for concatenating shapes:

**Definition 3.19** (Shape Concatenation). If \( \alpha_1 \) and \( \alpha_2 \) are two shapes such that \( V(\alpha_1) \cap V(\alpha_2) = V_{\alpha_1} = U_{\alpha_2} \) and either \( \alpha_1 \) or \( \alpha_2 \) contains no edges entirely within \( V_{\alpha_1} = U_{\alpha_2} \) then we define \( \alpha_1 \circ \alpha_2 \) to be the shape formed by gluing together \( \alpha_1 \) and \( \alpha_2 \) along \( V_{\alpha_1} = U_{\alpha_2} \). In other words,

1. \( V(\alpha_1 \circ \alpha_2) = V(\alpha_1) \cup V(\alpha_2) \)
2. \( E(\alpha_1 \circ \alpha_2) = E(\alpha_1) \cup E(\alpha_2) \)
3. \( U_{\alpha_1 \circ \alpha_2} = U_{\alpha_1} \) and \( V_{\alpha_1 \circ \alpha_2} = V_{\alpha_2} \).

The next proposition, again easy to check, shows that the shape concatenation operation respects the left/middle/right part decomposition.

**Proposition 3.20.** If \( \alpha \) can be decomposed into left, middle, and right parts \( \sigma, \tau, \sigma^T \) then \( \alpha = \sigma \circ \tau \circ \sigma^T \).
We now discuss why \( M_\alpha = M_{\sigma \circ \rho}^T \approx M_\sigma M_\tau M_{\rho^T} \) is only an approximation rather than an equality. Consider the difference \( M_\sigma M_\tau M_{\rho^T} - M_{\sigma \circ \rho}^T \). The graph matrix \( M_{\sigma \circ \rho}^T \) decomposes (by definition) into a sum over injective maps \( \phi : \mathcal{V}(\sigma \circ \tau \circ \rho^T) \to [n] \). Also by expanding definitions, the product \( M_\sigma M_\tau M_{\rho^T} \) expands into a sum over triples of injective maps \((\varphi_1, \varphi_2, \varphi_3)\), where \( \varphi_1 : \mathcal{V}(\sigma) \to [n], \varphi_2 : \mathcal{V}(\tau) \to [n], \varphi_3 : \mathcal{V}(\rho^T) \to [n] \) where \( \varphi_1 \) and \( \varphi_2 \) agree on \( \mathcal{V}_\sigma = U_\tau \) and \( \varphi_2 \) and \( \varphi_3 \) agree on \( \mathcal{V}_\tau = U_{\rho^T} \).

If they are combined into one map \( \varphi : \mathcal{V}(\sigma \cup \tau \cup \rho^T) \to [n] \), the resulting \( \varphi \) may not be injective because \( \varphi_1(\mathcal{V}(\sigma)), \varphi_2(\mathcal{V}(\tau)), \varphi_3(\mathcal{V}(\rho^T)) \) may have nontrivial intersection (beyond \( \varphi_1(\mathcal{V}_\sigma) \) and \( \varphi_2(\mathcal{V}_\tau) \)). We call the resulting terms intersection terms and handling them properly is a major part of the technical analysis.

**Remark 3.21.** Actually, the approximation \( M_\alpha = M_{\sigma \circ \rho}^T \approx M_\sigma M_\tau M_{\rho^T} \) is also off by a multiplicative constant because there is also a subtle issue involving the automorphism groups of these shapes. For now, we ignore this issue. For details about this issue, see Lemma 7.78.

### 3.3 Shape Coefficient Matrices

The idea for our analysis is as follows. Given a matrix-valued function \( \Lambda \) which is symmetric under permutations of \([n]\), we write \( \Lambda = \sum_\alpha \lambda_\alpha M_\alpha \). We then break each shape \( \alpha \) up into left, middle, and right parts \( \sigma, \tau, \) and \( \rho^T \).

For this analysis, we use shape coefficient matrices \( H_\tau \) whose rows and columns are indexed by left shapes and whose entries depend on the coefficients \( \lambda_\alpha \). We choose these matrices so that

\[
\Lambda = \sum_\tau H_\tau(\sigma, \rho^T) M_{\sigma \circ \rho^T} \approx \sum_\tau H_\tau(\sigma, \rho^T) M_\sigma M_\tau M_{\rho^T}
\]

To set this up, we separate the possible middle parts \( \tau \) into groups based on the size of \( U_\tau \) and whether or not they are trivial.

**Definition 3.22.** We define \( \mathcal{I}_{\text{mid}} \) to be the set of all possible \( U_\tau \). Here \( \mathcal{I}_{\text{mid}} \) is the set of tuples of unspecified vertices of the form \( U = (u_1, \ldots, u_k) \) where \( 0 \leq k \leq d \).

**Definition 3.23.** We say that a proper middle shape \( \tau \) is trivial if \( E(\tau) = \emptyset \) and \( |U_\tau \cap V_\tau| = |U_\tau| = |V_\tau| \) (i.e. \( V_\tau \) is a permutation of \( U_\tau \)).

For simplicity, the only proper trivial middle parts \( \tau \) we consider are shapes \( Id_{U_\tau} \) corresponding to identity matrices.

**Definition 3.24.** Given a tuple of unspecified vertices \( U = (u_1, \ldots, u_{|U|}) \) we define \( Id_U \) to be the shape where \( V(Id_U) = U, U_{Id_U} = V_{Id_U} = U, \) and \( E(Id_U) = \emptyset \).

We group all of the proper non-trivial middle parts \( \tau \) into sets \( \mathcal{M}_{U_\tau} \) based on the size of \( U_\tau \).

**Definition 3.25.** Given a tuple of unspecified vertices \( U = (u_1, \ldots, u_{|U|}) \), we define \( \mathcal{M}_{U_\tau} \) to be the set of proper non-trivial middle parts \( \tau \) such that \( U_\tau \) and \( V_\tau \) have the same size as \( U \). Note that \( U_\tau \) and \( V_\tau \) may intersect each other arbitrarily.

With these definitions, we can now define our shape coefficient matrices.
Definition 3.26. Given \( U \in \mathcal{I}_{mid} \), we define \( L_U \) to be the set of left shapes \( \sigma \) such that \( |V_\sigma| = |U| \).

Definition 3.27. For each \( U \in \mathcal{I}_{mid} \), we define the shape coefficient matrix \( H_{id_U} \) to be the matrix indexed by left shapes \( \sigma, \sigma' \in L_U \) with entries \( H_{id_U}(\sigma, \sigma') = \frac{1}{|U|!} \lambda_{\sigma \sigma'} t^{|U|} \).

Definition 3.28. For each \( U \in \mathcal{I}_{mid} \), for each \( \tau \in M_U \), we define the shape coefficient matrix \( H_\tau \) to be the matrix indexed by left shapes \( \sigma, \sigma' \in L_U \) with entries \( H_\tau(\sigma, \sigma') = \frac{1}{|U|!} \lambda_{\sigma \sigma'} t^{|U|} \).

With these shape coefficient matrices, we have the following decomposition of \( \Lambda = \sum_\alpha \lambda_\alpha M_\alpha \).

Lemma 3.29. \( \Lambda = \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in L_U} H_{id_U}(\sigma, \sigma') M_{\sigma \sigma'} t^{|U|} + \sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in M_U} \sum_{\sigma, \sigma' \in L_U} H_\tau(\sigma, \sigma') M_{\sigma \sigma'} t^{|U|} \).

We defer the proof of this lemma to Lemma 7.81.

For technical reasons, we need to define one more operation to handle intersection terms. We call this operation the \( -\gamma, \gamma \) operation.

Definition 3.30. Given \( U, V \in \mathcal{I}_{mid} \) where \( |U| > |V| \), we define \( \Gamma_{U, V} \) to be the set of left parts \( \gamma \) such that \( |U_\gamma| = |U| \) and \( V_\gamma = |V| \).

Definition 3.31. Given \( U, V \in \mathcal{I}_{mid} \) where \( |U| > |V| \), a shape coefficient matrix \( H_{id_V} \), and a \( \gamma \in \Gamma_{U, V} \), we define the shape coefficient matrix \( H_{id_V}^{-\gamma, \gamma} \) to be the matrix indexed by left shapes \( \sigma, \sigma' \in L_U \) with entries \( H_{id_V}^{-\gamma, \gamma}(\sigma, \sigma') = H(\sigma \circ \gamma, \sigma' \circ \gamma) \).

### 3.4 Informal Theorem Statement

We are now ready to state a simplified, qualitative version of our main theorem. For the full, quantitative version of our main theorem, see Theorem 7.101.

**Theorem 3.32.** There exist functions \( f(\tau) : \mathcal{M}_U \to \mathbb{R} \) and \( f(\gamma) : \Gamma_{U, V} \to \mathbb{R} \) depending on \( n \) and other parameters such that if \( \Lambda = \sum_\alpha \lambda_\alpha M_\alpha \) and the following conditions hold:

1. For all \( U \in \mathcal{I}_{mid} \), \( H_{id_U} \geq 0 \).
2. For all \( U \in \mathcal{I}_{mid} \) and all \( \tau \in \mathcal{M}_U \),
   \[
   \begin{bmatrix}
   H_{id_U} & f(\tau) H_\tau \\
   f(\tau) H_\tau^T & H_{id_U}
   \end{bmatrix} \succeq 0
   \]
3. For all \( U, V \in \mathcal{I}_{mid} \) such that \( |U| > |V| \) and all \( \gamma \in \Gamma_{U, V} \), \( H_{id_V}^{-\gamma, \gamma} \preceq f(\gamma) H_{id_V} \).

then with probability at least \( 1 - o(1) \) over \( G \sim \{\pm 1\}^n \) it holds that \( \Lambda(G) \geq 0 \).

**Remark 3.33.** As we will demonstrate in the remainder of this paper, our machinery works well when the coefficients \( \lambda_\alpha \) for each shape have an exponential decay in both \( |V(\alpha)| \) and \( |E(\alpha)| \). However, since our machinery is highly technical with many different parts, it does not work as well as if the coefficients have a different kind of decay.
3.5 An informal application to planted clique

Before we move on to further definitions needed for a more complete statement of the main theorem, we present an informal example.

Example 3.34. When the pseudo-calibration method is applied to prove an SoS lower bound for the planted clique problem in $n$ node graphs with clique size $k$, as in [BHK+16], the matrix-valued function which results is $\Lambda = \sum_{\alpha: |V(\alpha)| \leq t} \left( \frac{k}{n} \right)^{|V(\alpha)|} M_\alpha$ where $t \approx \log(n)$. One may then compute that the matrices $H_{id_{U\tau}}$ and $H_\tau$ are as follows (at least so long as $|V(\sigma)|, |V(\tau)|, |V(\sigma')| \ll t$; we ignore this detail for now). For all $r \in [0, \frac{t}{2}]$,

1. For $U$ with $|U| = r$, $H_{id_{U\tau}}(\sigma, \sigma') = \left( \frac{k}{n} \right)^{|V(\sigma)| + |V(\sigma')| - r}$
2. For all proper, non-trivial middle shapes $\tau$ such that $|U_\tau| = |V_\tau| = r$,
   $$H_\tau(\sigma, \sigma') = \left( \frac{k}{n} \right)^{|V(\sigma)| + |V(\sigma')| + |V(\tau)| - 2r}$$

Defining $v_\tau$ to be the vector such that $v_\tau(\sigma) = \left( \frac{k}{n} \right)^{|V(\sigma)| - \frac{r}{2}}$, we have that

1. For $U$ with $|U| = r$, $H_{id_{U\tau}} = v_\tau|U|v_\tau^T$
2. For all proper, non-trivial middle shapes $\tau$ such that $|U_\tau| = |V_\tau| = r$, $H_\tau = \left( \frac{k}{n} \right)^{|V(\tau)| - r} v_\tau v_\tau^T$
3. For all left parts $\gamma$, $H_{id_{U\gamma}}^{-\gamma, \gamma} = \left( \frac{k}{n} \right)^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} v_\gamma|U_\gamma|v_\gamma^T$

It turns out in this setting that we can take $f(\tau)$ to be $\tilde{O}(n^{|V(\tau)| - |U_\tau|})$ and $f(\gamma)$ to be $\tilde{O}(n^{|V(\gamma)| |U_\gamma|})$. Thus, as long as $k \ll \sqrt{n}$,

1. For any $U$ and all $\tau$ such that $V_\tau \neq U_\tau$ with $|U_\tau| = |V_\tau| = |U|$, $f(\tau)H_\tau \preceq H_{id_{U\tau}}$
2. For all non-trivial left parts $\gamma$, $H_{id_{U\gamma}}^{-\gamma, \gamma} \preceq f(\gamma)H_{id_{U\gamma}}$

Remark 3.35. This does not quite satisfy the conditions of Theorem 3.32 because there are $\tau$ such that $V_\tau = U_\tau$ but which are non-trivial because $E(\tau) \neq \emptyset$. For these $\tau$, condition 2 of Theorem 3.32 fails.

In order to prove their SoS lower bounds for planted clique, [BHK+16] handle this issue by grouping together all of the $\tau$ where $V_\tau = U_\tau$ into the indicator function for whether $V_\tau = U_\tau$ is a clique.

In this paper, we get around this issue by instead considering the planted slightly denser subgraph problem. This introduces an edge decay into the coefficients. For details, see Section 4.

3.6 Generalizing the machinery

In this section, we restricted ourselves to the case when the $\{-1, 1\}^{\binom{n}{2}}$ for simplicity. However, for our results we will need to handle more general types of inputs. We now briefly describe which kinds of inputs we will need to handle and how we handle them.
1. In general, the entries of the input may be labeled by more than 2 indices. For example, for tensor PCA on order 3 tensors, the entries of the input are indexed by 3 indices. To handle this, we will have shapes which have hyperedges rather than edges.

2. In general, the entries of the input will come from a distribution $\Omega$ rather than being $\pm 1$. To handle this, we will take an orthonormal basis $\{h_k\}$ for $\Omega$. We will then give each edge/hyperedge a label $l$ to specify which polynomial $h_l$ should be applied to that entry of the input.

3. In general, there may be $t$ different types of indices rather than just one type of index. In this case, the symmetry group will be $S_{n_1} \times \ldots \times S_{n_t}$ rather than $S_n$. To handle this, we will have shapes with different types of vertices.

We formally make these generalizations in Section 7.

3.7 Further definitions needed for our applications

We will describe some more notations and definitions that will be useful to us to describe the qualitative bounds for our applications. For each of our applications, we will describe the corresponding modifications needed to the definitions already in place and present new definitions where necessary.

3.7.1 Planted slightly denser subgraph

Since the input is a graph $G \in \{-1, 1\}^{\binom{n}{2}}$, most of what we introduced already apply to this setting. To describe the moment matrix, we need to define the truncation parameter.

**Definition 3.36 (Truncation parameters).** For integers $D_{sos}, D_V \geq 0$, say that a shape $\alpha$ satisfies the truncation parameters $D_{sos}, D_V$ if

- The degrees of the monomials that $U_\alpha$ and $V_\alpha$ correspond to, are at most $\frac{D_{sos}}{2}$
- The left part $\sigma$, the middle part $\tau$ and the right part $\sigma'$ of $\alpha$ satisfy $|V(\sigma)|, |V(\tau)|, |V(\sigma')| \leq D_V$

3.7.2 Tensor PCA

We consider the input to be a tensor $A \in \mathbb{R}^{[n]^k}$. The input entries are now sampled from the distribution $\mathcal{N}(0,1)$ instead of $\{-1,1\}$. So, we will work with the Hermite basis of polynomials.

Let the standard unnormalized Hermite polynomials be denoted as $h_0(x) = 1, h_1(x) = x, h_2(x) = x^2 - 1, \ldots$. Then, we work with the basis $h_a(A) := \prod_{e \in [n]^k} h_e(A_e)$ over $a \in \mathbb{N}^{[n]^k}$. Accordingly, we will modify the graphs that represent ribbons (and by extension, shapes), to have labeled hyperedges of arity $k$. So, an hyperedge $e$ with a label $t$ will correspond to the hermite polynomial $h_t(A_e)$.

**Definition 3.37 (Hyperedges).** Instead of standard edges, we will have labeled hyperedges of arity $k$ in the underlying graphs for our ribbons as well as shapes. The label for an hyperedge $e$, denoted $l_e$, is an element of $\mathbb{N}$ which will correspond to the Hermite polynomial being evaluated on that entry.

Note that our hyperedges are ordered since the tensor $A$ is not necessarily symmetric.
For variables \( x_1, \ldots, x_n \), the rows and columns of our moment matrix will now correspond to monomials of the form \( \prod_{i=1}^n x_i^{p_i} \) for \( p_i \geq 0 \). To capture this, we use the notion of index shape pieces and index shapes.

Informally, we split the above monomial product into groups based on their powers and each such group will form an index shape piece.

**Definition 3.38** (Index shape piece). An index shape piece \( U_i = ((U_{i,1}, \ldots, U_{i,t}), p_i) \) is a tuple of indices \( (U_{i,1}, \ldots, U_{i,t}) \) along with a power \( p_i \in \mathbb{N} \). Let \( V(U_i) \) be the set \{ \( U_{i,1}, \ldots, U_{i,t} \) \} of vertices of this index shape piece. When clear from context, we use \( U_i \) instead of \( V(U_i) \).

If we realize \( U_{i,1}, \ldots, U_{i,t} \) to be indices \( a_1, \ldots, a_t \in [n] \), then, this realization of this index shape piece corresponds to the monomial \( \prod_{i=1}^t x_{a_i}^{p_i} \).

**Definition 3.39** (Index shape). An index shape \( U \) is a set of index shape pieces \( U_i \) that have different powers. Let \( V(U) \) be the set of vertices \( \cup_i V(U_i) \). When clear from context, we use \( U \) instead of \( V(U) \).

Observe that each realization of an index shape corresponds to a row or column of the moment matrix.

**Definition 3.40.** For two index shapes \( U, V \), we write \( U \equiv V \) if for all powers \( p \), the index shape pieces of power \( p \) in \( U \) and \( V \) have the same length.

**Definition 3.41.** Define \( \mathcal{I}_{\text{mid}} \) to be the set of all index shapes \( U \) that contain only index shape pieces of power \( 1 \).

In the definition of shapes, the distinguished set of vertices should now be replaced by index shapes.

**Definition 3.42** (Shapes). Shapes are tuples \( \alpha = (H_\alpha, U_\alpha, V_\alpha) \) where \( H_\alpha \) is a graph with hyperedges of arity \( k \) and \( U_\alpha, V_\alpha \) are index shapes such that \( U_\alpha, V_\alpha \subseteq V(H_\alpha) \).

**Definition 3.43** (Proper shape). A shape \( \alpha \) is proper if it has no isolated vertices outside \( U_\alpha \cup V_\alpha \), no multi-edges and all the edges have a nonzero label.

To define the notion of vertex separators, we modify the notion of paths for hyperedges.

**Definition 3.44** (Path). A path is a sequence of vertices \( u_1, \ldots, u_t \) such that \( u_i, u_{i+1} \) are in the same hyperedge, for all \( i \leq t - 1 \).

The notions of vertex separator and decomposition into left, middle and right parts are identically defined with the above notion of hyperedges and paths. In Section 7, we will show that they are well defined.

In the definition of trivial shape \( \tau \), we now require \( U_\tau \equiv V_\tau \). For \( U \in \mathcal{I}_{\text{mid}} \), \( \mathcal{M}_U \) will be the set of proper non-trivial middle parts \( \tau \) with \( U_\tau \equiv V_\tau \equiv U \) and \( \mathcal{L}_U \) will be the set of left parts \( \sigma \) such that \( V_\sigma \equiv U \). Similarly, for \( U, V \in \mathcal{I}_{\text{mid}} \), \( \mathcal{L}_{U,V} \) will be the set of left parts \( \gamma \) such that \( U_\gamma \equiv U \) and \( V_\gamma \equiv V \).

In order to define the moment matrix, we need to truncate our shapes based on the number of vertices and the labels on our hyperedges. So, we make the following definition.

**Definition 3.45** (Truncation parameters). For integers \( D_{\text{sos}}, D_V, D_E \geq 0 \), say that a shape \( \alpha \) satisfies the truncation parameters \( D_{\text{sos}}, D_V, D_E \) if

- The degrees of the monomials that \( U_\alpha \) and \( V_\alpha \) correspond to, are at most \( \frac{D_{\text{sos}}}{2} \)
- The left part \( \sigma \), the middle part \( \tau \) and the right part \( \sigma^T \) of \( \alpha \) satisfy \( |V(\sigma)|, |V(\tau)|, |V(\sigma^T)| \leq D_V \)
- For each \( e \in E(\alpha) \), \( l_e \leq D_E \).
3.7.3 Sparse PCA

We consider the $m$ vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ to be the input. Similar to Tensor PCA, we will work with the Hermite basis of polynomials since the entries are sampled from the distribution $\mathcal{N}(0, 1)$.

In particular, if we denote the unnormalized Hermite polynomials by $h_0(x) = 1, h_1(x) = x, h_2(x) = x^2 - 1, \ldots$, then, we work with the basis $h_a(v) := \prod_{1 \leq i \leq n} h_{a_{i,j}}(v_{i,j})$ over $a \in \mathbb{N}^{m \times n}$. To capture these bases, we will modify the graphs that represent ribbons (and by extension, shapes), to be bipartite graphs with two types of vertices, and have labeled edges that go across vertices of different types. So, an edge $(i, j)$ with label $t$ between a vertex $i$ of type 1 and a vertex $j$ of type 2 will correspond to $h_t(v_{i,j})$.

**Definition 3.46** (Vertices). We will have two types of vertices, the vertices corresponding to the $m$ input vectors that we call type 1 vertices and the vertices corresponding to ambient dimension of the space that we call type 2 vertices.

**Definition 3.47** (Edges). Edges will go across vertices of different types, thereby forming a bipartite graph. An edge between a type 1 vertex $i$ and a type 2 vertex $j$ corresponds to the input entry $v_{i,j}$. Each edge will have a label in $\mathbb{N}$ corresponding to the Hermite polynomial evaluated on that entry.

We will have variables $x_1, \ldots, x_n$ in our SoS program, so we will work with index shape pieces and index shapes as in Tensor PCA, since the rows and columns of our moment matrix will now correspond to monomials of the form $\prod_{1 \leq i \leq n} x_i^{p_i}$ for $p_i \geq 0$. But since in our decompositions into left, right and middle parts, we will have type 2 vertices as well in the vertex separators, we will define a generalized notion of index shape pieces and index shapes.

**Definition 3.48** (Index shape piece). An index shape piece $U_i = ((U_{i,1}, \ldots, U_{i,t_i}), t_i, p_i)$ is a tuple of indices $(U_{i,1}, \ldots, U_{i,t_i})$ along a type $t_i \in \{1, 2\}$ with a power $p_i \in \mathbb{N}$. Let $V(U_i)$ be the set $\{U_{i,1}, \ldots, U_{i,t_i}\}$ of vertices of this index shape piece. When clear from context, we use $U_i$ instead of $V(U_i)$.

For an index shape piece $((U_{i,1}, \ldots, U_{i,t_i}), t_i, p_i)$ with type $t_i = 2$, if we realize $U_{i,1}, \ldots, U_{i,t_i}$ to be indices $a_1, \ldots, a_t \in [n]$, then, this index shape pieces correspond this to the monomial $\prod_{1 \leq i \leq n} x_{a_i}^{p_i}$.

**Definition 3.49** (Index shape). An index shape $U$ is a set of index shape pieces $U_i$ that have either have different types or different powers. Let $V(U)$ be the set of vertices $\sqcup_i V(U_i)$. When clear from context, we use $U$ instead of $V(U)$.

Observe that each realization of an index shape corresponds to a row or column of the moment matrix. For our moment matrix, the only nonzero rows correspond to index shapes that have only index shape pieces of type 2, since the only SoS variables are $x_1, \ldots, x_n$, but in order to do our analysis, we need to work with the generalized notion of index shape that allow index shape pieces of both types.

**Definition 3.50.** For two index shapes $U, V$, we write $U \equiv V$ if for all types $t$ and all powers $p$, the index shape pieces of type $t$ and power $p$ in $U$ and $V$ have the same length.

**Definition 3.51.** Define $\mathcal{I}_{mid}$ to be the set of all index shapes $U$ that contain only index shape pieces of power 1.
Section 7, the third qualitative condition we’d like to show is as follows: For all proper non-trivial middle parts \( \tau \) and truncation parameters \( D \), we make the following definition.

Definition 3.52 (Weight of an index shape). Suppose we have an index shape \( U = \{ U_1, U_2 \} \in \mathcal{I}_{mid} \) where
\[ U_1 = ((U_{1,1}, \ldots , U_{1,|U_1|}), 1, 1) \]
and \( U_2 = ((U_{2,1}, \ldots , U_{2,|U_2|}), 2, 1) \) is an index shape piece of type 1 and 2. Then, define the weight of this index shape to be
\[ w(U) = \sqrt{m_{U_1}} \sqrt{n_{U_2}}. \]

We now give the modified definition of shapes.

Definition 3.53 (Shapes). Shapes are tuples \( \alpha = (H_\alpha, U_\alpha, V_\alpha) \) where \( H_\alpha \) is a graph with two types of vertices, has labeled edges only across vertices of different types and \( U_\alpha, V_\alpha \) are index shapes such that \( U_\alpha \subseteq V(H_\alpha) \).

Definition 3.54 (Proper shape). A shape \( \alpha \) is proper if it has no isolated vertices outside \( U_\alpha \cup V_\alpha \), no multi-edges and all the edges have a nonzero label.

In Section 7, we will show that with this new definition of weight and shapes, any shape \( \alpha \) has a unique decomposition into \( \sigma \circ \tau \circ \sigma^T \) where \( \sigma, \tau, \sigma^T \) are left, middle and right parts respectively. Here, \( \tau \) may possibly be improper.

In the definition of trivial shape \( \tau \), we now require \( U_\tau \equiv V_\tau \). For \( U \in \mathcal{I}_{mid}, \mathcal{M}_U \) will be the set of proper non-trivial middle parts \( \tau \) with \( U_\tau \equiv V_\tau \equiv U \) and \( L_U \) will be the set of left parts \( \sigma \) such that \( V_\sigma \equiv U \). Similarly, for \( U, V \in \mathcal{I}_{mid} \), \( L_{U,V} \) will be the set of left parts \( \gamma \) such that \( U_\gamma \equiv U \) and \( V_\gamma \equiv V \).

Finally, in order to define the moment matrix, we need to truncate our shapes based on the number of vertices and the labels on our edges. So, we make the following definition.

Definition 3.55 (Truncation parameters). For integers \( D_{sos}, D_V, D_E \geq 0 \), say that a shape \( \alpha \) satisfies the truncation parameters \( D_{sos}, D_V, D_E \) if

1. The degrees of the monomials that \( U_\alpha \) and \( V_\alpha \) correspond to, are at most \( \frac{D_{sos}}{2} \).
2. The left part \( \sigma \), the middle part \( \tau \) and the right part \( \sigma^T \) of \( \alpha \) satisfy \( |V(\sigma)|, |V(\tau)|, |V(\sigma^T)| \leq D_V \).
3. For each \( e \in E(\alpha), I_e \leq D_E \).

3.7.4 Reloading the third condition

In Theorem 3.32, the third qualitative condition we’d like to show is as follows: For all \( U, V \in \mathcal{I}_{mid} \) such that \( |U| > |V| \) and all \( \gamma \in \Gamma_{U,V}, H_{Id_{U,V}}^{\gamma \gamma} \leq f(\gamma)H_{Id_{U,V}} \).

For technical reasons, we won’t be able to show this directly. To handle this, we instead work with a slight modification of \( H_{Id_{U,V}} \), a matrix \( H'_\gamma \) that’s very close to \( H_{Id_{U,V}} \). So, what we will end up showing is: For all \( \gamma \in \Gamma_{U,V}, H_{Id_{U,V}}^{\gamma \gamma} \leq f(\gamma)H'_\gamma \).

Let \( D_V \) be the truncation parameter. A canonical choice for \( H'_\gamma \) is to take

1. \( H'_\gamma(\sigma, \sigma') = H_{Id_{U,V}}(\sigma, \sigma') \) whenever \( |V(\sigma \circ \gamma)| \leq D_V \) and \( |V(\sigma' \circ \gamma)| \leq D_V \).
2. \( H'_\gamma(\sigma, \sigma') = 0 \) whenever \( |V(\sigma \circ \gamma)| > D_V \) or \( |V(\sigma' \circ \gamma)| > D_V \).

With this choice, \( H'_\gamma \) is the same as \( H_{Id_{U,V}} \) up to truncation error. We will formally bound the errors in the quantitative sections after we introduce the full machinery.
4 Qualitative bounds for Planted slightly denser subgraph

4.1 Pseudo-calibration

We will pseudo-calibrate with respect the following pair of random and planted distributions which we denote \( v \) and \( \mu \) respectively.

- Random distribution: Sample \( G \) from \( G(n, \frac{1}{2}) \)
- Planted distribution: Let \( k \) be an integer and let \( p > \frac{1}{2} \). Sample a graph \( G' \) from \( G(n, \frac{1}{2}) \). Choose a random subset \( S \) of the vertices, where each vertex is picked independently with probability \( \frac{1}{n} \). For all pairs \( i, j \) of vertices in \( S \), rerandomize the edge \((i, j)\) where the probability of \((i, j)\) being in the graph is now \( p \). Set \( G \) to be the resulting graph.

We assume that the input is given as \( G_{i,j} \) for \( i, j \in \binom{[n]}{2} \) where \( G_{i,j} \) is 1 if the edge \((i, j)\) is present in the graph and \(-1\) otherwise. We work with the Fourier basis \( \chi_E \) defined as \( \chi_E(G) := \prod_{(i,j) \in E} G_{i,j} \). For a subset \( I \subseteq [n] \), define \( x_I := \prod_{i \in I} x_i \).

**Lemma 4.1.** Let \( I \subseteq [n] \), \( E \subseteq \binom{[n]}{2} \). Then,

\[
\mathbb{E}_{\mu}[x_I \chi_E(G)] = \left( \frac{k}{n} \right)^{|I \cup V(E)|} (2p - 1)^{|E|}
\]

**Proof.** When we sample \( (G, S) \) from \( \mu \), we condition on whether \( I \cup V(E) \subseteq S \).

\[
\mathbb{E}_{(G,S) \sim \mu} [x_I \chi_E(G)] = \mathbb{E}_{(G,S) \sim \mu} [x_I \chi_E(G)| I \cup V(E) \subseteq S] \\
+ \mathbb{E}_{(G,S) \sim \mu} [x_I \chi_E(G)| I \cup V(E) \not\subseteq S]
\]

We claim that the second term is 0. In particular, \( \mathbb{E}_{(G,S) \sim \mu} [x_I \chi_E(G)| I \cup V(E) \not\subseteq S] = 0 \) because when \( I \cup V(E) \not\subseteq S \), either \( S \) doesn’t contain a vertex in \( I \) or an edge \((i,j) \in E \) is outside \( S \). If \( S \) doesn’t contain a vertex in \( I \), then \( x_I = 0 \) and hence, the quantity is 0. And if an edge \((i,j) \in E \) is outside \( S \), since this edge is sampled with probability \( \frac{1}{n} \) by taking expectations, the quantity \( \mathbb{E}_{(G,S) \sim \mu} [x_I \chi_E(G)| I \cup V(E) \not\subseteq S] \) is 0.

Finally, note that \( \mathbb{E}_{(G,S) \sim \mu} [I \cup V(E) \subseteq S] = \left( \frac{k}{n} \right)^{|I \cup V(E)|} \) and

\[
\mathbb{E}_{(G,S) \sim \mu} [x_I \chi_E(G)| I \cup V(E) \subseteq S] = \mathbb{E}_{(G,S) \sim \mu} [\chi_E(G)| V(E) \subseteq S] = (2p - 1)^{|E|}
\]

The last equality follows because for each edge \( e \in E \), since \( e \) is present independently with probability \( p \), the expected value of \( \chi_e \) is \( 1 \cdot p + (-1) \cdot (1 - p) = 2p - 1 \).

Now, we can write the moment matrix in terms of graph matrices.

**Definition 4.2.** Define the degree of SoS to be \( D_{sos} = n^{C_{sos}} \) for some constant \( C_{sos} > 0 \) that we choose later.
**Definition 4.3** (Truncation parameter). Define the truncation parameter to be \( D_V = n^{C_V \varepsilon} \) for some constant \( C_V > 0 \).

**Remark 4.4** (Choice of parameters). We first set \( \varepsilon \) to be a sufficiently small constant. Based on this choice, we will set \( C_V \) to be a sufficiently small constant to satisfy all the inequalities we use in our proof. Based on these choices, we can choose \( C_{sos} \) to be sufficiently small to satisfy the inequalities we use.

We will now describe the decomposition of the moment matrix \( \Lambda \).

**Definition 4.5.** If a shape \( \alpha \) satisfies the following properties:
- \( \alpha \) is proper,
- \( \alpha \) satisfies the truncation parameter \( D_{sos}, D_V \).

then define
\[
\lambda_\alpha = \left( \frac{k}{n} \right)^{|V(\alpha)|} (2p - 1)^{E(\alpha)}
\]

**Corollary 4.6.** \( \Lambda = \sum \lambda_\alpha M_\alpha \).

### 4.2 Proving positivity - Qualitative bounds

We use the canonical definition of \( H'_\gamma \) from Section 3.7.4. In this section, we will prove the main qualitative bounds Lemma 4.7, Lemma 4.9 and Lemma 4.11.

**Lemma 4.7.** For all \( U \in \mathcal{I}_{mid} \), \( H_{Id_U} \succeq 0 \)

We define the following quantity to capture the contribution of the vertices within \( \tau \) to the Fourier coefficients.

**Definition 4.8.** For \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), define
\[
S(\tau) = \left( \frac{k}{n} \right)^{|V(\tau)| - |U_\tau|} (2p - 1)^{E(\tau)}
\]

**Lemma 4.9.** For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \),
\[
\begin{bmatrix}
H_{Id_U}^{\tau} & H_{\tau} \\
H_{\tau}^T & H_{Id_U}^{\tau}
\end{bmatrix} \succeq 0
\]

We define the following quantity to capture the contribution of the vertices within \( \gamma \) to the Fourier coefficients.

**Definition 4.10.** For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \) and \( \gamma \in \Gamma_{U,V} \), define
\[
S(\gamma) = \left( \frac{k}{n} \right)^{|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}} (2p - 1)^{E(\gamma)}
\]
**Lemma 4.11.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,

$$
\frac{|\text{Aut}(V)|}{|\text{Aut}(U)|} \cdot \frac{1}{S(\gamma)^2} H_{\text{Id}_V}^{-\gamma} \gamma \leq H_{\gamma}
$$

In order to prove these bounds, we define the following quantity to capture the contribution of the vertices within $\sigma$ to the Fourier coefficients.

**Definition 4.12.** For a shape $\sigma \in \mathcal{L}$, define

$$
T(\sigma) = \left(\frac{k}{n}\right)^{|\mathcal{V}(\sigma)| - |\mathcal{U}_z|} (2p - 1)^{|E(\sigma)|}
$$

**Definition 4.13.** For $U \in \mathcal{I}_{\text{mid}}$, define $v_U$ to be the vector indexed by $\sigma \in \mathcal{L}$ such that $v_U(\sigma) = T(\sigma)$ if $\sigma \in \mathcal{L}_U$ and 0 otherwise.

**Proposition 4.14.** For all $U \in \mathcal{I}_{\text{mid}}$, $\rho \in \mathcal{P}_U$, $H_{\text{Id}_U} = \frac{1}{|\text{Aut}(U)|} v_U v_U^T$.

**Proof.** This follows by verifying the conditions of Definition 4.5.

This immediately implies that for all $U \in \mathcal{I}_{\text{mid}}$, $H_{\text{Id}_U} \succeq 0$, which is Lemma 4.7.

We restate Definition 4.8 for convenience.

**Definition 4.8.** For $U \in \mathcal{I}_{\text{mid}}$ and $\tau \in \mathcal{M}_U$, define

$$
S(\tau) = \left(\frac{k}{n}\right)^{|\mathcal{V}(\tau)| - |\mathcal{U}_z|} (2p - 1)^{|E(\tau)|}
$$

**Proposition 4.15.** For any $U \in \mathcal{I}_{\text{mid}}$ and $\tau \in \mathcal{M}_U$, $H_{\tau} = \frac{1}{|\text{Aut}(U)|} S(\tau) v_U v_U^T$.

**Proof.** This follows by a straightforward verification of the conditions of Definition 4.5.

**Lemma 4.9** immediately follows.

**Lemma 4.9.** For all $U \in \mathcal{I}_{\text{mid}}$ and $\tau \in \mathcal{M}_U$,

$$
\begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|} H_{\text{Id}_U}^T & \frac{H_{\tau}}{|\text{Aut}(U)|} \\
H_{\tau}^T & \frac{S(\tau)}{|\text{Aut}(U)|} H_{\text{Id}_U}
\end{bmatrix} \succeq 0
$$

**Proof.**

$$
\begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|} H_{\text{Id}_U}^T & \frac{H_{\tau}}{|\text{Aut}(U)|} \\
H_{\tau}^T & \frac{S(\tau)}{|\text{Aut}(U)|} H_{\text{Id}_U}
\end{bmatrix} = \begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|} v_U v_U^T & \frac{S(\tau)}{|\text{Aut}(U)|} v_U v_U^T \\
\frac{S(\tau)}{|\text{Aut}(U)|} v_U v_U^T & \frac{S(\tau)}{|\text{Aut}(U)|} v_U v_U^T
\end{bmatrix} \succeq 0
$$

We restate Definition 4.10 for convenience.
Definition 4.10. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and $\gamma \in \Gamma_{U, V}$, define

$$S(\gamma) = \left(\frac{k}{n}\right)^{|\mathcal{V}(\gamma)|-\frac{|U|+|V|}{2}} (2p - 1)|E(\gamma)|$$

Proposition 4.16. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for all $\gamma \in \Gamma_{U, V}$,

$$H_{id}^{-\gamma} = \left|\frac{\text{Aut}(U)}{\text{Aut}(V)}\right| S(\gamma)^2 H'_{\gamma}$$

Proof. Fix $\sigma, \sigma' \in \mathcal{L}_U$ such that $|\mathcal{V}(\sigma \circ \gamma)|, |\mathcal{V}(\sigma' \circ \gamma)| \leq D_V$. Note that $|\mathcal{V}(\sigma)| - \frac{|V|}{2} + |\mathcal{V}(\sigma')| - \frac{|V'|}{2} + 2(|\mathcal{V}(\gamma)| - \frac{|U|+|V|}{2}) = |\mathcal{V}(\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T)|$. Using Definition 4.5, we can easily verify that $\lambda_{\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T} = T(\sigma) T(\sigma') S(\gamma)^2$. Therefore, $H^{-\gamma}_{id} = \left|\frac{\text{Aut}(U)}{\text{Aut}(V)}\right| S(\gamma)^2 H_{id} (\sigma, \sigma')$. Since $H'_{\gamma}(\sigma, \sigma') = H_{id} (\sigma, \sigma')$ whenever $|\mathcal{V}(\sigma \circ \gamma)|, |\mathcal{V}(\sigma' \circ \gamma)| \leq D_V$, this completes the proof.

Rearranging this gives Lemma 4.11

Lemma 4.11. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U, V}$,

$$\left|\frac{\text{Aut}(V)}{\text{Aut}(U)}\right| \cdot \frac{1}{S(\gamma)^2} H^{-\gamma}_{id} \leq H'_{\gamma}$$

5 Qualitative bounds for Tensor PCA

5.1 Pseudo-calibration

Definition 5.1 (Slack parameter). Define the slack parameter to be $\Delta = n^{-C_A} \delta$ for a constant $C_A > 0$.

We will pseudo-calibrate with respect the following pair of random and planted distributions which we denote $\nu$ and $\mu$ respectively.

- Random distribution: Sample $A$ from $\mathcal{N}(0, I_{|\nu|^k})$.

- Planted distribution: Let $\lambda, \Delta > 0$. Sample $u$ from $\{-\frac{1}{\sqrt{\Delta n}}, 0, \frac{1}{\sqrt{\Delta n}}\}^n$ where the values are taken with probabilites $\frac{\Delta}{2}$, $1 - \Delta$, $\frac{\Delta}{2}$ respectively. Then sample $B$ from $\mathcal{N}(0, I_{|\nu|^k})$. Set $A = B + \lambda u \otimes k$.

Let the Hermite polynomials be $h_0(x) = 1, h_1(x) = x, h_2(x) = x^2 - 1, \ldots$. For $a \in \mathbb{N}^{[n]^k}$ and variables $A_e$ for $e \in [n]^k$, define $h(A) := \prod_{e \in [n]^k} h_e(A_e)$. We will work with this Hermite basis.

Lemma 5.2. Let $1 \in \mathbb{N}^n, a \in \mathbb{N}_{[n]^k}$. For $i \in [n]$, let $d_i = \sum_{e \in [n]^k} a_e$. Let $c$ be the number of $i$ such that $l_i + d_i$ is nonzero. Then, if $l_i + d_i$ are all even, we have

$$\mathbb{E}_{\mu}[u^l h(A)] = \Delta^c \left(\frac{1}{\sqrt{\Delta n}}\right)^{|l|} \prod_{e \in [n]^k} \left(\frac{\lambda}{(\Delta n)^{\frac{1}{2}}}\right)^{a_e}$$

Else, $\mathbb{E}_{\mu}[u^l h(A)] = 0$.

27
Proof. When $A \sim \mu$, for all $e \in [n]^k$, we have $A_e = B_e + \lambda \prod_{i \leq k} u_{e_i}$, where $B_e \sim \mathcal{N}(0,1)$.

Let’s analyze when the required expectation is nonzero. We can first condition on $u$ and use the fact that for a fixed $t$, $\mathbb{E}_{g \sim \mathcal{N}(0,1)}[h_k(g + t)] = t^k$ to obtain

$$\mathbb{E}_{(u_1, u_2) \sim \mu}[u^t h_{a}(A)] = \mathbb{E}_{(u_1) \sim \mu} \left[ u^t \prod_{e \in [n]^k} (\lambda \prod_{i \leq k} u_{e_i}) \right] = \mathbb{E}_{(u_1) \sim \mu} \left[ \prod_{i \in [n]} u_{i}^{I_i + d_i} \right] \prod_{e \in [n]^k} \lambda^a_e$$

Observe that this is nonzero precisely when all $I_i + d_i$ are in, in which case

$$\mathbb{E}_{(u_1) \sim \mu} \left[ \prod_{i \in [n]} u_{i}^{I_i + d_i} \right] = \Delta^c \left( \frac{1}{\sqrt{\Delta n}} \right)^{\sum_{i \leq n} I_i + d_i} = \Delta^c \left( \frac{1}{\sqrt{\Delta n}} \right)^{|I|} \prod_{e \in [n]^k} \left( \frac{1}{(\Delta n)^{\frac{a_e}{2}}} \right)^{a_e}$$

where we used the fact that $\sum_{e \in [n]^k} a_e = k \sum_{i \in [n]} d_i$. This completes the proof.

Now, we can write the moment matrix in terms of graph matrices.

Definition 5.3. Define the degree of SoS to be $D_{\text{sos}} = n^{C_{\text{sos}}}$ for some constant $C_{\text{sos}} > 0$ that we choose later.

Definition 5.4 (Truncation parameters). Define the truncation parameters to be $D_V = n^{C_{V}}$, $D_E = n^{C_{E}}$ for some constants $C_V, C_E > 0$.

Remark 5.5 (Choice of parameters). We first set $\varepsilon$ to be a sufficiently small constant. Based on the choice of $\varepsilon$, we will set the constant $C_{\varepsilon}$ sufficiently small so that the planted distribution is well defined. Based on these choices, we will set $C_V, C_E$ to be sufficiently small constants to satisfy all the inequalities we use in our proof. Based on these choices, we can choose $C_{\text{sos}}$ to be sufficiently small to satisfy the inequalities we use.

Remark 5.6. The underlying graphs for the graph matrices have the following structure; There will be $n$ vertices of a single type and the edges will be ordered hyperedges of arity $k$.

Definition 5.7. For the analysis of Tensor PCA, we will use the following notation.

- For an index shape $U$ and a vertex $i$, define $\deg^U(i)$ as follows: If $i \in V(U)$, then it is the power of the unique index piece $\Lambda \in U$ such that $i \in V(\Lambda)$. Otherwise, it is 0.
- For an index shape $U$, define $\deg(U) = \sum_{i \in V(U)} \deg^U(i)$. This is also the degree of the monomial that $U$ corresponds to.
- For a shape $\alpha$ and vertex $i$ in $\alpha$, let $\deg^\alpha(i) = \sum_{e \in E(\alpha)} I_e$.
- For any shape $\alpha$, let $\deg(\alpha) = \deg(U_\alpha) + \deg(V_\alpha)$.

We will now describe the decomposition of the moment matrix $\Lambda$.

Definition 5.8. If a shape $\alpha$ satisfies the following properties:

- $\deg^\alpha(i) + \deg^U(i) + \deg^V(i)$ is even for all $i \in V(\alpha)$,
- $\alpha$ is proper;
- $\alpha$ satisfies the truncation parameters $D_{\text{sos}}, D_V, D_E$. 

28
then define
\[
\lambda_\alpha = \Delta^{|V(\alpha)|} \left( \frac{1}{\sqrt{\Delta n}} \right)^{\deg(\alpha)} \prod_{e \in E(\alpha)} \left( \frac{\lambda}{(\Delta n)^{\frac{1}{2}}} \right)^{l_e}.
\]
Otherwise, define \( \lambda_\alpha = 0 \).

**Corollary 5.9.** \( \Lambda = \sum \lambda_\alpha M_\alpha \).

### 5.2 Proving positivity - Qualitative bounds

We use the canonical definition of \( H'_\gamma \) from Section 3.7.4. In this section, we will prove the following qualitative bounds.

**Lemma 5.10.** For all \( U \in \mathcal{I}_{mid} \), \( H_{Id_U} \succeq 0 \)

We define the following quantity to capture the contribution of the vertices within \( \tau \) to the Fourier coefficients.

**Definition 5.11.** For \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), if \( \deg^\tau(i) \) is even for all vertices \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \), define
\[
S(\tau) = \Delta^{|V(\tau)|-|U_\tau|} \prod_{e \in E(\tau)} \left( \frac{\lambda}{(\Delta n)^{\frac{1}{2}}} \right)^{l_e}.
\]
Otherwise, define \( S(\tau) = 0 \).

**Lemma 5.12.** For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \),
\[
\begin{bmatrix}
S(\tau) / |\text{Aut}(U)| & H_{Id_U} \\
H^T_{\tau} & S(\tau) / |\text{Aut}(U)| \end{bmatrix} \succeq 0
\]

We define the following quantity to capture the contribution of the vertices within \( \gamma \) to the Fourier coefficients.

**Definition 5.13.** For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \) and \( \gamma \in \Gamma_{U,V} \), if \( \deg^\gamma(i) \) is even for all vertices \( i \) in \( V(\gamma) \setminus U_\gamma \setminus V_\gamma \), define
\[
S(\gamma) = \Delta^{|V(\gamma)|-|U_\gamma|} \prod_{e \in E(\gamma)} \left( \frac{\lambda}{(\Delta n)^{\frac{1}{2}}} \right)^{l_e}.
\]
Otherwise, define \( S(\gamma) = 0 \).

**Lemma 5.14.** For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \) and all \( \gamma \in \Gamma_{U,V} \),
\[
\frac{|\text{Aut}(V)|}{|\text{Aut}(U)|} \cdot \frac{1}{S(\gamma)^2} H_{Id_V} \gamma^\gamma \preceq H'_{\gamma}
\]

29
5.2.1 Proof of Lemma 5.10

When we compose shapes \( \sigma, \sigma' \), from Definition 5.8, observe that all vertices \( i \) in \( \lambda_{\sigma \circ \sigma'} \) should have \( \deg_{\sigma \circ \sigma'}(i) \) to be even, in order for \( \lambda_{\sigma \circ \sigma'} \) to be nonzero. To partially capture this notion conveniently, we will introduce the notion of parity vectors.

**Definition 5.15.** Define a parity vector \( \rho \) to be a vector whose entries are in \( \{0,1\} \).

**Definition 5.16.** For \( U \in I_{mid} \), define \( \mathcal{P}_U \) to be the set of parity vectors \( \rho \) whose coordinates are indexed by \( U \).

**Definition 5.17.** For a left shape \( \sigma \), define \( \rho_\sigma \in \mathcal{P}_{V_\sigma} \), called the parity vector of \( \sigma \), to be the parity vector such that for each vertex \( i \in V_\sigma \), the \( i \)-th entry of \( \rho_\sigma \) is the parity of \( \deg_{V_\sigma}(i) + \deg_{\sigma}(i) \), that is \( (\rho_\sigma)_i \equiv \deg_{V_\sigma}(i) + \deg_{\sigma}(i) \pmod{2} \).

**Definition 5.18.** For \( U \in I_{mid} \) and \( \rho \in \mathcal{P}_U \), let \( \mathcal{L}_{U, \rho} \) be the set of all left shapes \( \sigma \in \mathcal{L}_U \) such that \( \rho_\sigma = \rho \), that is, the set of all left shapes with parity vector \( \rho \).

**Definition 5.19.** For a shape \( \tau \), for a \( \tau \) coefficient matrix \( H_\tau \) and parity vectors \( \rho, \rho' \in \mathcal{P}_{V_\tau} \), define the \( \tau \)-coefficient matrix \( H_{\tau, \rho, \rho'} \) as \( H_{\tau, \rho, \rho'}(\sigma, \sigma') = H_\tau(\sigma, \sigma') \) if \( \sigma \in \mathcal{L}_{U_{\rho, \rho'}} \) and \( \sigma' \in \mathcal{L}_{V_{\rho, \rho'}} \) and 0 otherwise.

**Proposition 5.20.** For any shape \( \tau \) and \( \tau \)-coefficient matrix \( H_\tau \), \( H_\tau = \sum_{\rho, \rho' \in \mathcal{P}_{V_\tau}} H_{\tau, \rho, \rho'} \rho \rho' \).

**Proposition 5.21.** For any \( U \in I_{mid} \), \( H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_{U, \rho}} \).

**Proof.** For any \( \sigma, \sigma' \in \mathcal{L}_{U}, \) using **Definition 5.8**, note that in order for \( H_{Id_U}(\sigma, \sigma') \) to be nonzero, we must have \( \rho_{\sigma} = \rho_{\sigma'} \).

We define the following quantity to capture the contribution of the vertices within \( \sigma \) to the Fourier coefficients.

**Definition 5.22.** For a shape \( \sigma \in \mathcal{L} \), if \( \deg_{\sigma}(i) + \deg_{U_\sigma}(i) \) is even for all vertices \( i \in V(\sigma) \setminus V_\sigma \), define

\[
T(\sigma) = \Delta^{|V(\sigma)| - |V_\sigma|} \left( \frac{1}{\sqrt{\Delta n}} \right)^{\deg_{U_\sigma}} \prod_{e \in E(\sigma)} \left( \frac{\lambda_e}{(\Delta n)^2} \right)^{l_e}
\]

Otherwise, define \( T(\sigma) = 0 \).

**Definition 5.23.** For \( U \in I_{mid} \) and \( \rho \in \mathcal{P}_U \), define \( v_\rho \) to be the vector indexed by \( \sigma \in \mathcal{L} \) such that \( v_\rho(\sigma) \) is \( T(\sigma) \) if \( \sigma \in \mathcal{L}_{U, \rho} \) and 0 otherwise.

**Proposition 5.24.** For all \( U \in I_{mid} \), \( \rho \in \mathcal{P}_U \), \( H_{Id_{U, \rho}} = \frac{1}{|\text{Aut}(U)|} v_\rho v_\rho^T \).

**Proof.** This follows by verifying the conditions of **Definition 5.8**.

**Lemma 5.10.** For all \( U \in I_{mid} \), \( H_{Id_U} \succeq 0 \).

**Proof.** We have \( H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_{U, \rho}} = \frac{1}{|\text{Aut}(U)|} \sum_{\rho \in \mathcal{P}_U} v_\rho v_\rho^T \succeq 0 \).
5.2.2 Proof of Lemma 5.12

The next proposition captures the fact that when we compose shapes $\sigma, \tau, \sigma'^T$, in order for $\lambda_{\sigma \tau \sigma'^T}$ to be nonzero, the parities of the degrees of the merged vertices should add up correspondingly.

**Proposition 5.25.** For all $U \in I_{mid}$ and $\tau \in M_U$, there exist two sets of parity vectors $P_\tau, Q_\tau \subseteq P_U$ and a bijection $\pi : P_\tau \rightarrow Q_\tau$ such that $H_\tau = \sum_{\rho \in P_\tau} H_{\tau, \rho, \pi(\rho)}$.

**Proof.** Using Definition 5.8, in order for $H_\tau(\sigma, \sigma')$ to be nonzero, in $\sigma \circ \tau \circ \sigma'$, we must have that for all $i \in U_\tau \cup V_\tau$, $deg_{H_\tau}(i) + deg_{H_{\sigma'}}(i) + deg_{H_{\sigma'^T}}(i)$ must be even. In other words, for any $\rho \in P_U$, there is at most one $\rho' \in P_U$ such that if we take $\sigma \in L_{U, \rho, \rho'}$, $\sigma' \in L_U$ with $H_\tau(\sigma, \sigma')$ nonzero, then the parity of $\sigma'$ is $\rho'$. Also, observe that $\rho'$ determines $\rho$. We then take $P_\tau$ to be the set of $\rho$ such that $\rho'$ exists, $Q_\tau$ to be the set of $\rho'$ and in this case, we define $\pi(\rho) = \rho'$.

We restate Definition 5.11 for convenience.

**Definition 5.11.** For $U \in I_{mid}$ and $\tau \in M_U$, if $deg_\tau(i)$ is even for all vertices $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, define

$$S(\tau) = \Delta^{|V(\tau)| - |U_\tau|} \prod_{e \in E(\tau)} \left( \frac{\lambda}{(\Delta n)^2} \right)^l_e$$

Otherwise, define $S(\tau) = 0$.

**Proposition 5.26.** For any $U \in I_{mid}$ and $\tau \in M_U$, suppose we take $\rho \in P_\tau$. Let $\pi$ be the bijection from Proposition 5.25 so that $\pi(\rho) \in Q_\tau$. Then, $H_{\tau, \rho, \pi(\rho)} = \frac{1}{|Aut(U)|} S(\tau) v_{\rho, \pi(\rho)}^T$.

**Proof.** This follows by a straightforward verification of the conditions of Definition 5.8.

**Lemma 5.12.** For all $U \in I_{mid}$ and $\tau \in M_U$,

$$\begin{bmatrix} S(\tau) & H_\tau \\ \frac{|Aut(U)|}{H_{Id_U}} & \frac{H_{beta}}{|Aut(U)|} \end{bmatrix} \succeq 0$$

**Proof.** Let $P_\tau, Q_\tau, \pi$ be from Proposition 5.25. For $\rho, \rho' \in P_U$, let $W_{\rho, \rho'} = v_{\rho}(v_{\rho'})^T$. Then, $H_{Id_U} = \sum_{\rho \in P_U} H_{Id_U, \rho, \rho'} = \frac{1}{|Aut(U)|} \sum_{\rho \in P_U} W_{\rho, \rho'}$ and $H_\tau = \sum_{\rho \in P_\tau} H_{\tau, \rho, \pi(\rho)} = \frac{1}{|Aut(U)|} S(\tau) \sum_{\rho \in P_\tau} W_{\rho, \pi(\rho)}$. We have

$$\begin{bmatrix} S(\tau) & H_\tau \\ \frac{|Aut(U)|}{H_{beta}} & \frac{H_{beta}}{|Aut(U)|} \end{bmatrix} = \frac{S(\tau)}{|Aut(U)|^2} \begin{bmatrix} \sum_{\rho \in P_U} W_{\rho, \rho} & \sum_{\rho \in P_\tau} W_{\rho, \pi(\rho)} \\ \sum_{\rho \in P_\tau} W_{\rho, \pi(\rho)} & \sum_{\rho \in P_U} W_{\rho, \rho} \end{bmatrix}$$
Since \( \frac{S(\tau)}{|\text{Aut}(U)|} \geq 0 \), it suffices to prove that \[
\begin{bmatrix}
\sum_{\rho \in P_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)} \\
\sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)} & \sum_{\rho \in P_U} W_{\rho,\rho}
\end{bmatrix} \succeq 0.
\]
Consider
\[
\begin{bmatrix}
\sum_{\rho \in P_U \setminus P_\tau} W_{\rho,\rho} & \sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)} \\
\sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)} & \sum_{\rho \in P_U \setminus Q, \rho} W_{\rho,\rho}
\end{bmatrix} \succeq 0.
\]

We have \( \sum_{\rho \in P_U \setminus P_\tau} W_{\rho,\rho} = \sum_{\rho \in P_U \setminus P_\tau} v_\rho v_\rho^T \succeq 0 \). Similarly, \( \sum_{\rho \in P_U \setminus Q, \rho} W_{\rho,\rho} \succeq 0 \) and so, the first term in the above expression, \[
\begin{bmatrix}
\sum_{\rho \in P_U \setminus P_\tau} W_{\rho,\rho} & 0 \\
0 & \sum_{\rho \in P_U \setminus Q, \rho} W_{\rho,\rho}
\end{bmatrix}
\]
is positive semidefinite. For the second term,
\[
\begin{bmatrix}
\sum_{\rho \in P_\tau} W_{\rho,\rho} & \sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)} \\
\sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)} & \sum_{\rho \in P_\tau} W_{\pi(\rho),\pi(\rho)}
\end{bmatrix} = \sum_{\rho \in P_\tau} \begin{bmatrix}
W_{\rho,\rho} & W_{\rho,\pi(\rho)} \\
W_{\rho,\pi(\rho)} & W_{\pi(\rho),\pi(\rho)}
\end{bmatrix}
\]
\[
= \sum_{\rho \in P_\tau} \begin{bmatrix}
v_\rho v_\rho^T \\
v_\pi(\rho) v_\pi(\rho)^T
\end{bmatrix}
\]
\[
\geq 0.
\]

### 5.2.3 Proof of Lemma 5.14

The next proposition captures the fact that when we compose shapes \( \sigma, \gamma, \gamma^T, \sigma^T \), in order for \( \lambda_{\sigma \otimes \gamma \otimes \gamma^T \otimes \sigma^T} \) to be nonzero, the parities of the degrees of the merged vertices should add up correspondingly.

**Definition 5.27.** For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \), for \( \gamma \in \Gamma_{U,V} \) and parity vectors \( \rho, \rho' \in P_U \), define the \( \gamma \circ \gamma^T \)-coefficient matrix \( H_{\gamma \circ \gamma^T}^{\gamma} \) as \( H_{\gamma \circ \gamma^T}^{\gamma}(\sigma, \sigma') = H_{\gamma \circ \gamma^T}^{\gamma}(\sigma, \sigma') \) if \( \sigma \in \mathcal{L}_{U, \rho}, \sigma' \in \mathcal{L}_{U, \rho'} \) and \( 0 \) otherwise.

**Proposition 5.28.** For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \), for all \( \gamma \in \Gamma_{U,V} \), there exists a set of parity vectors \( P_\gamma \subseteq P_U \) such that
\[
H_{\gamma \circ \gamma^T}^{\gamma} = \sum_{\rho \in P_\gamma} H_{\gamma \circ \gamma^T}^{\gamma}(\sigma, \sigma')
\]

**Proof.** Take any \( \rho \in P_U \). For \( \sigma \in \mathcal{L}_{U, \rho}, \sigma' \in \mathcal{L}_{U, \rho} \), since \( H_{\gamma \circ \gamma^T}^{\gamma}(\sigma, \sigma') = \frac{\lambda_{\sigma \otimes \gamma \otimes \gamma^T \otimes \sigma^T}}{|\text{Aut}(V)|} \), \( H_{\gamma \circ \gamma^T}^{\gamma}(\sigma, \sigma') \) is nonzero precisely when \( \lambda_{\sigma \otimes \gamma \otimes \gamma^T \otimes \sigma^T} \) is nonzero. For this quantity to be nonzero, using Definition 5.8, we get that it is necessary, but not sufficient, that the parity vector of \( \sigma' \) must also be \( \rho \). And also observe that there exists a set \( P_\gamma \) of parity vectors \( \rho \) for which \( H_{\gamma \circ \gamma^T}^{\gamma}(\sigma, \sigma') \) is nonzero and their sum is precisely \( H_{\gamma \circ \gamma^T}^{\gamma} \). □
**Definition 5.29.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$, for all $\gamma \in \Gamma_{U,V}$ and parity vector $\rho \in \mathcal{P}_U$, define the matrix $H'_{\rho_{U,V},\rho}$ as $H'_{\rho_{U,V},\rho}(\sigma, \sigma') = H'_{\gamma}(\sigma, \sigma')$ if $\sigma, \sigma' \in \mathcal{L}_{U,V}$ and 0 otherwise.

**Proposition 5.30.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$, for $\gamma \in \Gamma_{U,V}$, $H'_{\gamma} = \sum_{\rho \in \mathcal{P}_U} H'_{\rho_{U,V},\rho}$.

We restate **Definition 5.13** for convenience.

**Definition 5.13.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$ and $\gamma \in \Gamma_{U,V}$, if $\deg^\gamma(i)$ is even for all vertices $i$ in $V(\gamma) \setminus U_\gamma \setminus V_\gamma$, define

$$S(\gamma) = \Delta|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2} \prod_{e \in E(\gamma)} \left( \frac{\lambda}{(\Delta n)^2} \right)^{t_e}$$

Otherwise, define $S(\gamma) = 0$.

**Proposition 5.31.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$, for all $\gamma \in \Gamma_{U,V}$ and $\rho \in \mathcal{P}_U$,

$$H_{\rho_{U,V},\rho}^{-\gamma} = \frac{|\text{Aut}(U)|}{|\text{Aut}(V)|} S(\gamma)^2 H'_{\rho_{U,V},\rho}$$

**Proof.** Fix $\sigma, \sigma' \in \mathcal{L}_{U,V}$ such that $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$. Note that $|V(\sigma)| - \frac{|V_\gamma|}{2} + |V(\sigma')| - \frac{|V_{\gamma'}|}{2} + 2(|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}) = |V(\sigma \circ \gamma \circ \gamma' \circ \sigma'')|$. Using **Definition 5.8**, we can easily verify that $\lambda_{\sigma \circ \gamma \circ \gamma' \circ \sigma''} = T(\sigma)T(\sigma')S(\gamma)^2$. Therefore, $H_{\rho_{U,V},\rho}^{-\gamma}(\sigma, \sigma') = \frac{|\text{Aut}(U)|}{|\text{Aut}(V)|} S(\gamma)^2 H_{\rho_{U,V},\rho}(\sigma, \sigma')$. Since $H'_{\rho_{U,V},\rho}(\sigma, \sigma') = H_{\rho_{U,V},\rho}(\sigma, \sigma')$ whenever $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$, this completes the proof. ■

**Lemma 5.14.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,

$$\frac{|\text{Aut}(V)|}{|\text{Aut}(U)|} \cdot \frac{1}{S(\gamma)^2} H_{\rho_{U,V}}^{-\gamma} \preceq H'_{\gamma}$$

**Proof.** We have

$$\frac{|\text{Aut}(V)|}{|\text{Aut}(U)|} \cdot \frac{1}{S(\gamma)^2} H_{\rho_{U,V}}^{-\gamma} = \sum_{\rho \in \mathcal{P}_U} \frac{|\text{Aut}(V)|}{|\text{Aut}(U)|} \cdot \frac{1}{S(\gamma)^2} H'_{\rho_{U,V},\rho} = \sum_{\rho \in \mathcal{P}_U} H'_{\rho_{U,V},\rho} \preceq \sum_{\rho \in \mathcal{P}_U} H'_{\rho_{U,V},\rho} = H'_{\gamma},$$

where we used the fact that for all $\rho \in \mathcal{P}_U$, we have $H'_{\rho_{U,V},\rho} \geq 0$ which can be proved the same way as the proof of **Lemma 5.10**. ■
6 Qualitative bounds for Sparse PCA

6.1 Pseudo-calibration

Definition 6.1 (Slack parameter). Define the slack parameter to be $\Delta = d^{-c_\Delta}$ for a constant $C_\Delta > 0$.

We will pseudo-calibrate with respect the following pair of random and planted distributions which we denote $v$ and $\mu$ respectively.

- Random distribution: $v_1, \ldots, v_m$ are sampled from $\mathcal{N}(0, I_d)$ and we take $S$ to be the $m \times d$ matrix with rows $v_1, \ldots, v_m$.
- Planted distribution: Sample $u$ from $\{-\frac{1}{\sqrt{k}}, \frac{1}{\sqrt{k}}\}^d$ where the values are taken with probabilities $\frac{k}{2d}$, $\frac{1 - k}{2}$ respectively. Then sample $v_1, \ldots, v_m$ as follows. For each $i \in [m]$, with probability $\Delta$, sample $v_i$ from $\mathcal{N}(0, I_d + \lambda uu^T)$ and with probability $1 - \Delta$, sample $v_i$ from $\mathcal{N}(0, I_d)$. Finally, take $S$ to be the $m \times d$ matrix with rows $v_1, \ldots, v_m$.

We will again work with the Hermite basis of polynomials. For $a \in \mathbb{N}^{m \times d}$ and variables $v_{ij}$ for $i \in [m], j \in [n]$, define $h_a(v) := \prod_{i \in [m], j \in [n]} h_{a_{ij}}(v_{ij})$.

Definition 6.2. For a nonnegative integer $t$, define $t!!$ as $\begin{cases} \frac{(2t)!}{t!2^t} = 1 \times 3 \times \ldots \times t, & \text{if } t \text{ is odd} \\ 0, & \text{otherwise} \end{cases}$

Lemma 6.3. Let $I \in \mathbb{N}^d, a \in \mathbb{N}^{m \times d}$. For $i \in [m]$, let $e_i = \sum_{j \in [d]} a_{ij}$ and for $j \in [d]$, let $f_j = I_j + \sum_{i \in [m]} a_{ij}$. Let $c_1$ (resp. $c_2$) be the number of $i$ (resp. $j$) such that $e_i > 0$ (resp. $f_j > 0$). Then, if $e_i, f_j$ are all even, we have

$$\mathbb{E}_v[u^T h_a(v)] = \left( \frac{1}{\sqrt{k}} \right)^{|I|} \left( \frac{k}{d} \right)^{c_2} \Delta^{c_1} \prod_{i \in [m]} (e_i - 1)!! \prod_{i,j} \sqrt{\lambda_{ij}}$$

Else, $\mathbb{E}_v[u^T h_a(v)] = 0$.

Proof. $v_1, \ldots, v_m \sim \mu$ can be written as $v_i = g_i + \sqrt{\lambda} \sqrt{b_i} u$ where $g_i \sim \mathcal{N}(0, I_d), l_i \sim \mathcal{N}(0, 1), b_i \in \{0,1\}$ where $b_i = 1$ with probability $\Delta$.

Let’s analyze when the required expectation is nonzero. We can first condition on $b_i, l_i, u$ and use the fact that for a fixed $t$, $\mathbb{E}_{g \sim \mathcal{N}(0,1)}[h_k(g + t)] = t^k$ to obtain

$$\mathbb{E}_{(u,l,b_i,g)}[u^T h_a(v)] = \mathbb{E}_{(u,l,b_i)}[u^T \prod_{i,j} (\sqrt{\lambda} \sqrt{b_i} l_i u_j)^{a_{ij}}] = \mathbb{E}_{(u,l,b_i)}[\prod_{i \in [m]} (b_i l_i)^{c_i} \prod_{j \in [d]} u_j^{f_j}] \prod_{i,j} \sqrt{\lambda_{ij}}$$

For this to be nonzero, the set of $c_1$ indices $i$ such that $e_i > 0$, should not have been resampled otherwise $b_i = 0$, each of which happens independently with probability $\Delta$. And the set of $c_2$ indices $j$ such that $f_j > 0$ should have been such that $u_j$ is nonzero, each of which happens independently with probability $\frac{k}{d}$. Since $l_i, u_j$ are have zero expectation in $v$, we need $e_i, f_j$ to be even. The expectation then becomes

$$\Delta^{c_1} \left( \frac{k}{d} \right)^{c_2} \mathbb{E}_{(u,l,b_i)}[\prod_{i \in [m]} l_i^{c_i} \prod_{j \in [d]} u_j^{f_j}] \prod_{i,j} \sqrt{\lambda_{ij}} = \left( \frac{1}{\sqrt{k}} \right)^{|I|} \left( \frac{k}{d} \right)^{c_2} \Delta^{c_1} \prod_{i \in [m]} (e_i - 1)!! \prod_{i,j} \sqrt{\lambda_{ij}}$$
The last equality follows because, for each \( j \) such that \( u_j \) is nonzero, we have \( u_j^t = \left( \frac{1}{\sqrt{r}} \right)^t \) and \( \mathbb{E}_{Z \sim \mathcal{N}(0,1)}[g^t] = (t - 1)!! \) if \( t \) is even.

Now, we can write the moment matrix in terms of graph matrices.

**Definition 6.4.** Define the degree of SoS to be \( D_{\text{sos}} = d^{C_{\text{sos}}} \) for some constant \( C_{\text{sos}} > 0 \) that we choose later.

**Definition 6.5** (Truncation parameters). Define the truncation parameters to be \( D_V = d^{C_V}, D_E = d^{C_E} \) for some constants \( C_V, C_E > 0 \).

**Remark 6.6** (Choice of parameters). We first set \( \varepsilon > 0 \) to be a sufficiently small constant. Based on the choice of \( \varepsilon \), we will set the constant \( C_\Delta > 0 \) sufficiently small so that the planted distribution is well defined. Based on these choices, we will set \( C_V, C_E \) to be sufficiently small constants to satisfy all the inequalities we use in our proof. Based on these choices, we can choose \( C_{\text{sos}} \) to be sufficiently small to satisfy the inequalities we use.

**Remark 6.7.** The underlying graphs for the graph matrices have the following structure: There will be two types of vertices - \( d \) type 1 vertices corresponding to the dimensions of the space and \( m \) type 2 vertices corresponding to the different input vectors. The shapes will correspond to bipartite graphs with edges going between across of different types.

**Definition 6.8.** For the analysis of Sparse PCA, we will use the following notation.

- For a shape \( \alpha \) and type \( t \in \{1, 2\} \), let \( V_t(\alpha) \) denote the vertices of \( V(\alpha) \) that are of type \( t \). Let \( |\alpha|_t = |V_t(\alpha)| \).
- For an index shape \( U \) and a vertex \( i \), define \( \deg^U(i) \) as follows: If \( i \in V(U) \), then it is the power of the unique index shape piece \( A \in U \) such that \( i \in V(A) \). Otherwise, it is 0.
- For an index shape \( U \), define \( \deg(U) = \sum_{i \in V(U)} \deg^U(i) \). This is also the degree of the monomial \( p_U \).
- For a shape \( \alpha \) and vertex \( i \) in \( \alpha \), let \( \deg^\alpha(i) = \sum_{e \in E(\alpha)} |\alpha|_e \).
- For any shape \( \alpha \), let \( \deg(\alpha) = \deg(U_\alpha) + \deg(V_\alpha) \).
- For an index shape \( U \in \mathcal{I}_{\text{mid}} \) and type \( t \in \{1, 2\} \), let \( U_t \in U \) denote the index shape piece of type \( t \) in \( U \) if it exists, otherwise define \( U_t \) to be \( \emptyset \). Note that this is well defined since for each type \( t \), there is at most one index shape piece of type \( t \) in \( U \) since \( U \in \mathcal{I}_{\text{mid}} \). Also, denote by \( |U|_t \) the length of the tuple \( U_t \).

We will now describe the decomposition of the moment matrix \( \Lambda \).

**Definition 6.9.** If a shape \( \alpha \) satisfies the following properties:

- Both \( U_\alpha \) and \( V_\alpha \) only contain index shape pieces of type 1,
- \( \deg^\alpha(i) + \deg^{U_\alpha}(i) + \deg^{V_\alpha}(i) \) is even for all \( i \in V(\alpha) \),
- \( \alpha \) is proper,
- \( \alpha \) satisfies the truncation parameters \( D_{\text{sos}}, D_V, D_E \).
then define

\[
\lambda_\alpha = \left( \frac{1}{\sqrt{k}} \right)^{\deg(a)} \left( \frac{k}{d} \right)^{|\alpha|_1} \Delta^{|\alpha|_2} \prod_{j \in V_2(\alpha)} (\deg^a(j) - 1)!! \prod_{e \in E(\alpha)} \sqrt{\lambda_{le}^e} / \sqrt{k^e}.
\]

Otherwise, define \( \lambda_\alpha = 0 \).

**Corollary 6.10.** \( \Lambda = \sum \lambda_\alpha M_\alpha \).

### 6.2 Proving positivity - Qualitative bounds

We use the canonical definition of \( H'_\gamma \) from **Section 3.7.4**. In this section, we will prove the following qualitative bounds.

**Lemma 6.11.** For all \( U \in \mathcal{I}_{mid} \), \( H_{IdU} \succeq 0 \)

For technical reasons, it will be convenient to discretize the Normal distribution. The following fact follows from standard results on Gaussian quadrature, see e.g. [DKS17, Lemma 4.3].

**Fact 6.12 (Discretizing the Normal distribution).** There is an absolute constant \( C_{disc} \) such that, for any positive integer \( D \), there exists a distribution \( E \) over the real numbers supported on \( D \) points \( p_1, \ldots, p_D \), such that

- \( |p_i| \leq C_{disc} \sqrt{D} \) for all \( i \leq D \) and
- \( E_{g \sim E}[g^t] = E_{g \sim N(0,1)}[g^t] \) for all \( t = 0, 1, \ldots, 2D - 1 \)

We define the following quantity to capture the contribution of the vertices within \( \tau \) to the Fourier coefficients.

**Definition 6.13.** For \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), if \( \deg^\tau(i) \) is even for all vertices \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \), define

\[
S(\tau) = \left( \frac{k}{d} \right)^{|\tau_1| - |U_\tau|_1} \Delta^{|\tau_2| - |U_\tau|_2} \prod_{j \in V_2(\tau) \setminus U_\tau \setminus V_\tau} (\deg^\tau(j) - 1)!! \prod_{e \in E(\tau)} \sqrt{\lambda_{le}^e} / \sqrt{k^e}.
\]

Otherwise, define \( S(\tau) = 0 \).

**Definition 6.14.** For any shape \( \tau \), suppose \( U'_\tau = (U_\tau)_2, V'_\tau = (V_\tau)_2 \) are the type 2 vertices in \( U_\tau, V_\tau \) respectively. Define

\[
R(\tau) = C_{disc} \sqrt{D_E} \sum_{u \in U'_\tau, v \in V'_\tau} \deg^\tau(j)
\]

where \( C_{disc} \) is the constant from **Fact 6.12**.

**Lemma 6.15.** For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \),

\[
\begin{bmatrix}
\frac{S(\tau)R(\tau)}{|Aut(\tau)|} H_{IdU} & \frac{H_\tau}{|Aut(\tau)|} \\
\frac{H_\tau}{|Aut(\tau)|} & \frac{S(\tau)R(\tau)}{|Aut(\tau)|} H_{IdU}
\end{bmatrix} \succeq 0.
\]

We define the following quantity to capture the contribution of the vertices within \( \gamma \) to the Fourier coefficients.
**Definition 6.16.** For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and $\gamma \in \Gamma_{U,V}$, if $\deg^{\gamma}(i)$ is even for all vertices $i$ in $V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma}$, define

$$S(\gamma) = \left(\frac{k}{d}\right)^{|\gamma|} \Delta^{\gamma} \prod_{j \in V_{2}(\gamma) \setminus U_{\gamma}} (\deg^{\gamma}(j) - 1)! \prod_{e \in E(\gamma)} \frac{\sqrt{\lambda_{e}}}{\sqrt{k}}.$$

Otherwise, define $S(\gamma) = 0$.

**Lemma 6.17.** For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,

$$\frac{|Aut(V)|}{|Aut(U)|} \frac{1}{S(\gamma)^{2}R(\gamma)^{2}} H_{ld}^{-\gamma,\gamma} \leq H'_{\gamma}.$$

### 6.2.1 Proof of Lemma 6.11

When we compose shapes $\sigma, \sigma'$, from **Definition 6.9**, observe that all vertices $i$ in $\lambda_{\sigma \circ \sigma'}$ should have $\deg^{\sigma \circ \sigma'}(i) + \deg^{U_{\sigma \circ \sigma'}}(i) + \deg^{V_{\sigma \circ \sigma'}}(i)$ to be even, in order for $\lambda_{\sigma \circ \sigma'}$ to be nonzero. To partially capture this notion conveniently, we will introduce the notion of parity vectors.

**Definition 6.18.** Define a parity vector $\rho$ to be a vector whose entries are in \{0,1\}.

**Definition 6.19.** For $U \in \mathcal{I}_{mid}$, define $\mathcal{P}_{U}$ to be the set of parity vectors $\rho$ whose coordinates are indexed by $U_{1}$ followed by $U_{2}$.

**Definition 6.20.** For a left shape $\sigma$, define $\rho_{\sigma} \in \mathcal{P}_{V_{\sigma}}$, called the parity vector of $\sigma$, to be the parity vector such that for each vertex $i \in V_{\sigma}$, the $i$-th entry of $\rho_{\sigma}$ is the parity of $\deg^{U_{\sigma}}(i) + \deg^{U_{\sigma}}(i)$, that is, $(\rho_{\sigma})_{i} \equiv \deg^{U_{\sigma}}(i) + \deg^{U_{\sigma}}(i) \pmod{2}$.

**Definition 6.21.** For $U \in \mathcal{I}_{mid}$ and $\rho \in \mathcal{P}_{U}$, let $\mathcal{L}_{U,\rho}$ be the set of all left shapes $\sigma \in \mathcal{L}_{U}$ such that $\rho_{\sigma} = \rho$, that is, the set of all left shapes with parity vector $\rho$.

**Definition 6.22.** For a shape $\tau$, for a $\tau$ coefficient matrix $H_{\tau}$ and parity vectors $\rho \in \mathcal{P}_{U_{\tau}}, \rho' \in \mathcal{P}_{V_{\tau}}$, define the $\tau$-coefficient matrix $H_{\tau,\rho,\rho'}$ as $H_{\tau,\rho,\rho'}(\sigma, \sigma') \equiv H_{\tau}(\sigma, \sigma')$ if $\sigma \in \mathcal{L}_{U_{\tau},\rho}, \sigma' \in \mathcal{L}_{V_{\tau},\rho'}$ and 0 otherwise.

**Proposition 6.23.** For any shape $\tau$ and $\tau$-coefficient matrix $H_{\tau}$, $H_{\tau} = \sum_{\rho \in \mathcal{P}_{U_{\tau}}} \sum_{\rho' \in \mathcal{P}_{V_{\tau}}} H_{\tau,\rho,\rho'}$

**Proposition 6.24.** For any $U \in \mathcal{I}_{mid}$, $H_{ldU} = \sum_{\rho \in \mathcal{P}_{U}} H_{ldU,\rho,\rho}$

**Proof.** For any $\sigma, \sigma' \in \mathcal{L}_{U}$, using **Definition 6.9**, note that in order for $H_{ldU}(\sigma, \sigma')$ to be nonzero, we must have $\rho_{\sigma} = \rho_{\sigma'}$. 

We will now discretize the normal distribution while matching the first $2D_{E} - 1$ moments.

**Definition 6.25.** Let $\mathcal{D}$ be a distribution over the real numbers obtained by setting $D = D_{E}$ in **Fact 6.12**. So, in particular, for any $x$ sampled from $\mathcal{D}$, we have $|x| \leq C_{disc} \sqrt{D_{E}}$ and for $t \leq 2D_{E} - 1$, $E_{x \sim \mathcal{D}}[x^{t}] = (t - 1)!$.

We define the following quantity to capture the contribution of the vertices within $\sigma$ to the Fourier coefficients.
Definition 6.26. For a shape \( \sigma \in \mathcal{L} \), if \( \deg^\tau(i) + \deg^{U_2}(i) \) is even for all vertices \( i \in V(\sigma) \setminus V_\sigma \), define

\[
T(\sigma) = \left( \frac{1}{\sqrt{k}} \right)^{\deg(U_2)} \left( k \right)^{|\tau_1| - |U_1|} \left( \frac{1}{2} \right)^{|\tau_2| - |U_2|} \prod_{j \in V_2(\sigma) \setminus V_\sigma} (\deg^\tau(j) - 1)! \prod_{e \in E(\tau)} \sqrt{\lambda_e} \prod_{e \in E(\sigma)} \sqrt{k^e} .
\]

Otherwise, define \( T(\sigma) = 0 \).

Definition 6.27. Let \( U \in \mathcal{I}_{mid} \). Let \( x_i \) for \( i \in U_2 \) be variables. Denote them collectively as \( x_{U_2} \). For \( \rho \in \mathcal{P}_U \), define \( v_{\rho,x_{U_2}} \) to be the vector indexed by left shapes \( \sigma \in \mathcal{L} \) such that the \( \sigma \)th entry is \( T(\sigma) \prod_{i \in U_2} x_i^{\deg(i)} \) if \( \sigma \in \mathcal{L}_{U,\rho} \) and 0 otherwise.

Proposition 6.28. For any \( U \in \mathcal{I}_{mid} \), \( \rho \in \mathcal{P}_U \), suppose \( x_i \) for \( i \in U_2 \) are random variables sampled from \( D \). Then,

\[
H_{Id_{U,\rho,\rho}} = \frac{1}{|\text{Aut}(U)|} \mathbb{E}_{x} [v_{\rho,x_{U_2}} v_{\rho,x_{U_2}}^T] .
\]

Proof. Observe that for \( \sigma, \sigma' \in \mathcal{L}_{U,\rho} \) and \( t \in \{1, 2\} \), \( (|\sigma|_t - |U_1|_t) + (|\sigma'|_t - |U_2|_t) = |\sigma \circ \sigma'|_t \). The result follows by verifying the conditions of Definition 6.9 and using Definition 6.25.

Lemma 6.11. For all \( U \in \mathcal{I}_{mid} \), \( H_{Id_{U}} \geq 0 \).

Proof. We have \( H_{Id_{U}} = \sum_{\rho \in \mathcal{P}_U} H_{Id_{U,\rho,\rho}} = \frac{1}{|\text{Aut}(U)|} \mathbb{E}_{x} [v_{\rho,x_{U_2}} v_{\rho,x_{U_2}}^T] \geq 0 \).

6.2.2 Proof of Lemma 6.15

The next proposition captures the fact that when we compose shapes \( \sigma, \tau, \sigma'^T \), in order for \( \lambda_{\sigma \circ \tau \circ \sigma'^T} \) to be nonzero, the parities of the degrees of the merged vertices should add up correspondingly.

Proposition 6.29. For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), there exist two sets of parity vectors \( P_\tau, Q_\tau \subseteq \mathcal{P}_U \) and a bijection \( \pi : P_\tau \rightarrow Q_\tau \) such that \( H_\tau = \sum_{\rho \in P_\tau} H_{\tau,\rho,\pi(\rho)} \).

Proof. Using Definition 6.9, in order for \( H_\tau(\sigma, \sigma') \) to be nonzero, we must have that, in \( \sigma \circ \tau \circ \sigma' \), for all \( i \in U_\tau \cup V_\tau \), \( \deg_{U_\tau}(i) + \deg_{U_\tau'}(i) + \deg_{\tau \circ \sigma \circ \tau'^T}(i) \) must be even. In other words, for any \( \rho \in \mathcal{P}_U \), there is at most one \( \rho' \in \mathcal{P}_U \) such that if we take \( \sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_{U_2} \) with \( H_\tau(\sigma, \sigma') \) nonzero, then the parity of \( \sigma' \) is \( \rho' \). Also, observe that \( \rho' \) determines \( \rho \). We then take \( P_\tau \) to be the set of \( \rho \) such that \( \rho' \) exists, \( Q_\tau \) to be the set of \( \rho' \) and in this case, we define \( \pi(\rho) = \rho' \).

We restate Definition 6.13 for convenience.

Definition 6.13. For \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), if \( \deg^\tau(i) \) is even for all vertices \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \), define

\[
S(\tau) = \left( \frac{k}{d} \right)^{|\tau_1| - |U_1|} \left( k \right)^{|\tau_2| - |U_2|} \prod_{j \in V_2(\tau) \setminus U_\tau \setminus V_\tau} (\deg^\tau(j) - 1)! \prod_{e \in E(\tau)} \sqrt{\lambda_e} \prod_{e \in E(\tau)} \sqrt{k^e} .
\]

Otherwise, define \( S(\tau) = 0 \).
Proposition 6.30. For any $U \in \mathcal{I}_{\text{mid}}$ and $\tau \in \mathcal{M}_U$, suppose we take $\rho \in P_{\tau}$. Let $\pi$ be the bijection from Proposition 6.29 so that $\pi(\rho) \in Q_{\tau}$. Let $U' = (U_{\tau})_2, V' = (V_{\tau})_2$ be the type 2 vertices in $U_{\tau}, V_{\tau}$ respectively. Let $x_i$ for $i \in U' \cup V'$ be random variables independently sampled from $D$. Define $x_{U'}$ (resp. $x_{V'}$) to be the subset of variables $x_i$ for $i \in U'$ (resp. $i \in V'$). Then,

$$H_{\tau, \rho, \pi(\rho)} = \frac{1}{|Aut(U)|^2} \mathbf{E}[\mathbf{S}(\tau) \mathbf{x}_\rho \mathbf{x}_{U'} \left( \prod_{i \in U' \cup V'} x_i^{deg(i)} \right) \mathbf{v}_{\pi(\rho), x_V'}^T]$$

Proof. For $\sigma \in L_{U, \rho}, \sigma' \in L_{U, \pi(\rho)}$ and $t \in \{1, 2\}$, we have $|\tau|_1 - |U_t|_t + (|\sigma|_1 - \frac{|V_{\tau}|}{2}) + (|\sigma'|_1 - \frac{|V_{\tau}|}{2}) = |\sigma \circ \tau \circ \sigma'|_t$. The result then follows by a straightforward verification of the conditions of Definition 6.9 using Definition 6.25.

Lemma 6.15. For all $U \in \mathcal{I}_{\text{mid}}$ and $\tau \in \mathcal{M}_U$,

$$\begin{bmatrix} S(\tau) R(\tau) \mathbf{H}_{Id_U} & \mathbf{H}_\tau \mathbf{H}_{Id_U}^T \\ \mathbf{H}_\tau^T & \mathbf{H}_{Id_U} \end{bmatrix} \geq 0$$

Proof. Let $P_{\tau}, Q_{\tau}, \pi$ be from Proposition 6.29. Let $U' = (U_{\tau})_2, V' = (V_{\tau})_2$ be the type 2 vertices in $U_{\tau}, V_{\tau}$ respectively. Let $x_i$ for $i \in U' \cup V'$ be random variables independently sampled from $D$. Define $x_{U'}$ (resp. $x_{V'}$) to be the subset of variables $x_i$ for $i \in U'$ (resp. $i \in V'$).

For $\rho \in P_{U}$, define $W_{\rho, \rho} = \mathbf{E}[y_{U_r} \mathbf{D}_U \mathbf{V}_{U_{\tau}} \mathbf{V}_{U_{\tau}}^T]$, so that $H_{Id_U, \rho, \rho} = \frac{1}{|Aut(U)|^2} W_{\rho, \rho}$. Observe that $W_{\rho, \rho} = \mathbf{E}[\mathbf{V}_{\rho, x_{U'}} \mathbf{V}_{\rho, x_{U'}}^T] = \mathbf{E}[\mathbf{V}_{\rho, x_{V'}} \mathbf{V}_{\rho, x_{V'}}^T]$ because $x_{U'}$ and $x_{V'}$ are also sets of variables sampled from $D$ and $U', V'$ have the same size as $U_2$ because $U_\pi = V = U$.

For $\rho, \rho' \in P_{U}$, define $Y_{\rho, \rho'} = \mathbf{E}[\mathbf{V}_{\rho, x_{U'}} \left( \prod_{i \in U' \cup V'} x_i^{deg(i)} \right) \mathbf{V}_{\rho', x_{V'}}^T]$. Then, $H_{\tau} = \sum_{\rho \in P_{\tau}} H_{\tau, \rho, \pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)}$. We have

$$\begin{bmatrix} S(\tau) R(\tau) \mathbf{H}_{Id_U} & \mathbf{H}_\tau \mathbf{H}_{Id_U}^T \\ \mathbf{H}_\tau^T & \mathbf{H}_{Id_U} \end{bmatrix} = \frac{S(\tau)}{|Aut(U)|^2} \begin{bmatrix} R(\tau) \sum_{\rho \in P_{U}} W_{\rho, \rho} & \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)} \\ \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)}^T & R(\tau) \sum_{\rho \in P_{U}} W_{\rho, \rho} \end{bmatrix} \begin{bmatrix} \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} \\ 0 \end{bmatrix}$$

Since $\frac{S(\tau)}{|Aut(U)|^2} \geq 0$, it suffices to prove that $R(\tau) \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)} R(\tau) \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} \geq 0$. Consider

$$\begin{bmatrix} R(\tau) \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} & \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)} \\ \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)}^T & R(\tau) \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} \end{bmatrix} = R(\tau) \begin{bmatrix} \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} & 0 \\ 0 & \sum_{\rho \in P_{\tau} \setminus Q_{\tau}} W_{\rho, \rho} \end{bmatrix}$$

$$+ R(\tau) \begin{bmatrix} \sum_{\rho \in P_{\tau} \setminus Q_{\tau}} W_{\rho, \rho} & \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)} \\ \sum_{\rho \in P_{\tau}} Y_{\rho, \pi(\rho)}^T & R(\tau) \sum_{\rho \in P_{U \setminus Q_{\tau}}} W_{\rho, \rho} \end{bmatrix}$$

We have $\sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} = \sum_{\rho \in P_{U \setminus P_{\tau}}} \mathbf{E}[\mathbf{V}_{\rho, x_{U'}} \mathbf{V}_{\rho, x_{U'}}^T] \geq 0$. Similarly, $\sum_{\rho \in P_{\tau} \setminus Q_{\tau}} W_{\rho, \rho} \geq 0$. Also, $R(\tau) \geq 0$ and so, the first term in the above expression, $R(\tau) \begin{bmatrix} \sum_{\rho \in P_{U \setminus P_{\tau}}} W_{\rho, \rho} & 0 \\ 0 & \sum_{\rho \in P_{\tau} \setminus Q_{\tau}} W_{\rho, \rho} \end{bmatrix}$ is positive.
semidefinite. For the second term,

$$\left[ R(\tau) \sum_{\rho \in \mathcal{P}_{\tau}} W_{\rho, \rho} + \sum_{\rho \in \mathcal{P}_{\tau}} Y_{\rho, \pi(\rho)}^T R(\tau) \sum_{\rho \in \mathcal{P}_{\tau}} W_{\rho, \pi(\rho)} \right]$$

$$= \sum_{\rho \in \mathcal{P}_{\tau}} \left[ R(\tau) \mathbb{E} \left[ v_{\rho, x_{i'}}^T \left( \prod_{i \in U \cup V} x_i^{deg(i)} \right) v_{\pi(\rho), x_{i'}} \right] \right]$$

$$= \sum_{\rho \in \mathcal{P}_{\tau}} \mathbb{E} \left[ v_{\rho, x_{i'}}^T v_{\pi(\rho), x_{i'}} \left( \prod_{i \in U \cup V} x_i^{deg(i)} \right) v_{\pi(\rho), x_{i'}} \right]$$

We will prove that the term inside the expectation is positive semidefinite for each $\rho \in \mathcal{P}_{\tau}$ and each sampling of the $x_i$ from $\mathcal{D}$, which will complete the proof. Fix $\rho \in \mathcal{P}_{\tau}$ and any sampling of the $x_i$ from $\mathcal{D}$. Let $w_1 = v_{\rho, x_{i'}}$, $w_2 = v_{\pi(\rho), x_{i'}}$. Let $E = \prod_{i \in U \cup V} x_i^{deg(i)}$. We would like to prove that

$$\left[ R(\tau)w_1w_1^T \quad Ew_1w_1^T \right] \begin{bmatrix} \quad \quad Ew_1w_1^T \\
Ew_1w_1^T & R(\tau)w_2w_2^T \end{bmatrix} \succeq 0.$$  

For all $y$ sampled from $\mathcal{D}$, $|y| \leq C_{\text{disc}} \sqrt{\mathcal{D}} E$ and so, $|E| \leq (C_{\text{disc}} \sqrt{\mathcal{D}} E) \sum_{i \in U \cup V} x_i^{deg(i)} = R(\tau)$.

If $E \geq 0$, then

$$\left( R(\tau) - E \right) \begin{bmatrix} w_1w_1^T & 0 \\
0 & w_2w_2^T \end{bmatrix}$$

$$\begin{bmatrix} w_1w_1^T & w_1w_1^T \\
w_1w_1^T & w_2w_2^T \end{bmatrix} E = \begin{bmatrix} w_1w_1^T & 0 \\
0 & w_2w_2^T \end{bmatrix} + E \begin{bmatrix} w_1w_1^T & w_1w_1^T \\
w_1w_1^T & w_2w_2^T \end{bmatrix}$$

$$\succeq 0$$

since $R(\tau) - E \succeq 0$ and if $E < 0$,

$$\left( R(\tau) + E \right) \begin{bmatrix} w_1w_1^T & 0 \\
0 & w_2w_2^T \end{bmatrix}$$

$$\begin{bmatrix} w_1w_1^T & w_1w_1^T \\
w_1w_1^T & w_2w_2^T \end{bmatrix} E = \begin{bmatrix} w_1w_1^T & 0 \\
0 & w_2w_2^T \end{bmatrix} + E \begin{bmatrix} w_1w_1^T & w_1w_1^T \\
w_1w_1^T & w_2w_2^T \end{bmatrix}$$

$$\succeq 0$$

since $R(\tau) + E \succeq 0$.




6.2.3 Proof of Lemma 6.17

The next proposition captures the fact that when we compose shapes $\sigma, \gamma, \gamma^T, \sigma^T$, in order for $\lambda_{\sigma \gamma \gamma^T \sigma^T}$ to be nonzero, the parities of the degrees of the merged vertices should add up correspondingly.

**Definition 6.31.** For all $U, V \in \mathcal{I}_{\text{mid}}$ where $w(U) > w(V)$, for $\gamma \in \Gamma_{U, V}$ and parity vectors $\rho, \rho' \in \mathcal{P}_U$, define the $\gamma^T$-coefficient matrix $H_{\rho \rho'}^{\gamma^T}$ as $H_{\rho \rho'}^{\gamma^T}(\sigma, \sigma') = H_{\rho \rho'}^{\gamma^T}(\sigma, \sigma')$ if $\sigma \in \mathcal{L}_{U, \rho}, \sigma' \in \mathcal{L}_{U, \rho'}$ and 0 otherwise.
Proposition 6.32. For all \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \), for all \( \gamma \in \Gamma_{U, V} \), there exists a set of parity vectors \( P_\gamma \subseteq P_U \) such that
\[
H_{Id_{V, \gamma}}^{-\gamma} = \sum_{\rho \in P_\gamma} H_{Id_{V, \rho, \rho}}^{-\gamma}
\]

Proof. Take any \( \rho \in P_U \). For \( \sigma \in \mathcal{L}_{\rho, \rho}, \sigma' \in \mathcal{L}_{\rho, \rho} \), since \( H_{Id_{V, \gamma}}^{-\gamma}(\sigma, \sigma') = \frac{\lambda_{\sigma \sigma' \gamma \gamma} T_{\sigma' \sigma}}{|\text{Aut}(V)|} \), \( H_{Id_{V, \gamma}}^{-\gamma}(\sigma, \sigma') \) is nonzero precisely when \( \lambda_{\sigma \sigma' \gamma \gamma} T_{\sigma' \sigma} \) is nonzero. For this quantity to be nonzero, using Definition 6.9, we get that it is necessary, but not sufficient, that the parity vector of \( \sigma' \) must also be \( \rho \). And also observe that there exists a set \( P_\gamma \) of parity vectors \( \rho \) for which \( H_{Id_{V, \rho, \rho}}^{-\gamma} \) is nonzero and their sum is precisely \( H_{Id_{V, \gamma}}^{-\gamma} \). \( \blacksquare \)

Definition 6.33. For all \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \), for all \( \gamma \in \Gamma_{U, V} \) and parity vector \( \rho \in P_U \), define the matrix \( H_{Id_{V, \rho, \rho}}^{-\gamma} \) as \( H_{Id_{V, \rho, \rho}}^{-\gamma}(\sigma, \sigma') = H_{\gamma}(\sigma, \sigma') \) if \( \sigma, \sigma' \in \mathcal{L}_{\rho, \rho} \) and 0 otherwise.

Proposition 6.34. For all \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \), for \( \gamma \in \Gamma_{U, V} \), \( H_{\gamma} = \sum_{\rho \in P_\gamma} H_{Id_{V, \rho, \rho}}^{-\gamma} \)

We will now define vectors which are truncations of \( \nu_{\rho, \chi_{U, 2}} \).

Definition 6.35. Let \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \), and let \( \gamma \in \Gamma_{U, V} \). Let \( x_i \) for \( i \in U_2 \) be variables. Denote them collectively as \( \chi_{U, 2} \). For \( \rho \in P_U \), define \( \nu_{\rho, \chi_{U, 2}}^{\gamma} \) to be the vector indexed by left shapes \( \sigma \in \mathcal{L} \) such that the \( \sigma \)th entry is \( \nu_{\rho, \chi_{U, 2}}^{\gamma}(\sigma) \) if \( |\nu(\sigma \circ \gamma)| \leq \deg(\gamma) \) and parity vector \( \rho \) and \( \gamma \) such that the \( \sigma \)th entry is \( \nu_{\rho, \chi_{U, 2}}^{\gamma}(\sigma) \).

We restate Definition 6.16 for convenience.

Definition 6.16. For all \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \) and \( \gamma \in \Gamma_{U, V} \), if \( \deg(\gamma) \) is even for all vertices \( i \) in \( V(\gamma) \setminus U_\gamma \setminus V_\gamma \), define
\[
S(\gamma) = \left( \frac{k}{d} \right)^{|\gamma_1| - \frac{|V_1| + |V_2|}{2}} \Delta|\gamma_2| - \frac{|V_1| + |V_2|}{2} \prod_{j \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} (\deg(\gamma) - 1)!! \prod_{e \in E(\gamma)} \frac{\sqrt{|V_1|}}{\sqrt{K^e}}
\]

Otherwise, define \( S(\gamma) = 0 \).

Proposition 6.36. For any \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \), and for any \( \gamma \in \Gamma_{U, V} \), suppose we take \( \rho \in P_\gamma \). When we compose \( \gamma \) with \( \gamma^T \) to get \( \gamma \circ \gamma^T \), let \( U' = (U \gamma\gamma^T)_2 \), \( V' = (V \gamma\gamma^T)_2 \) be the type 2 vertices in \( U \gamma\gamma^T, V \gamma\gamma^T \) respectively. And let \( W' \) be the set of type 2 vertices in \( \gamma \circ \gamma^T \) that were identified in the composition when we set \( V_\gamma = \gamma_\gamma^T \). Let \( x_i \) for \( i \in U' \cup W' \cup V' \) be random variables independently sampled from \( D \). Define \( x_{U'} \) (resp. \( x_{V'} \), \( x_{W'} \)) to be the subset of variables \( x_i \) for \( i \in U' \) (resp. \( i \in V', i \in W' \)). Then,
\[
H_{Id_{V, \rho, \rho}}^{-\gamma} = \frac{1}{|\text{Aut}(V)|} S(\gamma) ^2 \mathbb{E}_x \left[ (\nu_{\rho, \chi_{U, 2}}^{\gamma}(x)) \left( \prod_{i \in U' \cup W' \cup V'} x_i^{\deg(\gamma)(i)} \right) \right]
\]

Proof. Fix \( \sigma, \sigma' \in \mathcal{L}_{\rho, \rho} \) such that \( |\nu(\sigma \circ \gamma)|, |\nu(\sigma' \circ \gamma)| \leq \deg(\gamma) \). Note that for \( t \in \{1, 2\}, |\sigma_t - \frac{|V_1|}{2} + |\sigma'|_t - \frac{|V_2|}{2} + 2(|\gamma_t| - \frac{|U_1| + |V_1|}{2}) = |\sigma \circ \gamma \circ \gamma^T \circ \sigma' T| \). We can easily verify the equality using Definition 6.9 and Definition 6.25. \( \blacksquare \)
**Proposition 6.37.** For any $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, and for any $\gamma \in \Gamma_{U,V}$, suppose we take $\rho \in P_{U}$. Then,

$$H'_{\gamma, \rho, \rho} = \frac{1}{|Aut(U)|} \mathbb{E}_{y_{U_2} \sim D_{U_2}} \left[ (v_{\rho, y_{U_2}})^{\gamma} (v_{\rho, y_{U_2}}')^T \right]$$

We can now prove Lemma 6.17.

**Lemma 6.17.** For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,

$$\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H_{Id_{U}}^{-\gamma, \gamma} \leq H'_{\gamma, \rho, \rho}$$

**Proof.** Let $U', V', W'$ be as in Proposition 6.36. We have

$$\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H_{Id_{U}}^{-\gamma, \gamma} = \sum_{\rho \in P_{\gamma}} \frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H_{Id_{U}, \rho, \rho}$$

$$= \sum_{\rho \in P_{\gamma}} \frac{1}{|Aut(U)|} \cdot \frac{1}{R(\gamma)^2} \mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) \left( \prod_{i \in U' \cup W \cup V'} x_i^{deg_{U'}^\gamma T(i)} \right) \right]$$

We will now prove that, for all $\rho \in P_{\gamma}$,

$$\frac{1}{|Aut(U)|} \cdot \frac{1}{R(\gamma)^2} \mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) \left( \prod_{i \in U' \cup W \cup V'} x_i^{deg_{U'}^\gamma T(i)} \right) \right] \leq H'_{\gamma, \rho, \rho}$$

This reduces to proving that

$$\frac{2}{R(\gamma)^2} \mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) \left( \prod_{i \in U' \cup W \cup V'} x_i^{deg_{U'}^\gamma T(i)} \right) \right] \leq 2 \mathbb{E}_{y_{U_2} \sim D_{U_2}} \left[ (v_{\rho, y_{U_2}}) (v_{\rho, y_{U_2}}')^T \right]$$

where the last equality followed from linearity of expectation and the fact that $U' \equiv V' \equiv U_2$.

Since $H_{Id_{U}, \rho, \rho}$ is symmetric, we have

$$\mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) \left( \prod_{i \in U' \cup W \cup V'} x_i^{deg_{U'}^\gamma T(i)} \right) \right] = \mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) \left( \prod_{i \in U' \cup W \cup V'} x_i^{deg_{U'}^\gamma T(i)} \right) \right]$$

So, it suffices to prove

$$\frac{1}{R(\gamma)^2} \mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) \left( \prod_{i \in U' \cup W \cup V'} x_i^{deg_{U'}^\gamma T(i)} \right) \right] \leq \mathbb{E}_{x} \left[ (v_{\rho, x_{U'}}) (v_{\rho, x_{U'}}')^T \right]$$

42
We will prove that for every sampling of the \( x_i \) from \( \mathcal{D} \), we have

\[
\frac{1}{R(\gamma)^2} \left( (v_{p,x_{i'}}^{-\gamma}) \prod_{i \in U' \cup W' \cup V'} x_i^{\deg_{\gamma}^T(i)} (v_{p,x_{i'}}^{-\gamma})^T + (v_{p,x_{i'}}^{-\gamma}) \prod_{i \in U' \cup W' \cup V'} x_i^{\deg_{\gamma}^T(i)} (v_{p,x_{i'}}^{-\gamma})^T \right) 
\]

Then, taking expectations will give the result. Indeed, fix a sampling of the \( x_i \) from \( \mathcal{D} \). Let \( E = \prod_{i \in U' \cup W' \cup V'} x_i^{\deg_{\gamma}^T(i)} \) and let \( w_1 = v_{p,x_{i'}}^{-\gamma}, w_2 = v_{p,x_{i'}}^{-\gamma} \). Then, the inequality we need to show is

\[
\frac{E}{R(\gamma)^2} (w_1 w_2^T + w_2 w_2^T) \leq w_1 w_1^T + w_2 w_2^T 
\]

Now, since \( |x_i| \leq C_{\text{disc}} \sqrt{D_E} \) for all \( i \), we have \( |E| \leq \prod_{i \in U' \cup W' \cup V'} (C_{\text{disc}} \sqrt{D_E})^{\deg_{\gamma}^T(i)} = R(\gamma)^2 \).

If \( E \geq 0 \), using \( \frac{E}{R(\gamma)^2} (w_1 w_2^T + w_2 w_2^T) \geq 0 \) gives

\[
\frac{E}{R(\gamma)^2} (w_1 w_2^T + w_2 w_2^T) \geq \frac{E}{R(\gamma)^2} (w_1 w_1^T + w_2 w_2^T) 
\]

since \( 0 \leq E \leq R(\gamma)^2 \).

And if \( E < 0 \), using \( \frac{-E}{R(\gamma)^2} (w_1 w_2^T + w_2 w_2^T) \leq 0 \) gives

\[
\frac{E}{R(\gamma)^2} (w_1 w_2^T + w_2 w_2^T) \leq \frac{-E}{R(\gamma)^2} (w_1 w_1^T + w_2 w_2^T) 
\]

since \( 0 \leq -E \leq R(\gamma)^2 \).

Finally, we use the fact that for all \( \rho \in \mathcal{P}_{\mathcal{U}} \), we have \( H_{\mathcal{T}', \rho \cdot \rho} \geq 0 \) which can be proved the same way as the proof of Lemma 6.11. Therefore,

\[
\frac{|\text{Aut}(V)|}{|\text{Aut}(\mathcal{U})|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H_{\mathcal{I}'_{\mathcal{T}}} \leq \sum_{\rho \in \mathcal{P}_{\mathcal{T}}} H_{\mathcal{T}'_{\rho \cdot \rho}} 
\]

\[
= \sum_{\rho \in \mathcal{P}_{\mathcal{U}}} H_{\mathcal{T}'_{\rho \cdot \rho}} = H_{\mathcal{T}'} 
\]

7 Technical Definitions and Theorem Statement

7.1 Section Introduction

In this section, we make our definitions and results more precise. We also generalize our definitions and results to handle problems where one or more of the following is true:

43
1. The input entries correspond to hyperedges rather than edges.
2. We have different types of indices.
3. $\Omega$ is a more complicated distribution than $\{-1, +1\}$.
4. We have to consider matrix indices which are not multilinear.

Throughout this section and the remainder of this manuscript, we give the reader a choice for the level of generality of this machinery. In particular, we will first recall our definition for the simpler case when our input is $\{-1, +1\}^{(2)}$ and we only consider multilinear indices. We will then discuss how this simpler definition generalizes. We denote these generalizations with an asterix $\ast$.

### 7.1.1 Additional Parameters for the General Case $\ast$

In the general case we will need a few additional parameters which we define here.

**Definition 7.1.**

1. We define $k$ to be the arity of the hyperedges corresponding to the input.
2. We define $t_{\text{max}}$ to be the number of different types of indices. We define $n_i$ to be the number of possibilities for indices of type $i$ and we define $n = \max \{ n_i : i \in [t_{\text{max}}] \}$.

### 7.2 Indices, Input Entries, Vertices, and Edges

Note: For this section, we use $X$ to denote the input, we use $x$ to denote entries of the input and we use $y$ to denote solution variables.

**Definition 7.2** (Vertices: Simplified Case). When the input and solution variables are indexed by one type of index which takes values in $[n]$ then we represent the index $i$ by a vertex labeled $i$.

If we want to leave an index unspecified, we instead represent it by a vertex labeled with a variable (we will generally use $u$, $v$, or $w$ for these variables).

**Definition 7.3** (Vertices: General Case $\ast$). When the input and solution variables are indexed by several types of indices where indices of type $t$ take values in $[n_t]$, we represent an index of type $t$ with value $i$ as a vertex labeled by the tuple $(t, i)$. We say that such a vertex has type $t$.

If we want to leave an index of type $t$ unspecified, we instead represent it by a vertex labeled with a tuple $(t, ?)$ where $?$ is a variable (which will generally be $u$, $v$, or $w$).

**Definition 7.4** (Edges: Simplified Case). When the input is $X \in \{-1, +1\}^{(2)}$, we represent the entries of the input by the undirected edges $\{(i, j) : i < j \in [n]\}$. Given an edge $e = (i, j)$, we take $x_e = x_{ij}$ to be the input entry corresponding to $e$.

**Definition 7.5** (Edges: General Case $\ast$). In general, we represent the entries of the input by hyperedges whose form depends on nature of the input. We still take $x_e$ to be the input entry corresponding to $e$.

**Example 7.6.** If the input is an $n_1 \times n_2$ matrix $X$ then we will have two types of indices, one for the row and one for the column. Thus, we will have the vertices $\{(1, i) : i \in [n_1]\} \cup \{(2, j) : j \in [n_2]\}$. In this case, we have an edge $((1, i), (2, j))$ for each entry $x_{ij}$ of the input.
Example 7.7. If the input is an \( n \times n \) matrix \( X \) which is not symmetric then we only need the indices \([n]\). In this case, we have a directed edge \((i, j)\) for each entry \(x_{ij}\) where \(i \neq j\). If the entries \(x_{ii}\) are also part of the input then we also have loops \((i, i)\) for these entries.

Example 7.8. If our input is a symmetric \( n \times n \times n \) tensor \( X \), i.e., \( x_{ijk} = x_{jik} = x_{jki} = x_{kij} \) and \( x_{ijk} = 0 \) whenever \(i, j, k\) are not distinct then we only need the indices \([n]\). In this case, we have an undirected hyperedge \( e = (i, j, k) \) for each entry \(x_e = x_{ijk}\) of the input where \(i, j, k\) are distinct.

Example 7.9. If the input is an \( n_1 \times n_2 \times n_3 \) tensor \( X \) then we will have three types of indices. Thus, we will have the vertices \(\{ (1, i) : i \in [n_1]\} \cup \{ (2, j) : j \in [n_2]\} \cup \{ (3, k) : k \in [n_3]\}\). In this case, we have a hyperedge \( e = ((1, i), (2, j), (3, k))\) for each entry \(x_e = x_{ijk}\) of the input.

### 7.3 Matrix Indices and Monomials

In this subsection, we discuss how our matrices are indexed and how we associate matrix indices with monomials. We also describe the automorphism groups of matrix indices.

**Definition 7.10** (Matrix Indices: Simplified Case). If there is only one type of index and we have the constraints \(y_i^2 = 1\) or \(y_i^2 = y_i\) on the solution variables then we define a matrix index \(A\) to be a tuple of indices \((a_1, \ldots, a_{|A|})\). We make the following definitions about matrix indices:

1. We associate the monomial \(\prod_{i=1}^{|A|} y_{a_i}\) to \(A\).
2. We define \(V(A)\) to be the set of vertices \(\{a_i : i \in [|A|]\}\). For brevity, we will often write \(A\) instead of \(V(A)\) when it is clear from context that we are referring to \(A\) as a set of vertices rather than a matrix index.
3. We take the automorphism group of \(A\) to be \(\text{Aut}(A) = S_{|A|}\) (the permutations of the elements of \(A\)).

**Example 7.11.** The matrix index \(A = (4, 6, 1)\) represents the monomial \(y_4 y_6 y_1 = y_1 y_4 y_6\) and \(\text{Aut}(A) = S_3\).

**Remark 7.12.** We take \(A\) to be an ordered tuple rather than a set for technical reasons.

In general, we need a more intricate definition for matrix indices. We start by defining matrix index pieces.

**Definition 7.13** (Matrix Index Piece Definition*). We define a matrix index piece \(A_i = ((a_{i1}, \ldots, a_{i|A_i|}), t_i, p_i)\) to be a tuple of indices \((a_{i1}, \ldots, a_{i|A_i|})\) together with a type \(t_i\) and a power \(p_i\). We make the following definitions about matrix index pieces:

1. We associate the monomial \(p_{A_i} = \prod_{j=1}^{|A_i|} y_{t_{ij}}^{p_i}\) with \(A_i\).
2. We define \(V(A_i)\) to be the set of vertices \(\{(t_i, a_{ij}) : j \in [|A_i|]\}\).
3. We take the automorphism group of \(A_i\) to be \(\text{Aut}(A_i) = S_{|A_i|}\).
4. We say that \(A_i\) and \(A_j\) are disjoint if \(V(A_i) \cap V(A_j) = \emptyset\) (i.e., \(t_i \neq t_j\) or \(\{a_{i1}, \ldots, a_{i|A_i|}\} \cap \{a_{j1}, \ldots, a_{j|A_j|}\} = \emptyset\)).

**Definition 7.14** (General Matrix Index Definition*). We define a matrix index \(A = \{A_i\}\) to be a set of disjoint matrix index pieces. We make the following definitions about matrix indices:
1. We associate the monomial \( p_A = \prod_{A_i \in A} p(A_i) \) with \( A \).

2. We define \( V(A) \) to be the set of vertices \( \cup_{A_i \in A} V(A_i) \). For brevity, we will often write \( A \) instead of \( V(A) \) when it is clear from context that we are referring to \( A \) as a set of vertices rather than a matrix index.

3. We take the automorphism group of \( A \) to be \( \text{Aut}(A) = \prod_{A_i \in A} \text{Aut}(A_i) \)

Example 7.15 (*). If \( A_1 = ((2), 1, 1) \), \( A_2 = ((3, 1), 1, 2) \), and \( A_3 = ((1, 2), 3, 2, 1) \) then \( A = \{ A_1, A_2, A_3 \} \) represents the monomial \( p = y_{12}y_{13}^2y_{11}^2y_{21}y_{22}y_{23} \) and we have \( \text{Aut}(A) = S_1 \times S_2 \times S_3 \).

### 7.4 Fourier Characters and Ribbons

A key idea is to analyze Fourier characters of the input.

Definition 7.16 (Simplified Fourier Characters). If the input distribution is \( \Omega = \{-1, 1\} \) then given a multi-set of edges \( E \), we define \( \chi_E(X) = \prod_{e \in E} x_e \).

Example 7.17. If the input is a graph \( G \in \{-1, 1\}^{|G|} \) and \( E \) is a set of potential edges of \( G \) (with no multiple edges) then \( \chi_E(G) = (-1)^{|E \setminus E(G)|} \).

In general, the Fourier characters are somewhat more complicated.

Definition 7.18 (Orthonormal Basis for \( \Omega^* \)). We define the polynomials \( \{ h_i : i \in \mathbb{Z} \cap [0, \text{supp}(\Omega) - 1] \} \) to be the unique polynomials (which can be found through the Gram-Schmidt process) such that

1. \( \forall i, E_\Omega[h_i^2(x)] = 1 \)
2. \( \forall i \neq j, E_\Omega[h_i(x)h_j(x)] = 0 \)
3. For all \( i \), the leading coefficient of \( h_i(x) \) is positive.

Example 7.19. If \( \Omega \) is the normal distribution then the polynomials \( \{ h_i \} \) are the Hermite polynomials with the appropriate normalization so that for all \( i \), \( E_\Omega[h_i^2(x)] = 1 \). In particular, \( h_0(x) = 1 \), \( h_1(x) = x \), \( h_2(x) = \frac{x^2 - 1}{\sqrt{2}} \), \( h_3(x) = \frac{x^3 - 3x}{\sqrt{6}} \), etc.

Definition 7.20 (General Fourier Characters (*)). Given a multi-set of hyperedges \( E \), each of which has a label \( l(e) \in \text{[support}(\Omega) - 1] \) (or \( \mathbb{N} \) if \( \Omega \) has infinite support), we define \( \chi_E = \prod_{l(e) \in E} h_{l(e)}(X_e) \).

We say that such a multi-set of hyperedges \( E \) is proper if it contains no duplicate hyperedges, i.e. it is a set (though the labels on the hyperedges can be arbitrary non-negative integers). Otherwise, we say that \( E \) is improper.

Remark 7.21. The Fourier characters are \( \{ \chi_E : E \text{ is proper} \} \). For improper \( E \), \( \chi_E \) can be decomposed as a linear combination of \( \chi_{E_j} \) where each \( E_j \) is proper. We allow improper \( E \) because it is sometimes more convenient to have improper \( E \) in the middle of the analysis and then do this decomposition at the end.

Definition 7.22 (Ribbons). A ribbon \( R \) is a tuple \( (H_R, A_R, B_R) \) where \( H_R \) is a multi-graph (*or multi-hypergraph with labeled edges in the general case) whose vertices are indices of the input and \( A_R \) and \( B_R \) are matrix indices such that \( V(A_R) \subseteq V(H_R) \) and \( V(B_R) \subseteq V(H_R) \). We make the following definitions about ribbons:

1. We define \( V(R) = V(H_R) \) and \( E(R) = E(H_R) \)
2. We define $\chi_R = \chi_{E(R)}$.
3. We define $M_R$ to be the matrix such that $(M_R)_{A_R B_R} = \chi_R$ and $M_{AB} = 0$ whenever $A \neq A_R$ or $B \neq B_R$.

We say that $R$ is a proper ribbon if $H_R$ contains no isolated vertices outside of $A_R \cup B_R$ and $E(R)$ is proper. If there is an isolated vertex in $(V(R) \setminus A_R) \setminus B_R$ or $E(R)$ is improper then we say that $R$ is an improper ribbon.

Proper ribbons are useful because they give an orthonormal basis for the space of matrix valued functions.

**Definition 7.23** (Inner products of matrix functions). For a pair of real matrices $M_1, M_2$ of the same dimension, we write $\langle M_1, M_2 \rangle = \text{tr}(M_1 M_2^T)$ (i.e. $\langle M_1, M_2 \rangle$ is the entrywise dot product of $M_1$ and $M_2$). For a pair of matrix-valued functions $M_1, M_2$ (of the same dimensions), we define

$$\langle M_1, M_2 \rangle = E_x [(M_1(X), M_2(X))]$$

**Proposition 7.24.** If $R$ and $R'$ are two proper ribbons then $\langle M_R, M_{R'} \rangle = 1$ if $R = R'$ and is 0 otherwise.

7.5 Shapes

In this subsection, we describe a basis for $S$-invariant matrix valued functions where each matrix in this basis can be described by a relatively small shape $\alpha$. The fundamental idea behind shapes is that we keep the structure of the objects we are working with but leave the elements of the object unspecified.

7.5.1 Simplified Index Shapes

**Definition 7.25** (Simplified Index shapes). With our simplifying assumptions, an index shape $U$ is a tuple of unspecified indices $(u_1, \ldots, u_{|U|})$. We make the following definitions about index shapes:

1. We define $V(U)$ to be the set of vertices $\{u_i : i \in [|U|]\}$. For brevity, we will often write $U$ instead of $V(U)$ when it is clear from context that we are referring to $U$ as a set of vertices rather than an index shape.
2. We define the weight of $U$ to be $\omega(U) = |U|$.
3. We take the automorphism group of $U$ to be $\text{Aut}(U) = S_{|U|}$ (the permutations of the elements of $U$).

**Definition 7.26.** We say that a matrix index $A = (a_1, \ldots, a_{|A|})$ has index shape $U = (u_1, \ldots, u_{|U|})$ if $|U| = |A|$. Note that in this case, if we take the map $\varphi : \{u_j : j \in [|U|]\} \rightarrow [n]$ where $\varphi(u_j) = a_j$ then $\varphi(U) = (\varphi(u_1), \ldots, \varphi(u_{|U|})) = (a_1, \ldots, a_{|A|}) = A$

**Definition 7.27.** We say that index shapes $U = (u_1, \ldots, u_{|U|})$ and $V = (v_1, \ldots, v_{|V|})$ are equivalent (which we write as $U \equiv V$) if $|U| = |V|$. If $U \equiv V$ then we can set $U = V$ by setting $v_j = u_j$ for all $j \in [|U|]$.

**Example 7.28.** The matrix index $A = \{4, 6, 1\}$ has shape $U = \{u_1, u_2, u_3\}$ which has weight 3.
7.5.2 General Index Shapes*

In general, we define general index shapes in the same way that we defined general matrix indices (just with unspecified indices)

**Definition 7.29** (Index Shape Piece Definition). We define a index shape piece \( U_i = ((u_{i_1}, \ldots, u_{i_{|U_i|}}), t_i, p_i) \) to be a tuple of indices \( (u_{i_1}, \ldots, u_{i_{|U_i|}}) \) together with a type \( t_i \) and a power \( p_i \). We make the following definitions about index shape pieces:

1. We define \( V(U_i) \) to be the set of vertices \( \{(t_i, u_{ij}) : j \in [|U_i|]\} \).
2. We define \( w(U_i) = |U_i| \log_\alpha(n_i) \)
3. We take the automorphism group of \( U_i \) to be \( \text{Aut}(U_i) = S_{|U_i|} \)

**Definition 7.30** (General Index Shape Definition). We define an index shape \( U = \{U_i\} \) to be a set of index shape pieces such that for all \( i' \neq i \), either \( t_{i'} \neq t_i \) or \( p_{i'} \neq p_i \). We make the following definitions about index shapes:

1. We define \( V(U) \) to be the set of vertices \( \bigcup_{U_i \in U} V(U_i) \). For brevity, we will often write \( U \) instead of \( V(U) \) when it is clear from context that we are referring to \( U \) as a set of vertices rather than an index shape.
2. We define \( w(U) = \sum_{U_i \in U} w(U_i) \)
3. We take the automorphism group of \( U \) to be \( \text{Aut}(U) = \prod_{U_i \in U} \text{Aut}(U_i) \)

**Remark 7.31.** For technical reasons, we want to ensure that if two index shapes \( U \) and \( U' \) have the same weight then \( U \) and \( U' \) have the same number of each type of vertex. To ensure this, we add an infinitesimal perturbation to each \( n_i \) if necessary.

**Definition 7.32.** We say that a matrix index \( A \) has index shape \( U \) if there is an assignment of values to the unspecified indices of \( U \) which results in \( A \). More precisely, we say that \( A \) has index shape \( U \) if there is a map \( \varphi : \{u_{ij}\} \to \mathbb{N} \) such that if we define \( \varphi(U_i) \) to be \( \varphi(U_i) = ((\varphi(u_{i_1}), \ldots, \varphi(u_{i_{|U_i|}})), t_i, p_i) \) then \( \varphi(U) = \{\varphi(U_i)\} = \{A_i\} = A \).

**Definition 7.33.** If \( U \) and \( V \) are two index shapes, we say that \( U \) is equivalent to \( V \) (which we write as \( U \equiv V \)) if \( U \) and \( V \) have the same number of index shape pieces and we can order the index shape pieces of \( U \) and \( V \) so that writing \( U = \{U_i\} \) and \( V = \{V_i\} \) where \( U_i = ((u_{i_1}, \ldots, u_{i_{|U_i|}}), t_i, p_i) \) and \( V_i = ((v_{i_1}, \ldots, v_{i_{|V_i|}}), t'_i, p'_i) \), we have that for all \( i \), \( |V_i| = |U_i| \), \( t'_i = t_i \), and \( p'_i = p_i \). If \( U \equiv V \) then we can set \( U = V \) by setting \( u_{ij} = v_{ij} \) for all \( i \) and all \( j \in [|U_i|] \).

7.5.3 Ribbon Shapes

With these definitions, we are now ready to define shapes and the matrices associated to them.

**Definition 7.34** (Shapes). A ribbon shape \( \alpha \) (which we call a shape for brevity) is a tuple \( \alpha = (H_\alpha, U_\alpha, V_\alpha) \) where \( H_\alpha \) is a multi-graph (*or multi-hypergraph with labeled edges in the general case) whose vertices are unspecified distinct indices of the input (*whose type is specified in the general case) and \( U_\alpha \) and \( V_\alpha \) are index shapes such that \( V(U_\alpha) \subseteq V(H_\alpha) \) and \( V(V_\alpha) \subseteq V(H_\alpha) \). We make the following definitions about shapes:
1. We define \( V(\alpha) = V(H_\alpha) \) (note that \( V(\alpha) \) and \( V_\alpha \) are not the same thing) and we define \( E(\alpha) = E(H_\alpha) \).

2. We say that a shape \( \alpha \) is proper if it contains no isolated vertices outside of \( V(U_\alpha) \cup V(V_\alpha) \). \( E(\alpha) \) has no multiple edges/hyperedges and edges in \( E(\alpha) \) do not have label 0. If there is an isolated vertex in \( V(\alpha) \setminus V(U_\alpha) \setminus V(V_\alpha) \) or \( E(\alpha) \) has a multiple edge/hyperedge then we say that \( \alpha \) is an improper shape.

Note: For brevity, we will often write \( U_\alpha \) and \( V_\alpha \) instead of \( V(U_\alpha) \) and \( V(V_\alpha) \) when it is clear from context that we are referring to \( U_\alpha \) and \( V_\alpha \) as sets of vertices rather than index shapes.

**Definition 7.35** (Trivial shapes). We say that a shape \( \alpha \) is trivial if \( V(\alpha) = V(U_\alpha) = V(V_\alpha) \) and \( E(\alpha) = \emptyset \). Otherwise, we say that \( \alpha \) is non-trivial.

**Remark 7.36.** Note that all trivial shapes can do is permute the order of the vertices in \( V(U_\alpha) = V(V_\alpha) \).

**Definition 7.37.** Informally, we say that a ribbon \( R \) has shape \( \alpha \) if replacing the indices in \( R \) with unspecified labels results in \( \alpha \). Formally, we say that \( R \) has shape \( \alpha \) if there is an injective mapping \( \varphi : V(\alpha) \rightarrow [n] \) (or \([t_{\max}] \times [n] \) in the general case) such that \( \varphi(\alpha) = R \), i.e. \( \varphi(H_\alpha) = H_R \), \( \varphi(U_\alpha) = A_R \), and \( \varphi(V_\alpha) = B_R \).

**Definition 7.38.** We say that two shapes \( \alpha \) and \( \beta \) are equivalent (which we write as \( \alpha \equiv \beta \)) if they are the same up to renaming their indices. More precisely, we say that \( \alpha \equiv \beta \) if there is a bijective map \( \pi : V(H_\alpha) \rightarrow V(H_\beta) \) such that \( \pi(H_\alpha) = H_\beta \), \( \pi(U_\alpha) = U_\beta \), and \( \pi(V_\alpha) = V_\beta \).

**Definition 7.39.** Given a shape \( \alpha \) and matrix indices \( A, B \) of shapes \( U_\alpha \) and \( V_\alpha \) respectively, we define \( R(\alpha, A, B) \) to be the set of ribbons \( R \) such that \( R \) has shape \( \alpha \), \( A_R = A \), and \( B_R = B \).

**Definition 7.40.** For a shape \( \alpha \), we define the matrix-valued function \( M_\alpha \) to have entries \( M_\alpha(A, B) \) given by

\[
(M_\alpha)_{A,B}(X) = \sum_{R \in R(\alpha, A, B)} \chi_R(X)
\]

For examples of \( M_\alpha \), see [AMP20].

**Proposition 7.41.** The \( M_\alpha \)'s for proper shapes \( \alpha \) are an orthogonal basis for the \( S \)-invariant functions.\(^1\)

**Remark 7.42.** Conceptually, one may think of forming an orthonormal basis for this space with the functions \( M_\alpha / \sqrt{\langle M_\alpha, M_\alpha \rangle} \), but for technical reasons it is easiest to work with these functions without normalizing them to 1. By orthogonality and the fact that every Boolean function is a polynomial, any \( S \)-invariant matrix-valued function \( \Lambda \) is expressible as

\[
\Lambda = \sum_\alpha \frac{\langle \Lambda, M_\alpha \rangle}{\langle M_\alpha, M_\alpha \rangle} \cdot M_\alpha
\]

### 7.6 Composing Ribbons and Shapes

**Definition 7.43** (Composing Ribbons). We say that ribbons \( R_1 \) and \( R_2 \) are composable if \( B_{R_1} = A_{R_2} \). Note that this definition is not symmetric so we may have that \( R_1 \) and \( R_2 \) are composable but \( R_2 \) and \( R_1 \) are not composable.

\(^1\)Because of orthogonality of the underlying Fourier characters, it is not hard to check that when \( \alpha \neq \alpha' \) and \( M_\alpha, M_{\alpha'} \) have the same dimensions, \( \langle M_\alpha, M_{\alpha'} \rangle = 0 \).
We say that \( R_1 \) and \( R_2 \) are properly composable if we also have that \( V(R_1) \cap V(R_2) = V(B_{R_1}) = V(A_{R_2}) \) (there are no unexpected intersections between \( R_1 \) and \( R_2 \)).

If \( R_1 \) and \( R_2 \) are composable ribbons then we define the composition of \( R_1 \) and \( R_2 \) to be the ribbon \( R_1 \circ R_2 \) such that

1. \( A_{R_1 \circ R_2} = A_{R_1} \) and \( B_{R_1 \circ R_2} = B_{R_2} \)
2. \( V(R_1 \circ R_2) = V(R_1) \cup V(R_2) \)
3. \( E(R_1 \circ R_2) = E(R_1) \cup E(R_2) \) (and thus \( \chi_{R_1 \circ R_2} = \chi_{R_1} \chi_{R_2} \))

We say that ribbons \( R_1, \ldots, R_k \) are composable/properly composable if for all \( j \in [k-1] \), \( R_1 \circ \ldots \circ R_j \) and \( R_{j+1} \) are composable/properly composable. If \( R_1, \ldots, R_k \) are composable then we define \( R_1 \circ \ldots \circ R_k \) to be \( R_1 \circ \ldots \circ R_k = (R_1 \circ \ldots \circ R_{k-1}) \circ R_k \)

**Proposition 7.44.** Ribbon composition is associative, i.e. if \( R_1, R_2, R_3 \) are composable/properly composable ribbons then \( R_2 \circ R_3 \) are composable/properly composable, \( R_1, (R_2 \circ R_3) \) are composable/properly composable, and \( R_1 \circ (R_2 \circ R_3) = (R_1 \circ R_2) \circ R_3 \)

**Proposition 7.45.** If \( R_1 \) and \( R_2 \) are composable ribbons then \( M_{R_1 \cup R_2} = M_{R_1} M_{R_2} \)

We have similar definitions for composing shapes.

**Definition 7.46 (Composing Shapes).** We say that shapes \( \alpha \) and \( \beta \) are composable if \( U_\beta \equiv V_\alpha \). Note that this definition is not symmetric so we may have that \( \alpha \) and \( \beta \) are composable but \( \beta \) and \( \alpha \) are not composable.

If \( \alpha \) and \( \beta \) are composable shapes then we define the composition of \( \alpha \) and \( \beta \) to be the shape \( \alpha \circ \beta \) such that

1. \( U_{\alpha \circ \beta} = U_\alpha \) and \( V_{\alpha \circ \beta} = V_\beta \)
2. After setting \( U_\beta = V_\alpha \), we take \( V(\alpha \circ \beta) = V(\alpha) \cup V(\beta) \)
3. \( E(\alpha \circ \beta) = E(\alpha) \cup E(\beta) \)

We say that shapes \( \alpha_1, \ldots, \alpha_k \) are composable if for all \( j \in [k-1] \), \( \alpha_1 \circ \ldots \circ \alpha_j \) and \( \alpha_{j+1} \) are composable. If \( \alpha_1, \ldots, \alpha_k \) are composable then we define the shape \( \alpha_1 \circ \ldots \circ \alpha_k \) to be \( \alpha_1 \circ \ldots \circ \alpha_k = (\alpha_1 \circ \ldots \circ \alpha_{k-1}) \circ \alpha_k \)

**Proposition 7.47.** Shape composition is associative, i.e. if \( \alpha_1, \alpha_2, \alpha_3 \) are composable shapes then \( \alpha_2, \alpha_3 \) are composable, \( \alpha_1, (\alpha_2 \circ \alpha_3) \) are composable, and \( \alpha_1 \circ (\alpha_2 \circ \alpha_3) = (\alpha_1 \circ \alpha_2) \circ \alpha_3 \)

### 7.7 Decomposition of Shapes into Left, Middle, and Right parts

In this subsection, we describe how shapes can be decomposed into left, middle, and right parts based on the leftmost and rightmost *minimum vertex separators*, which is a crucial idea for our analysis.

**Definition 7.48 (Paths).** A path in a shape \( \alpha \) is a sequence of vertices \( v_1, \ldots, v_t \) such that \( v_i, v_{i+1} \) are in some edge/hyperedge together. A pair of paths is vertex-disjoint if the corresponding sequences of vertices are disjoint.

**Definition 7.49 (Vertex separators).** Let \( \alpha \) be a shape and let \( U \) and \( V \) be sets of vertices in \( \alpha \). We say that a set of vertices \( S \subseteq V(\alpha) \) is a vertex separator of \( U \) and \( V \) if every path in \( \alpha \) from \( U \) to \( V \) contains at least one vertex in \( S \). Note that any vertex separator \( S \) of \( U \) and \( V \) must contain all of the vertices in \( U \cap V \).

As a special case, we say that \( S \) is a vertex separator of \( \alpha \) if \( S \) is a vertex separator of \( U_\alpha \) and \( V_\alpha \)
We define the weight of a set of vertices $S \subseteq V(\alpha)$ in the same way that weight is defined for index shapes.

**Definition 7.50** (Simplified Weight). When there is only one type of index, the weight of a set of vertices $S \subseteq V(\alpha)$ is simply $|S|$.

**Definition 7.51** (General Weight*). In general, given a set of vertices $S \subseteq V(\alpha)$, writing $S = \cup_t S_t$ where $S_t$ is the set of vertices of type $t$ in $S$, we define the weight of $S$ to be $w(S) = \sum_t |S_t| \log_2(n_t)$

**Remark 7.52** (*). Again, if necessary, we add an infinitesimal perturbation to $n_1, n_2, \ldots, n_{1\text{max}}$ so that if two separators $S$ and $S'$ have the same weight then $S$ and $S'$ have the same number of each type of vertex.

**Definition 7.53** (Leftmost and rightmost minimum vertex separators). The leftmost minimum vertex separator is the vertex separator $S$ of minimum weight such that for every other minimum-weight vertex separator $S'$, $S$ is a separator of $U_\alpha$ and $S'$. The rightmost minimum vertex separator is the vertex separator $T$ of minimum weight such that for every other minimum-weight vertex separator $T'$, $T$ is a separator of $T'$ and $V_\alpha$.

The work [BHK+16] showed that under our simplifying assumptions, leftmost and rightmost minimum vertex separators are well defined. For a general proof that leftmost and rightmost minimum vertex separators are well defined, see Appendix A.

We now have the following crucial idea. Every shape $\alpha$ can be decomposed into the composition of three composable shapes $\sigma, \tau, \sigma^T$ based on the leftmost and rightmost minimum vertex separators $S, T$ of $\alpha$ together with orderings of $S$ and $T$.

**Definition 7.54** (Simplified Separators With Orderings). Under our simplifying assumptions, given a set of vertices $S \subseteq V(\alpha)$ and an ordering $O_S = s_1, \ldots, s_{|S|}$ of the vertices of $S$, we define the index shape $(S, O_S)$ to be $(S, O_S) = (s_1, \ldots, s_{|S|})$.

**Definition 7.55** (General Separators With Orderings*). In the general case, we need to give an ordering for each type of vertex. Let $S \subseteq V(\alpha)$ be a subset of the vertices of $\alpha$ and write $S = \cup_t S_t$ where $S_t$ is the set of vertices in $S$ of type $t$. Given $O_S = \{O_t\}$ where $O_t = s_{t|t|}, \ldots, s_{|S_t|}$ is an ordering of the vertices of $S_t$, we define the index shape piece $(S_t, O_t)$ to be $(S_t, O_t) = ((s_{t1}, \ldots, s_{t|S_t|}), t, 1)$ and we define the index shape $(S, O_S)$ to be $(S, O_S) = \{(S_t, O_t)\}$.

**Proposition 7.56**. The number of possible orderings $O$ for $S$ is equal to $|\text{Aut}((S, O_S))|$.

**Definition 7.57** (Shape transposes). Given a shape $\alpha$, we define $\alpha^T$ to be the shape $\alpha$ with $U_\alpha$ and $V_\alpha$ swapped i.e. $U_{\alpha^T} = V_\alpha$ and $V_{\alpha^T} = U_\alpha$.

**Definition 7.58** (Left, middle, and right parts). Let $\alpha$ be a shape. Let $S$ and $T$ be the leftmost and rightmost minimal vertex separators of $\alpha$ together with orderings $O_S, O_T$ of $S$ and $T$.

- We define the left part $\sigma_\alpha$ of $\alpha$ to be the shape such that
  1. $H_{\sigma_\alpha}$ is the induced subgraph of $H_\alpha$ on all of the vertices of $\alpha$ reachable from $U_\alpha$ without passing through $S$ (note that $H_{\sigma_\alpha}$ includes the vertices of $S$) except that we remove any edges/hyperedges which are contained entirely within $S$.
  2. $U_{\sigma_\alpha} = U_\alpha$ and $V_{\sigma_\alpha} = (S, O_S)$.
We define the right part $\sigma^T_\alpha$ of $\alpha$ to be the shape such that

1. $H_{\sigma^T_\alpha}$ is the induced subgraph of $H_\alpha$ on all of the vertices of $\alpha$ reachable from $V_\alpha$ without passing through $T$ (note that $H_{\sigma^T_\alpha}$ includes the vertices of $T$) except that we remove any edges/hyperedges which are contained entirely within $T$.
2. $V_{\sigma^T_\alpha} = V_\alpha$ and $U_{\sigma^T_\alpha} = (T, O_T)$

We define the middle part $\tau_\alpha$ of $\alpha$ to be the shape such that

1. $H_{\tau_\alpha}$ is the induced subgraph of $H_\alpha$ on all of the vertices of $\alpha$ which are not reachable from $U_\alpha$ and $V_\alpha$ without touching $S$ and $T$ (note that $H_{\tau_\alpha}$ includes the vertices of $S$ and $T$). $H_{\tau_\alpha}$ also includes the hyperedges entirely within $S$ and the hyperedges entirely within $T$.
2. $U_{\tau_\alpha} = (S, O_S)$ and $V_{\tau_\alpha} = (T, O_T)$

**Proposition 7.59.** If $\sigma, \tau, \sigma^T_\alpha$ are the left, middle, and right parts for $\alpha$ for given orderings $O_S, O_T$ of $S$ and $T$ then $\alpha = \sigma \circ \tau \circ \sigma^T_\alpha$.

**Remark 7.60.** One may ask which ordering(s) we should take of $S$ and $T$. The answer is that we will take all of the possible orderings of $S$ and $T$ simultaneously, giving equal weight to each.

Based on this decomposition and the following claim, we make the following definitions for what it means for a shape to be a left, middle, or right part.

**Claim 7.61** (Proved in Section 6.1 in [BHK+16]).

- Every shape $\sigma$ which is the left part of some other shape $\alpha$ has that $V_\sigma$ is its left-most and right-most minimum-weight separator.
- Every shape $\sigma^T$ which is the right part of some other shape $\alpha$ has that $U_{\sigma^T}$ is its left-most and right-most minimum-weight separator.
- Every shape $\tau$ which is the middle part of some other shape $\alpha$ has $U_\tau$ as its left-most minimum size separator and $V_\tau$ as its right-most minimum-weight separator.

**Definition 7.62.**

1. We say that a shape $\sigma$ is a left shape if $\sigma$ is a proper shape, $V_\sigma$ is the left-most and right-most minimum-weight separator of $\sigma$, every vertex in $V(\sigma) \setminus V_\sigma$ is reachable from $U_\sigma$ without touching $V_\sigma$, and $\sigma$ has no hyperedges entirely within $V_\sigma$.
2. We say that a shape $\tau$ is a proper middle shape if $\tau$ is a proper shape, $U_\tau$ is the left-most minimum-weight separator of $\tau$, and $V_\tau$ is the right most minimum-weight separator of $\tau$. In the analysis, we will also need to consider improper middle shapes $\tau$ which may not be proper shapes and which may have smaller separators between $U_\tau$ and $V_\tau$.
3. We say that a shape $\sigma^T$ is a right shape if $\sigma^T$ is a proper shape, $U_{\sigma^T}$ is the left-most and right-most minimum-weight separator of $\sigma^T$, every vertex in $V(\sigma^T) \setminus U_{\sigma^T}$ is reachable from $V_{\sigma^T}$ without touching $U_{\sigma^T}$, and $\sigma^T$ has no hyperedges entirely within $U_{\sigma^T}$.

---

The proof in [BHK+16] only explicitly treats the case when the shapes $\alpha$ are graphs, but the proof easily generalizes to the case when the $\alpha$ are hypergraphs.
Proposition 7.63. For all shapes $\sigma$, $\sigma$ is a left shape if and only if $\sigma^T$ is a right shape.

Remark 7.64. As the reader has likely guessed, throughout this section we use $\sigma$ to denote left parts and $\tau$ to denote middle parts. Instead of having a separate letter for right parts, we express right parts as the transpose of a left part.

7.8 Coefficient matrices

We will have that $\Lambda = \sum_\alpha \lambda_\alpha M_\alpha$. To analyze $\Lambda$, it is extremely useful to express these coefficients in terms of matrices. To do this, we will need a few more definitions. We start by defining the sets of index shapes that can appear when analyzing $\Lambda$.

Definition 7.65. Given a moment matrix $\Lambda$, we define the following sets of index shapes.

1. We define $\mathcal{I}(\Lambda) = \{U : \exists \text{ matrix } A : A \text{ has shape } U\}$ to be the set of index shapes which describe row and column indices of $\Lambda$.
2. We define $w_{\text{max}}$ to be $w_{\text{max}} = \max \{w(U) : U \in \mathcal{I}(\Lambda)\}$.
3. With our simplifying assumptions, we define $\mathcal{I}_{\text{mid}}$ to be $\mathcal{I}_{\text{mid}} = \{U : |U| \leq w_{\text{max}}\}$

3*. In general, we define $\mathcal{I}_{\text{mid}}$ to be $\mathcal{I}_{\text{mid}} = \{U : w(U) \leq w_{\text{max}}, \forall U_i \in U, p_i = 1\}$

We also need to define the sets of shapes which can appear when analyzing $\Lambda$.

Definition 7.66 (Truncation Parameters). Given a moment matrix $\Lambda = \sum_\alpha \lambda_\alpha M_\alpha$, we define $D_V, D_E$ to be the smallest natural numbers such that for all shapes $\alpha$ such that $\lambda_\alpha \neq 0$, decomposing $\alpha$ as $\alpha = \sigma \circ \tau \circ \sigma^T$,

1. $|V(\sigma)| \leq D_V$, $|V(\tau)| \leq D_V$, and $|V(\sigma')| \leq D_V$.
2. For all edges $e \in E(\sigma) \cup E(\tau) \cup E(\sigma')$, $l_e \leq D_E$.

Remark 7.67. Under our simplifying assumptions, all edges have label 1 so we will take $D_E = 1$ and ignore conditions involving $D_E$.

Definition 7.68. Given a moment matrix $\Lambda$, we define the following sets of shapes:

1. $\mathcal{L} = \{\sigma : \sigma \text{ is a left shape}, U_\sigma \in \mathcal{I}(\Lambda), V_\sigma \in \mathcal{I}_{\text{mid}}, V(\sigma) \leq D_V, \forall \tau \in \mathcal{E}(\sigma), l_e \leq D_E\}$
2. Given $V \in \mathcal{I}_{\text{mid}}$, we define $\mathcal{E}_V = \{\sigma \in \mathcal{L} : V_e \equiv V\}$
3. Given $U \in \mathcal{I}_{\text{mid}}$, we define $\mathcal{M}_U = \{\tau : \tau \text{ is a non-trivial proper middle shape}, U_\tau \equiv V_\tau \equiv U, |V(\tau)| \leq D_V, \forall \tau \in \mathcal{E}(\tau), l_e \leq D_E\}$

Definition 7.69. Given a moment matrix $\Lambda$, we define a $\Lambda$-coefficient matrix (which we call a coefficient matrix for brevity) to be a matrix whose rows and columns are indexed by left shapes $\sigma, \sigma' \in \mathcal{L}$.

We say that a coefficient matrix $H$ is SOS-symmetric if $H(\sigma, \sigma')$ is invariant under permuting the vertices of $U_\sigma$ and permuting the vertices of $U_{\sigma'}$ (*more precisely, for the general case we permute the vertices within each index shape piece of $U_\sigma$ and permute the vertices within each index shape piece of $U_{\sigma'}$).

Definition 7.70. Given a shape $\tau$, we say that a coefficient matrix $H$ is a $\tau$-coefficient matrix if $H(\sigma, \sigma') = 0$ whenever $V_\sigma \not\equiv U_\tau$ or $V_\tau \not\equiv U_{\sigma'}^T$.
Definition 7.71. Given an index shape $U$, we define $\text{Id}_U$ to be the shape with $\text{Id}_U = V_{\text{Id}_U} = U$, no other vertices, and no edges.

Given a shape $\tau$ and a $\tau$-coefficient matrix $H$, we create two different matrix-valued functions, $M_{\tau}^{\text{fact}}(H)$ and $M_{\tau}^{\text{orth}}(H)$. As we will see, we can express $\Lambda$ in terms of $M^{\text{orth}}$ but to show PSDness we will need to shift to $M^{\text{act}}$. We analyze the difference between $M^{\text{fact}}$ and $M^{\text{orth}}$ in subsections 8.2, 8.3, and 8.4.

Definition 7.72. Given a shape $\tau$ and a $\tau$-coefficient matrix $H$, define

$$M_{\tau}^{\text{fact}}(H) = \sum_{\sigma \in \mathcal{L}_{\tau}, \sigma' \in \mathcal{L}_{\tau}} H(\sigma, \sigma') M_{\tau} M_{\tau}^T.$$ 

Proposition 7.73. For all $A$ and $B$ with shapes in $\mathcal{I}(\Lambda)$,

$$\left( M_{\tau}^{\text{fact}}(H) \right)(A, B) = \sum_{\sigma \in \mathcal{L}_{\tau}, \sigma' \in \mathcal{L}_{\tau}} H(\sigma, \sigma') \sum_{A', B'} \sum_{R_1 \in \mathcal{R}(\tau, A, A'), R_2 \in \mathcal{R}(\tau, A', B'), R_3 \in \mathcal{R}(\sigma', B', B)} M_{R_1}(A, A') M_{R_2}(A', B') M_{R_3}(B', B).$$

If $R_1, R_2, R_3$ are properly composable then $R = R_1 \circ R_2 \circ R_3$ has the expected shape $\sigma \circ \tau \circ \sigma^T$. Otherwise, $R_1 \circ R_2 \circ R_3$ will have a different shape. We define $M_{\tau}^{\text{orth}}(H)$ to be the same sum as $M_{\tau}^{\text{fact}}(H)$ except that it is restricted to properly composable ribbons $R_1, R_2, R_3$.

Definition 7.74. We define $M_{\tau}^{\text{orth}}(H)$ so that for all $A$ and $B$ with shapes in $\mathcal{I}(\Lambda)$,

$$\left( M_{\tau}^{\text{orth}}(H) \right)(A, B) = \sum_{\sigma \in \mathcal{L}_{\tau}, \sigma' \in \mathcal{L}_{\tau}} H(\sigma, \sigma') \sum_{A', B'} \sum_{R_1 \in \mathcal{R}(\sigma', B', B), R_3 \in \mathcal{R}(\sigma', B', B)} M_{R_1}(A, A') M_{R_2}(A', B') M_{R_3}(B', B)$$

$$= \sum_{\sigma \in \mathcal{L}_{\tau}, \sigma' \in \mathcal{L}_{\tau}} H(\sigma, \sigma') \sum_{A', B'} \sum_{R_1 \in \mathcal{R}(\sigma', B', B), R_3 \in \mathcal{R}(\sigma', B', B)} M_{R_1 \circ R_2 \circ R_3}(A, B)$$

It would be nice if we had that $M_{\tau}^{\text{orth}}(H) = \sum_{\sigma \in \mathcal{R}_{\tau}, \sigma' \in \mathcal{R}_{\tau}} H(\sigma, \sigma') M_{\sigma \circ \tau \circ \sigma^T}$. However, this is not quite correct because there is an additional term related to automorphism groups.

Definition 7.75. Given a shape $\alpha$, define $\text{Aut}(\alpha)$ to be the set of mappings from $\alpha$ to itself which keep $U_{\alpha}$ and $V_{\alpha}$ fixed.

Definition 7.76. Given composable shapes $\sigma, \tau, \sigma^T$, we define

$$\text{Decomp}(\sigma, \tau, \sigma^T) = \text{Aut}(\sigma \circ \tau \circ \sigma^T) / (\text{Aut}(\sigma) \times \text{Aut}(\tau) \times \text{Aut}(\sigma^T))$$

Remark 7.77. Each element $\pi \in \text{Decomp}(\sigma, \tau, \sigma^T)$ decomposes $\sigma \circ \tau \circ \sigma^T$ into $\sigma$, $\tau$, and $\sigma^T$ by specifying copies $\pi(\sigma)$, $\pi(\tau)$, $\pi(\sigma^T)$ of $\sigma$, $\tau$, and $\sigma^T$ such that $\pi(\sigma) \circ \pi(\tau) \circ \pi(\sigma^T) = \pi(\sigma \circ \tau \circ \sigma^T) = \sigma \circ \tau \circ \sigma^T$. Thus, $|\text{Decomp}(\sigma, \tau, \sigma^T)|$ is the number of ways to decompose $\sigma \circ \tau \circ \sigma^T$ into $\sigma$, $\tau$, and $\sigma^T$. 

54
Lemma 7.78.  
\[ M^\text{orth}_\tau(H) = \sum_{\sigma \in \mathcal{L}_U, \sigma' \in \mathcal{L}_V} H(\sigma, \sigma') |\text{Decomp}(\sigma, \tau, \sigma'^T)| M_{\sigma \otimes \tau \otimes \sigma'^T} \]

Proof sketch. Observe that there is a bijection between ribbons \( R \) with shape \( \sigma \circ \tau \circ \sigma'^T \) together with an element \( \pi \in \text{Decomp}(\sigma, \tau, \sigma') \) and triples of ribbons \((R_1, R_2, R_3)\) such that

1. \( R_1, R_2, R_3 \) have shapes \( \sigma, \tau, \) and \( \sigma'^T \), respectively.
2. \( V(R_1) \cap V(R_2) = A_{R_2} = B_{R_1}, V(R_2) \cap V(R_3) = A_{R_3} = B_{R_2}, \) and \( V(R_1) \cap V(R_3) = A_{R_2} \cap B_{R_2} \)

To see this, note that given such ribbons \( R_1, R_2, R_3 \), the ribbon \( R = R_1 \circ R_2 \circ R_3 \) has shape \( \sigma \circ \tau \circ \sigma'^T \) and the ribbons \( R_1, R_2, R_3 \) specify a decomposition of \( \sigma \circ \tau \circ \sigma'^T \) into \( \sigma, \tau, \) and \( \sigma'^T \).

Conversely, given \( R \) and an element \( \pi \in \text{Decomp}(\sigma, \tau, \sigma') \), \( \pi \) specifies how to decompose \( R \) into ribbons \( R_1, R_2, R_3 \) of shapes \( \sigma, \tau, \) and \( \sigma'^T \).

For a more rigorous proof, see Appendix B.

\[ \square \]

Remark 7.79. As this lemma shows, we have to be very careful about symmetry groups in our analysis. For accuracy, it is safest to check that the coefficients for each individual ribbon match.

Given a matrix-valued function \( \Lambda \), we can associate coefficient matrices to \( \Lambda \) as follows:

Definition 7.80. Given a matrix-valued function \( \Lambda = \sum_{\alpha, \lambda} \lambda \alpha M_\alpha \),

1. For each index shape \( U \in \mathcal{I}_{\text{mid}} \) and every \( \sigma, \sigma' \in \mathcal{L}_U \), we take \( H_{\text{Id}_{\lambda U}}(\sigma, \sigma') = \frac{1}{|\text{Aut}(U)|^{\lambda}} M_{\sigma \otimes \sigma'^T} \)
2. For each \( U \in \mathcal{I}_{\text{mid}}, \tau \in \mathcal{M}_U \) and \( \sigma, \sigma' \in \mathcal{L}_U \), we take \( H_T(\sigma, \sigma') = \frac{1}{|\text{Aut}(U)| |\text{Aut}(V)|^{\lambda}} M_{\sigma \otimes \tau \otimes \sigma'^T} \)

Lemma 7.81. \( \Lambda = \sum_{U \in \mathcal{I}_{\text{mid}}} M^\text{orth}_{\text{Id}_{\lambda U}}(H_{\text{Id}_{\lambda U}}) + \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_U} M^\text{orth}_T(H_T) \)

Proof. We check that the coefficients for each individual ribbon \( R \) match. There are two cases to consider.

If \( R \) has shape \( \alpha \) where \( \alpha \) has a unique minimum vertex separator \( S \), then there is a bijection between orderings \( O_S \) for \( S \) and pairs of ribbons \( R_1, R_2 \) such that \( R_1 \circ R_2 = R \) and the shapes \( \sigma, \sigma'^T \) of \( R_1, R_2 \) are left and right shapes respectively.

To see this, observe that when we concatenate \( R_1 \) and \( R_2 \), this assigns the matrix index \( B_{R_1} = A_{R_2} \) to \( S \), which is equivalent to specifying an ordering \( O_S \) for \( S \). Conversely, given an ordering \( O_S \) for \( S \), we take \( R_1 \) to be the part of \( R \) between \( A_R \) and \( (S, O_S) \) and we take \( R_2 \) to be the part of \( R \) between \( (S, O_S) \) and \( B_R \).

From this bijection, it follows that the coefficient of \( M_R \) is \( \lambda_\alpha \) on both sides of the equation.

Similarly, if \( R \) has shape \( \alpha \) where \( \alpha \) does not have a unique minimal vertex separator, then there is a bijection between orderings \( O_S, O_T \) for the leftmost and rightmost minimum vertex separators \( S, T \) of \( R \) and triples of ribbons \( R_1, R_2, R_3 \) such that \( R_1 \circ R_2 \circ R_3 = R \) and the shapes \( \sigma, \tau, \sigma'^T \) of \( R_1, R_2, R_3 \) are left, proper middle, and right shapes respectively.

To see this, observe that when we concatenate \( R_1, R_2, \) and \( R_3 \), this assigns the matrix index \( B_{R_1} = A_{R_2} \) to \( S \) and assigns the matrix index \( B_{R_2} = A_{R_3} \) to \( T \), which is equivalent to specifying orderings \( O_S, O_T \) for \( S, T \). Conversely, given orderings \( O_S, O_T \) for \( S, T \), we take \( R_1 \) to be the part of \( R \) between \( A_R \) and \( (S, O_S) \),
we take \( R_2 \) to be the part of \( R \) between \((S,O_S)\) and \((T,O_T)\), and we take \( R_2 \) to be the part of \( R \) between \((T,O_T)\) and \( B_R \).

From this bijection, it again follows that the coefficient of \( M_R \) is \( \lambda_n \) on both sides of the equation.

7.9 The \(-\gamma, -\gamma\) operation and qualitative theorem statement

In the intersection term analysis (see subsections 8.2, 8.3, and 8.4), we will need to further decompose left shapes \( \sigma \) as \( \sigma = \sigma_2 \circ \gamma \) where \( \sigma_2 \) and \( \gamma \) are themselves left shapes. Accordingly, we make the following definitions

**Definition 7.82.** Given a moment matrix \( \Lambda \), we define the following sets of left shapes:

1. \( \Gamma = \{ \gamma : \gamma \text{ is a non-trivial left shape, } U_\gamma, V_\gamma \in \mathcal{I}_{mid}, |V(\gamma)| \leq D_v, \forall e \in E(\gamma), l_e \leq D_E \} \)
2. Given \( U, V \in \mathcal{I}_{mid} \) such that \( w(U) > w(V) \), define \( \Gamma_{U,V} = \{ \gamma \in \Gamma : U_\gamma \equiv U, V_\gamma \equiv V \} \).
3. Given \( U \in \mathcal{I}_{mid} \), define \( \Gamma_{U,\ast} = \{ \gamma \in \Gamma : U_\gamma \equiv U \} \).
4. Given \( V \in \mathcal{I}_{mid} \), define \( \Gamma_{\ast,V} = \{ \gamma \in \Gamma : V_\gamma \equiv V \} \).

**Remark 7.83.** Under our simplifying assumptions, \( \Gamma \) is the same as \( \mathcal{L} \) except that \( \Gamma \) excludes the trivial shapes. In general, while \( \mathcal{L} \) requires that \( U_\gamma \in \mathcal{I}(\Lambda) \), \( \Gamma \) requires that \( U_\gamma \in \mathcal{I}_{mid} \). Note that \( \mathcal{I}(\Lambda) \) and \( \mathcal{I}_{mid} \) may be incomparable because

1. There may be index shapes \( U \in \mathcal{I}_{mid} \) such that no matrix index of \( \Lambda \) has shape \( U \).
2. All index shape pieces \( U_i \) for index shapes \( U \in \mathcal{I}_{mid} \) must have \( p_i = 1 \) while this is not the case for \( \mathcal{I}(\Lambda) \).

We now state our theorem qualitatively after giving one more definition.

**Definition 7.84.** Given a shape \( \tau \), left shapes \( \gamma \in \Gamma_{\ast,U} \), and \( \gamma' \in \Gamma_{\ast,V} \), and a \( \tau \)-coefficient matrix \( H \), define \( H^{-\gamma,\gamma'} \) to be the \((\gamma \circ \tau \circ \gamma'^T)\)-coefficient matrix with entries

1. \( H^{-\gamma,\gamma'}(\sigma,\sigma') = H(\sigma \circ \gamma, \sigma' \circ \gamma') \) if \( |V(\sigma \circ \gamma)| \leq D_v \) and \( |V(\sigma' \circ \gamma')| \leq D_v \).
2. \( H^{-\gamma,\gamma'}(\sigma,\sigma') = 0 \) if \( |V(\sigma \circ \gamma)| > D_v \) or \( |V(\sigma' \circ \gamma')| > D_v \).

**Remark 7.85.** For the theorem, we will only need the case when \( \gamma' = \gamma \)

Our qualitative theorem statement is as follows:

**Theorem 7.86.** Let \( \Lambda = \sum_{U \in \mathcal{I}_{mid}} M_{id_{U}}^{orth}(H_{id_{U}}) + \sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in \mathcal{M}_{U}} M_{\tau}^{orth}(H_{\tau}) \) be an SOS-symmetric matrix valued function.

There exist functions \( f(\tau) \) and \( f(\gamma) \) depending on \( n \) and other parameters such that if the following conditions hold:

1. For all \( U \in \mathcal{I}_{mid}, H_{id_{U}} \succeq 0 \)
2. For all \( U \in \mathcal{I}_{mid} \) and all \( \tau \in \mathcal{M}_{U} \),

\[
\begin{bmatrix}
H_{id_{U}} & f(\tau)H_{\tau} \\
 f(\tau)H_{\tau}^T & H_{id_{U}}
\end{bmatrix} \succeq 0
\]
3. For all $U, V \in \mathcal{I}_{\text{mid}}$ where $\omega(U) > \omega(V)$ and all $\gamma \in \Gamma_{U,V}$, $H_{\text{Id}_V}^{-\gamma} \preceq f(\gamma)H_{\text{Id}_U}$

then with high probability $\Lambda \succeq 0$

**Remark 7.87.** Roughly speaking, conditions 1 and 2 give us an approximate PSD decomposition for the moment matrix $M$. Condition 3 comes from the intersection term analysis, which is the most technically intensive part of the proof.

### 7.10 Quantitative theorem statement

To state our theorem quantitatively, we will need a few more things. First, the conditions of the theorem will involve functions $B_{\text{norm}}(\alpha), B(\gamma), N(\gamma)$, and $c(\alpha)$. Roughly speaking, these functions will be used as follows in the analysis:

1. $B_{\text{norm}}(\alpha)$ will bound the norms of the matrices $M_\alpha$
2. $B(\gamma)$ and $N(\gamma)$ will help us bound the intersection terms (see Section 8.4).
3. $c(\alpha)$ will help us sum over the possible $\gamma$ and $\tau$.

Second, for technical reasons it turns out that comparing $H^{-\gamma}_{\text{Id}_V}$ to $H_{\text{Id}_U}$ doesn’t quite work. Instead, we compare $H^{-\gamma}_{\text{Id}_V}$ to a matrix $H'_\gamma$ of our choice where $H'_\gamma$ is very close to $H_{\text{Id}_U}$ (up to truncation error).

**Definition 7.88.** Given a function $B_{\text{norm}}(\alpha)$, we define the distance $d_\tau(H, H')$ between two $\tau$-coefficient matrices $H$ and $H'$ to be

$$d_\tau(H, H') = \sum_{\sigma \in \mathcal{L}_{U,\tau}, \sigma' \in \mathcal{L}_{V,\tau}} |H'_\tau(\sigma, \sigma') - H_\tau(\sigma, \sigma')|B_{\text{norm}}(\sigma)B_{\text{norm}}(\tau)B_{\text{norm}}(\sigma')$$

Third, we need an SOS-symmetric analogue of the identity matrix.

**Definition 7.89.** We define $I_{\text{Sym}}$ to be the matrix such that

1. The rows and columns of $I_{\text{Sym}}$ are indexed by the matrix indices $A, B$ whose index shape is in $\mathcal{I}(\Lambda)$.
2. $I_{\text{Sym}}(A, B) = 1$ if $p_A = p_B$ and $I_{\text{Sym}}(A, B) = 0$ if $p_A \neq p_B$.

**Proposition 7.90.** If $M$ has SOS-symmetry and the rows and columns of $I_{\text{Sym}}$ are indexed by matrix indices $A, B$ whose index shape is in $\mathcal{I}(\Lambda)$ then $M \preceq \|M\| I_{\text{Sym}}$

**Corollary 7.91.** For all $\tau$ and all SOS-symmetric $\tau$-coefficient matrices $H_\tau$ and $H'_\tau$,

$$M_{\tau}^{\text{fact}}(H'_\tau) + M_{\tau}^{\text{fact}}(H'^{\tau}_\tau) - M_{\tau}^{\text{fact}}(H_\tau) - M_{\tau}^{\text{fact}}(H^{\tau}_\tau) \preceq 2d_\tau(H_\tau, H'_\tau)I_{\text{Sym}}$$

Note that if $\tau, H_\tau$ and $H'_\tau$ are all symmetric then

$$M_{\tau}^{\text{fact}}(H'_\tau) - M_{\tau}^{\text{fact}}(H_\tau) \preceq d_\tau(H_\tau, H'_\tau)I_{\text{Sym}}$$

Finally, we need a few more definitions about shapes $\alpha$. 57
Definition 7.92 ($\mathcal{M}'$). We define $\mathcal{M}'$ to be the set of all shapes $\alpha$ such that

1. $|V(\alpha)| \leq 3D_V$
2. $\forall e \in E(\alpha), l_e \leq D_E$
3. All edges $e \in E(\alpha)$ have multiplicity at most $3D_V$.

Definition 7.93 ($S_\alpha$). Given a shape $\alpha$, define $S_\alpha$ to be the leftmost minimum vertex separator of $\alpha$.

Definition 7.94 ($I_\alpha$). Given a shape $\alpha$, define $I_\alpha$ to be the set of vertices in $V(\alpha) \setminus (U_\alpha \cup V_\alpha)$ which are isolated.

We can now state our main theorem.

Theorem 7.95. Given the moment matrix $\Lambda = \sum_{U \in \mathcal{I}_{mid}} M_{Idu}^{orth}(H_{Idu}) + \sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in \mathcal{M}_U} M_\tau^{orth}(H_\tau)$, for all $\varepsilon > 0$, if we take

1. $q = 3 \left[ D_V \ln(n) + \frac{\ln(\frac{1}{\varepsilon})}{3} + D_V \ln(5) + 3D_V^2 \ln(2) \right]$
2. $B_{\text{vertex}} = 6D_V \sqrt{2eq}$
3. $B_{\text{norm}}(\alpha) = B_{\text{vertex}}^{\frac{|V(\alpha)\setminus U_\alpha| + |V(\alpha)\setminus V_\alpha|}{n} \frac{\omega(\alpha) + \omega(U_\alpha) - \omega(S_\alpha)}{2}}$
4. $B(\gamma) = B_{\text{vertex}}^{\frac{|V(\gamma)\setminus V_\alpha| + |V(\gamma)\setminus U_\alpha|}{n} \frac{\omega(V(\gamma)\setminus U_\alpha)}{2}}$
5. $N(\gamma) = (3D_V)^2 |V(\gamma)\setminus V_\alpha| + |V(\gamma)\setminus U_\alpha|$
6. $c(\alpha) = 100(3D_V)^2 |U_\alpha\setminus V_\alpha| + |V_\alpha\setminus U_\alpha| + 2|E(\alpha)| 2|V(\alpha)\setminus (U_\alpha \cup V_\alpha)|$

and we have SOS-symmetric coefficient matrices $\{H'_\gamma : \gamma \in \Gamma\}$ such that the following conditions hold:

1. For all $U \in \mathcal{I}_{mid}$, $H_{Idu} \succeq 0$
2. For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, $B_{\text{norm}}(\tau) H_\tau \succeq 0$
3. For all $U, V \in \mathcal{I}_{mid}$ where $\omega(U) > \omega(V)$ and all $\gamma \in \Gamma_{U,V}$, $c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Idu}^{-1}\gamma \succeq H'_\gamma$

then with probability at least $1 - \varepsilon$,

$$\Lambda \succeq \frac{1}{2} \left( \sum_{U \in \mathcal{I}_{mid}} M_{Idu}^{fact}(H_{Idu}) \right) - 3 \left( \sum_{U \in \mathcal{I}} \sum_{\gamma \in \Gamma_{U,V}} d_{Idu}(H'_\gamma, H_{Idu}) \right) \text{Id}_{\text{sym}}$$

If it is also true that whenever $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}'$,

$$\sum_{U \in \mathcal{I}_{mid}} M_{Idu}^{fact}(H_{Idu}) \succeq 6 \left( \sum_{U \in \mathcal{I}} \sum_{\gamma \in \Gamma_{U,V}} d_{Idu}(H'_\gamma, H_{Idu}) \right) \text{Id}_{\text{sym}}$$

then with probability at least $1 - \varepsilon$, $\Lambda \succeq 0$.  

58
7.10.1 General Main Theorem

Before stating the general main theorem, we need to modify a few definitions for $\alpha$ and give a few definitions for $\Omega$

**Definition 7.96** ($S_{a,\text{min}}$ and $S_{a,\text{max}}$). Given a shape $\alpha \in \mathcal{M}'$, define $S_{a,\text{min}}$ to be the leftmost minimum vertex separator of $\alpha$ if all edges with multiplicity at least 2 are deleted and define $S_{a,\text{max}}$ to be the leftmost minimum vertex separator of $\alpha$ if all edges with multiplicity at least 2 are present.

**Definition 7.97** (General $I_\alpha$). Given a shape $\alpha$, define $I_\alpha$ to be the set of vertices in $V(\alpha) \setminus (U_\alpha \cup V_\alpha)$ such that all edges incident with that vertex have multiplicity at least 2.

**Definition 7.98** ($B_\Omega$). We take $B_\Omega(j)$ to be a non-decreasing function such that for all $j \in \mathbb{N}$, $E_\Omega[x^j] \leq B_\Omega(j)^j$.

**Definition 7.99** ($h^+_j$). For all $j$, we define $h^+_j$ to be the polynomial $h_j$ where we make all of the coefficients have positive sign.

**Lemma 7.100.** If $\Omega = N(0, 1)$ then we can take $B_\Omega(j) = \sqrt{j}$ and we have that

$$h^+_j(x) \leq \frac{1}{\sqrt{j}}(x^2 + j)^{\frac{1}{2}} \leq \left(\frac{e}{j}(x^2 + j)\right)^{\frac{1}{2}}$$

For a proof, see [AMP20, Lemma 8.15].

**Theorem 7.101.** Given the moment matrix $\Lambda = \sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{orth}}^H(H_{ldU}) + \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_U} M_{\tau}^H(H_\tau)$, for all $\epsilon > 0$, if we take

1. $q = \left[3D_V\ln(n) + \ln\left(\frac{1}{\epsilon}\right) + (3D_V)^k\ln(D_E + 1) + 3D_V\ln(5)\right]$
2. $B_{\text{vertex}} = 6qD_V$
3. $B_{\text{edge}}(e) = 2h^+_j(B_\Omega(6D_VD_E)) \max_{f \in [0, 3D_VD_E]} \left\{ \left( h^+_j(B_\Omega(2lj)) \right) \right\}$
   
   As a special case, if $\Omega = N(0, 1)$ then we can take $B_{\text{edge}}(e) = (400D_V^2D_E^2)^{lj}$
4. $B_{\text{norm}}(\alpha) = 2eB_{\text{vertex}}^{[V(\alpha)\cup U_\alpha]+[V(\alpha)\setminus V_\alpha]} \left( \prod_{e \in E(\alpha)} B_{\text{edge}}(e) \right) n_{\frac{\nu(\alpha)+\nu(\alpha)-\nu(S_{a,\text{min}})}}$
5. $B(\gamma) = B_{\text{vertex}}^{[V(\gamma)\cup U_\gamma]+[V(\gamma)\setminus V_\gamma]} \left( \prod_{e \in E(\gamma)} B_{\text{edge}}(e) \right) n_{\frac{\nu(\gamma)+\nu(\gamma)-\nu(S_{a,\text{min}})}}$
6. $N(\gamma) = (3D_V)^2[\nu(\gamma)\setminus V_\gamma]+[\nu(\gamma)\cup U_\gamma]$
7. $c(\alpha) = 100(3l_{\text{max}}D_V)^{[U_\alpha\cup V_\alpha]+[V_\alpha\cup U_\alpha]+k[\nu(\alpha)+2l_{\text{max}}]}[V(\alpha)\setminus (U_\alpha \cup V_\alpha)]$

and we have SOS-symmetric coefficient matrices $\{H'_\gamma : \gamma \in \Gamma\}$ such that the following conditions hold:

1. For all $U \in \mathcal{I}_{\text{mid}}$, $H_{ldU} \succeq 0$
2. For all $U \in \mathcal{I}_{\text{mid}}$ and $\tau \in \mathcal{M}_U$, $\frac{1}{\vert\text{Aut}(U)\vert c(\tau)} H_{ldU}^T B_{\text{norm}}(\tau) H_\tau \frac{1}{\vert\text{Aut}(U)\vert c(\tau)} H_{ldU} \succeq 0$
3. For all $U,V \in \mathcal{I}_{\text{mid}}$ where $\omega(U) > \omega(V)$ and all $\gamma \in \Gamma_{U,V}$,
\[
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{\text{Id}_{U,V}}^{-T \gamma} \leq H_{\gamma}'
\]
then with probability at least $1 - \varepsilon$,
\[
\Lambda \geq \frac{1}{2} \left( \sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_{U}}^{\text{fact}}(H_{\text{Id}_{U}}) \right) - 3 \left( \sum_{U \in \mathcal{I}} \sum_{\gamma \in \Gamma_{U,V}} \frac{d_{\text{Id}_{U}}(H_{\gamma}' \gamma H_{\text{Id}_{U}})}{|\text{Aut}(U)|c(\gamma)} \right) \text{Id}_{\text{sym}}
\]
If it is also true that whenever $\|M_{\alpha}\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}'$,
\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_{U}}^{\text{fact}}(H_{\text{Id}_{U}}) \geq 6 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,V}} \frac{d_{\text{Id}_{U}}(H_{\gamma}' \gamma H_{\text{Id}_{U}})}{|\text{Aut}(U)|c(\gamma)} \right) \text{Id}_{\text{sym}}
\]
then with probability at least $1 - \varepsilon$, $\Lambda \geq 0$.

7.11 Choosing $H_{\gamma}'$ and Truncation Error

A canonical choice for $H_{\gamma}'$ is to take

1. $H_{\gamma}'(\sigma,\sigma') = H_{\text{Id}_{U}}(\sigma,\sigma')$ whenever $|V(\sigma \gamma)| \leq D_V$ and $|V(\sigma' \gamma)| \leq D_V$.
2. $H_{\gamma}'(\sigma,\sigma') = 0$ whenever $|V(\sigma \gamma)| > D_V$ or $|V(\sigma' \gamma)| > D_V$.

With this choice, the truncation error is
\[
d_{\text{Id}_{U}}(H_{\text{Id}_{U}}', H_{\gamma}') = \sum_{\sigma,\sigma' \in \mathcal{U}, \gamma \in \Gamma_{U,V}} |V(\sigma \gamma) - D_V| \text{ and } |V(\sigma' \gamma) - D_V|
\]

8 Proof of the Main Theorem

In this section, we prove the main theorem under the assumption that the functions $B_{\text{norm}}(\alpha)$, $B(\gamma)$, $N(\gamma)$, and $c(\gamma)$ have certain properties. More precisely, we prove the following theorem.

**Theorem 8.1.** For all $\varepsilon > 0$ and all $\varepsilon' \in (0, \frac{1}{25}]$, for any moment matrix
\[
\Lambda = \sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_{U}}^{\text{orth}}(H_{\text{Id}_{U}}) + \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}} M_{\tau}^{\text{orth}}(H_{\tau})
\]
if $B_{\text{norm}}(\alpha)$, $B(\gamma)$, $N(\gamma)$, and $c(\alpha)$ are functions such that

1. With probability at least $(1 - \varepsilon)$, for all shapes $\alpha \in \mathcal{M}'$, $\|M_{\alpha}\| \leq B_{\text{norm}}(\alpha)$.
2. For all $\tau \in \mathcal{M}'$, $\gamma \in \Gamma_{\tau,V}$, $\gamma' \in \Gamma_{\tau,V'}$, and all intersection patterns $P \in \mathcal{P}_{\tau,V}$,
\[
B_{\text{norm}}(\tau_P) \leq B(\gamma)B(\gamma')B_{\text{norm}}(\tau)
\]

Note: Intersection patterns and $\mathcal{P}_{\tau,V}$ will be defined later, see Definitions 8.8 and 8.9.
3. For all composable $\gamma_1, \gamma_2$, $B(\gamma_1)B(\gamma_2) = B(\gamma_1 \circ \gamma_2)$.
4. $\forall U \in I_{mid}, \sum_{\gamma \in \Gamma_{U,\gamma}} \frac{1}{|Aut(U)\mid c(\gamma)} < \epsilon'$
5. $\forall V \in I_{mid}, \sum_{\gamma \in \Gamma_{+V}} \frac{1}{|Aut(U)\mid c(\gamma)} < \epsilon'$
6. $\forall U \in I_{mid}, \sum_{\tau \in M_{U}} \frac{1}{|Aut(U)\mid c(\tau)} < \epsilon'$
7. For all $\tau \in M'$, $\gamma \in \Gamma_{+U} \cup \{Id_U\}$, and $\gamma' \in \Gamma_{+V} \cup \{Id_V\}$,

$$\sum_{j>0} \sum_{\tau_1,\tau_2,\ldots,\tau_j \in \Gamma_{+\tau} \mid \gamma_i \text{ is non-trivial}} \frac{1}{|Aut(U)\mid} \prod_{i} \frac{1}{|Aut(U)\mid} \sum_{p_i \in P_{\gamma_2}, \gamma, \gamma'} \left( \prod_{i=1}^{j} N(P_i) \right)$$

$$\leq \frac{N(\gamma)N(\gamma')}{(|Aut(U)\mid)^{\gamma} \text{ is non-trivial}}$$

Note: $\Gamma_{\gamma,\gamma'}$ will be defined later, see Definition 8.18.

and we have SOS-symmetric coefficient matrices $\{H'_{\gamma} : \gamma \in \Gamma\}$ such that the following conditions hold:
1. For all $U \in I_{mid}, H_{Id_U} \geq 0$
2. For all $U \in I_{mid}$ and $\tau \in M_{U}$,

$$\left[ \frac{1}{|Aut(U)\mid c(\tau)} H_{Id_U} \quad B_{\text{norm}}(\tau)H_{\tau} \right] \geq 0$$

3. For all $U, V \in I_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U, V}$,

$$c(\gamma)^2 N(\gamma)B(\gamma)^2 H_{Id_U}^{\gamma,\gamma'} \notin H_{\gamma}$$

then with probability at least $1 - \epsilon$,

$$\Lambda \geq \frac{1}{2} \left( \sum_{U \in I_{mid}} M_{Id_U}^{\text{fact}}(H_{Id_U}) - 3 \sum_{U \in I_{mid}} \sum_{\gamma \in \Gamma_{U,\gamma}} \frac{d_{Id_U}(H_{\gamma}, H_{Id_U})}{|Aut(U)\mid c(\gamma)} \right) Id_{sym}$$

If it is also true that whenever $|M_{\alpha}| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in M'$,

$$\sum_{U \in I_{mid}} M_{Id_U}^{\text{fact}}(H_{Id_U}) \geq 6 \left( \sum_{U \in I_{mid}} \sum_{\gamma \in \Gamma_{U,\gamma}} \frac{d_{Id_U}(H_{\gamma}, H_{Id_U})}{|Aut(U)\mid c(\gamma)} \right) Id_{sym}$$

then with probability at least $1 - \epsilon$, $\Lambda \geq 0$.

Throughout this section, we assume that we have functions $B_{\text{norm}}(\alpha), B(\gamma), N(\gamma), \text{ and } c(\gamma)$. If $\forall \alpha \in M', |M_{\alpha}| \leq B_{\text{norm}}(\alpha)$ then we say that the norm bounds hold. For the other properties of these functions, we will either restate these properties in our intermediate results to highlight where these properties are needed or just state that the conditions on these functions are satisfied for brevity.
8.1 Warm-up: Analysis with no intersection terms

In this subsection, we show how the analysis works if we ignore the difference between $M^{\text{fact}}$ and $M^{\text{orth}}$.

**Theorem 8.2.** For all $\epsilon' \in (0, \frac{1}{2}]$, if the norm bounds hold and the following conditions hold

1. For all $U \in \mathcal{I}_{\text{mid}}$, $H_{Id_U} \succeq 0$
2. For all $U \in \mathcal{I}_{\text{mid}}$ and all $\tau \in \mathcal{M}_U$
   
   $\begin{bmatrix}
   \frac{1}{|\text{Aut}(U)|}H_{Id_U} & \frac{1}{|\text{Aut}(U)|}H_{Id_U}^T
   
   B_{\text{norm}}(\tau)H_{\tau} & B_{\text{norm}}(\tau)H_{\tau}^T
   \end{bmatrix} \succeq 0$
3. $\forall U \in \mathcal{I}_{\text{mid}}, \sum_{\tau \in \mathcal{M}_U} \frac{1}{|\text{Aut}(U)|} \leq \epsilon'$.

then

$$\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{fact}}^{\text{Id}_U}(H_{Id_U}) + \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_U} M_{\text{fact}}^{\text{Id}_U}(H_{\tau}) \succeq (1 - 2\epsilon') \sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{fact}}^{\text{Id}_U}(H_{Id_U}) \succeq 0$$

**Proof.** We first show how a single term $M_\sigma M_\tau M_{\sigma'^T}$ plus its transpose $M_\sigma' M_\tau M_{\sigma'^T}$ can be bounded.

**Lemma 8.3.** If the norm bounds hold then for all $\tau \in \mathcal{M}^l$ and shapes $\sigma, \sigma'$ such that $\sigma, \tau, \sigma'^T$ are composable, for all $a, b$ such that $a > 0, b > 0$, and $ab = B_{\text{norm}}(\tau)^2$,

$$M_\sigma M_\tau M_{\sigma'^T} + M_{\sigma'} M_{\tau} M_{\sigma'^T} \preceq a M_\sigma M_{\sigma'^T} + b M_{\sigma'} M_{\sigma'^T}$$

**Proof.** Observe that

$$0 \preceq \begin{bmatrix} \sqrt{a} M_\sigma - \frac{\sqrt{b}}{B_{\text{norm}}(\tau)} M_{\sigma'} M_{\tau} \\ \sqrt{a} M_{\sigma'^T} - \frac{\sqrt{b}}{B_{\text{norm}}(\tau)} M_{\sigma'} M_{\tau} \end{bmatrix}^T =

\begin{bmatrix} \sqrt{a} M_\sigma - \frac{\sqrt{b}}{B_{\text{norm}}(\tau)} M_{\sigma'} M_{\tau} \\ \sqrt{a} M_{\sigma'^T} - \frac{\sqrt{b}}{B_{\text{norm}}(\tau)} M_{\sigma'} M_{\tau} \end{bmatrix} =

a M_\sigma M_{\sigma'^T} - M_\sigma M_{\tau} M_{\sigma'^T} - M_{\sigma'} M_{\tau} M_{\sigma'^T} + \frac{b}{B_{\text{norm}}(\tau)^2} M_{\sigma'} M_{\tau} M_{\sigma'^T} \preceq

a M_\sigma M_{\sigma'^T} - M_\sigma M_{\tau} M_{\sigma'^T} - M_{\sigma'} M_{\tau} M_{\sigma'^T} + \frac{b}{B_{\text{norm}}(\tau)^2} M_{\sigma'} (B_{\text{norm}}(\tau)^2 I_d) M_{\sigma'^T}$$

Thus, $M_\sigma M_\tau M_{\sigma'^T} + M_{\sigma'} M_{\tau} M_{\sigma'^T} \preceq a M_\sigma M_{\sigma'^T} + b M_{\sigma'} M_{\sigma'^T}$, as needed. \hfill \blacksquare

Unfortunately, if we try to bound everything term by term, there may be too many terms to bound. Instead, we generalize this argument for vectors and coefficient matrices.

**Definition 8.4.** Let $\tau$ be a shape. We say that a vector $v$ is a left $\tau$-vector if the coordinates of $v$ are indexed by left shapes $\sigma \in \mathcal{L}_{U_\tau}$. We say that a vector $w$ is a right $\tau$-vector if the coordinates of $w$ are indexed by left shapes $\sigma' \in \mathcal{L}_{\tau V}$. 

62
**Lemma 8.5.** For all $\tau \in \mathcal{M}'$, if the norm bounds hold, $v$ is a left $\tau$-vector, and $w$ is a right $\tau$-vector then

$$M^\text{fact}_\tau (vw^T) + M^\text{fact}_\tau (\omega \overline{\omega}^T) \leq B_{\text{norm}}(\tau) \left( M^\text{fact}_{\text{Id}_{\mathcal{U}_T}} (v\overline{v}^T) + M^\text{fact}_{\text{Id}_{\mathcal{V}_T}} (\omega \overline{\omega}^T) \right)$$

and

$$-M^\text{fact}_\tau (vw^T) - M^\text{fact}_\tau (\omega \overline{\omega}^T) \leq B_{\text{norm}}(\tau) \left( M^\text{fact}_{\text{Id}_{\mathcal{U}_T}} (v\overline{v}^T) + M^\text{fact}_{\text{Id}_{\mathcal{V}_T}} (\omega \overline{\omega}^T) \right)$$

**Proof.** Observe that

$$0 \leq \left( \sum_{\sigma} v_{\sigma} M_{\sigma} \right) \left( \sum_{\sigma} \frac{w_{\sigma} M_{\sigma}}{B_{\text{norm}}(\tau)} \right)^T = \sum_{\sigma} \left( \sum_{\sigma'} \frac{w_{\sigma} M_{\sigma}}{B_{\text{norm}}(\tau)} \right)^T \sum_{\sigma'} \left( \sum_{\sigma} v_{\sigma} M_{\sigma} \right) \frac{w_{\sigma} M_{\sigma}}{B_{\text{norm}}(\tau)} M_{\sigma'} + \frac{1}{B_{\text{norm}}(\tau)^2} \sum_{\sigma, \sigma'} (v_{\sigma} v_{\sigma'}) M_{\sigma} M_{\sigma'} M_{\sigma', T}$$

Further observe that

1. $\sum_{\sigma, \sigma'} (v_{\sigma} v_{\sigma'}) M_{\sigma} M_{\sigma', T} = M^\text{fact}_{\text{Id}_{\mathcal{U}_T}} (v\overline{v}^T)$
2. $\sum_{\sigma, \sigma'} (v_{\sigma} w_{\sigma'}) M_{\sigma} M_{\sigma', T} = M^\text{fact}_\tau (vw^T)$
3. $\sum_{\sigma, \sigma'} (w_{\sigma} v_{\sigma'}) M_{\sigma} M_{\sigma', T} = M^\text{fact}_\tau (\omega \overline{\omega}^T)$
4. $\sum_{\sigma, \sigma'} (w_{\sigma} w_{\sigma'}) M_{\sigma} M_{\sigma', T} M_{\sigma', T} = \left( \sum_{\sigma} w_{\sigma} M_{\sigma} \right) M_{\tau} \left( \sum_{\sigma} w_{\sigma} M_{\sigma} \right)^T \leq \left( \sum_{\sigma} w_{\sigma} M_{\sigma} \right) B_{\text{norm}}(\tau)^2 \text{Id} \left( \sum_{\sigma} w_{\sigma} M_{\sigma} \right)^T = B_{\text{norm}}(\tau)^2 \sum_{\sigma, \sigma'} (w_{\sigma} w_{\sigma'}) M_{\sigma} M_{\sigma', T} = B_{\text{norm}}(\tau)^2 M^\text{fact}_{\text{Id}_{\mathcal{V}_T}} (\omega \overline{\omega}^T)$

Putting everything together,

$$\frac{M^\text{fact}_\tau (vw^T) + M^\text{fact}_\tau (\omega \overline{\omega}^T)}{B_{\text{norm}}(\tau)} \leq M^\text{fact}_{\text{Id}_{\mathcal{U}_T}} (v\overline{v}^T) + M^\text{fact}_{\text{Id}_{\mathcal{V}_T}} (\omega \overline{\omega}^T)$$

and

$$\frac{-M^\text{fact}_\tau (vw^T) - M^\text{fact}_\tau (\omega \overline{\omega}^T)}{B_{\text{norm}}(\tau)} \leq M^\text{fact}_{\text{Id}_{\mathcal{U}_T}} (v\overline{v}^T) + M^\text{fact}_{\text{Id}_{\mathcal{V}_T}} (\omega \overline{\omega}^T)$$

as needed.

$\blacksquare$
Corollary 8.6. For all $\tau \in \mathcal{M}'$, if the norm bounds hold and $H_U$ and $H_V$ are matrices such that

\[
\begin{bmatrix}
H_U & B_{\text{norm}}(\tau) H_T \\
B_{\text{norm}}(\tau) H_T^T & H_V
\end{bmatrix} \succeq 0
\]

then $M^\text{fact}_\tau (H_\tau) + M^\text{fact}_\tau (H_\tau) \preceq M^\text{fact}_{\text{id}_U}(H_U) + M^\text{fact}_{\text{id}_V}(H_V)$

Proof. If \[
\begin{bmatrix}
H_U & B_{\text{norm}}(\tau) H_T \\
B_{\text{norm}}(\tau) H_T^T & H_V
\end{bmatrix} \succeq 0
\]
then we can write

\[
\begin{bmatrix}
H_U & B_{\text{norm}}(\tau) H_T \\
B_{\text{norm}}(\tau) H_T^T & H_V
\end{bmatrix} = \sum_i (v_i, w_i) (v_i, w_i)^T
\]

Since the $M^\text{fact}$ operations are linear, the result now follows by summing the equation

\[
M^\text{fact}_\tau (v_i w_i^T) + M^\text{fact}_\tau (w_i v_i^T) \preceq B_{\text{norm}}(\tau) \left( M^\text{fact}_{\text{id}_U}(v_i v_i^T) + M^\text{fact}_{\text{id}_V}(w_i w_i^T) \right)
\]

over all $i$.

Theorem 8.2 now follows directly. For all $U \in \mathcal{I}_\text{mid}$ and all $\tau \in \mathcal{M}_U$, using Corollary 8.6 with $H_U = H_V = \frac{1}{|\text{Aut}(U)| c(\tau)} H_{\text{id}_U}$,

\[
M^\text{fact}_\tau (H_\tau) + M^\text{fact}_\tau (H_\tau) \preceq \frac{1}{|\text{Aut}(U)| c(\tau)} M^\text{fact}_{\text{id}_U}(H_{\text{id}_U}) + \frac{1}{|\text{Aut}(U)| c(\tau)} M^\text{fact}_{\text{id}_U}(H_{\text{id}_U})
\]

Summing this equation over all $U \in \mathcal{I}_\text{mid}$ and all $\tau \in \mathcal{M}_U$, we obtain that

\[
\sum_{U \in \mathcal{I}_\text{mid}} \sum_{\tau \in \mathcal{M}_U} M^\text{fact}_\tau (H_\tau) \preceq 2\varepsilon' \sum_{U \in \mathcal{I}_\text{mid}} M^\text{fact}_{\text{id}_U}(H_{\text{id}_U})
\]

as needed.

### 8.2 Intersection Term Analysis Strategy

As we saw in the previous subsection, the analysis works out nicely if we work with $M^\text{fact}$. Unfortunately, our matrices are expressed in terms of $M^\text{orth}$. In this subsection, we describe our strategy for analyzing the difference between $M^\text{fact}$ and $M^\text{orth}$.

Recall the following expressions for \( M^\text{fact}_\tau (H)(A, B) \) and \( M^\text{orth}_\tau (H)(A, B) \) where $A$ has shape $U_\tau$ and $B$ has shape $V_\tau$:

\[
\left( M^\text{fact}_\tau (H) \right)(A, B) = \sum_{\sigma \in \mathcal{L}_U, \sigma' \in \mathcal{L}_V} H(\sigma, \sigma') \sum_{A', B'} \sum_{R_1, R_2, R_3} M_{R_1}(A, A') M_{R_2}(A', B') M_{R_3}(B', B)
\]

\[
\left( M^\text{orth}_\tau (H) \right)(A, B) = \sum_{\sigma \in \mathcal{L}_U, \sigma' \in \mathcal{L}_V} H(\sigma, \sigma') \sum_{A', B'} \sum_{R_1, R_2, R_3} M_{R_1}(A, A') M_{R_2}(A', B') M_{R_3}(B', B)
\]

\[
\text{where } R_1, R_2, R_3 \text{ are properly composable}
\]
Remark 8.7. A key feature of our analysis is that it will work the same way regardless of the shapes \( R_1, R_2, R_3 \) which are composable but not properly composable. These terms, which we call intersection terms, are not negligible and must be analyzed carefully. In particular, we decompose each resulting ribbon \( R = R_1 \circ R_2 \circ R_3 \) into new left, middle, and right parts. We do this as follows:

1. Let \( V_s \) be the set of vertices which appear more than once in \( V(R_1 \circ R_2 \circ R_3) \). In other words, \( V_s \) is the set of vertices involved in the intersections between \( R_1, R_2, \) and \( R_3 \) (not counting the facts that \( B_{R_1} = A_{R_2} \) and \( B_{R_2} = A_{R_3} \) because we expect these intersections).
2. Let \( A' \) be the leftmost minimum vertex separator of \( A_{R_1} \) and \( B_{R_1} \cup V_s \) in \( R_1 \). We turn \( A' \) into a matrix index by specifying an ordering \( O_{A'} \) for the vertices in \( A' \).
3. Let \( B' \) be the leftmost minimum vertex separator of \( A_{R_3} \cup V_s \) and \( B_{R_3} \) in \( R_2 \). We turn \( B' \) into a matrix index by specifying an ordering \( O_{B'} \) for the vertices in \( B' \).
4. Decompose \( R_1 \) as \( R_1 = R'_1 \cup R_4 \) where \( R'_1 \) is the part of \( R_1 \) between \( A_{R_1} \) and \( A' \) and \( R_4 \) is the part of \( R_1 \) between \( B' \) and \( B_{R_1} \). Similarly, decompose \( R_3 \) as \( R_3 = R'_3 \cup R_5 \) where \( R_5 \) is the part of \( R_3 \) between \( B' \) and \( B_{R_3} \).
5. Take \( R'_2 = R_4 \circ R_2 \circ R_5 \) and note that \( R'_1 \circ R'_2 \circ R'_3 = R_1 \circ R_2 \circ R_3 \). We view \( R'_1, R'_2, R'_3 \) as the left, middle, and right parts of \( R = R_1 \circ R_2 \circ R_3 \)

While we will verify our analysis by checking the coefficients of the ribbons, we want to express everything in terms of shapes. We use the following conventions for the names of the shapes:

1. As usual, we let \( \sigma, \tau, \) and \( \sigma'^T \) be the shapes of \( R_1, R_2, \) and \( R_3 \).
2. We let \( \gamma \) and \( \gamma'^T \) be the shapes of \( R_4 \) and \( R_5 \).
3. We let \( \sigma_2, \tau_p, \) and \( \sigma_2'^T \) be the shapes of \( R'_1, R'_2, \) and \( R'_3 \). Here \( P \) is the intersection pattern induced by \( R_4, R_2, \) and \( R_5 \) which we define in the next subsection.

**Remark 8.7.** A key feature of our analysis is that it will work the same way regardless of the shapes \( \sigma_2, \sigma_2'^T \) of \( R'_1 \) and \( R'_3 \). In other words, if we replace \( \sigma_2 \) by \( \sigma_2a \) and \( \sigma_2'^T \) by \( \sigma_2a'^T \) for a given intersection term, this just replaces \( \sigma = \sigma_2 \cup \gamma \) with \( \sigma_a = \sigma_2a \cup \gamma \) and \( \sigma'^T = \sigma'^T_a \cup \gamma'^T \) with \( \sigma'^T_a = \sigma'^T_{2a} \cup \gamma'^T \). This allows us to focus on the shapes \( \gamma, \tau, \) and \( \gamma'^T \) and is the reason why the \( -\gamma, \gamma \) operation appears in our results.

### 8.3 Intersection Term Analysis

In this section, we implement our strategy for analyzing intersection terms. For simplicity, we only give rough definitions and proof sketches here. For a more rigorous treatment, see Appendix B.

We begin by defining intersection patterns which describe how the ribbons \( R_1, R_2, \) and \( R_3 \) intersect.

**Definition 8.8** (Rough Definition of Intersection Patterns). Given \( \tau \in \mathcal{M}', \gamma \in \Gamma_s, \Upsilon \cup \{Id_U\}, \gamma' \in \Gamma_s, \Upsilon \cup \{Id_V\} \), and ribbons \( R_1, R_2, \) and \( R_3 \) of shapes \( \gamma, \tau, \) and \( \gamma'^T \) which are composable but not properly composable, we define the intersection pattern \( P \) induced by \( R_1, R_2, \) and \( R_3 \) and the resulting shape \( \tau_p \) as follows:
1. We take \( V(P) = V(\gamma \circ \tau \circ \gamma' T) \).
2. We take \( E(P) \) to be the set of edges \((u, v)\) such that \( u, v \) are distinct vertices in \( V(\sigma \circ \tau \circ \sigma' T) \) but \( u \) and \( v \) correspond to the same vertex in \( R_1 \circ R_2 \circ R_3 \).
3. We define \( \tau_P \) to be the shape of the ribbon \( R = R_1 \circ R_2 \circ R_3 \).

**Definition 8.9.** Given \( \tau \in \mathcal{M}' \), \( \gamma \in \Gamma_{U_t} \cup \{Id_{U_t}\} \), and \( \gamma' \in \Gamma_{V_t} \cup \{Id_{V_t}\} \), we define \( \mathcal{P}_{\gamma, \tau, \gamma'^T} \) to be the set of all possible intersection patterns \( P \) which can be induced by ribbons \( R_1, R_2, R_3 \) of shapes \( \gamma, \tau, \) and \( \gamma' T \).

**Remark 8.10.** Note that if \( \gamma = Id_{U_t} \) and \( \gamma' = Id_{V_t} \) then \( \mathcal{P}_{\gamma, \tau, \gamma'^T} = \emptyset \) as every intersection pattern must have an unexpected intersection so either \( \gamma \) or \( \gamma' \) must be non-trivial.

It would be nice if the intersection pattern \( P \) together with the ribbon \( R \) allowed us to recover the original ribbons \( R_1, R_2, \) and \( R_3 \). Unfortunately, it is possible for different triples of ribbons to result in the same intersection pattern \( P \) and ribbon \( R \). That said, the number of such triples cannot be too large, and this is sufficient for our purposes.

**Definition 8.11.** Given an intersection pattern \( P \in \mathcal{P}_{\gamma, \tau, \gamma'^T} \), let \( R \) be a ribbon of shape \( \tau_P \). We define \( N(P) \) to be the number of different triples of ribbons \( R_1, R_2, R_3 \) such that \( R_1 \circ R_2 \circ R_3 = R \) and \( R_1, R_2, R_3 \) induce the intersection pattern \( P \).

**Lemma 8.12.** For all intersection patterns \( P \in \mathcal{P}_{\gamma, \tau, \gamma'^T} \), \( N(P) \leq |V(\tau_P)| |V(\gamma) \cup U_t| + |V(\gamma') \cup U_t| \)

**Proof sketch.** This can be proved by making the following observations:

1. \( A_{R_1} = A_R \) and \( B_{R_3} = B_R \).
2. All of the remaining vertices in \( V(R_1) \) and \( V(R_3) \) must be equal to some vertex in \( V(R) \).
3. Once \( R_1 \) and \( R_3 \) are determined, there is at most one ribbon \( R_2 \) such that \( R_1 \circ R_2 \circ R_3 \) is composable, \( R = R_1 \circ R_2 \circ R_3 \), and \( R_1, R_2, R_3 \) induce the intersection pattern \( P \).

With these definitions, we can now analyze the intersection terms.

**Definition 8.13.** Given a left shape \( \sigma \), define \( e_\sigma \) to be the vector which has a 1 in coordinate \( \sigma \) and has a 0 in all other coordinates.

**Lemma 8.14.** For all \( \tau \in \mathcal{M}' \), \( \sigma \in \mathcal{L}_{U_t} \), and \( \sigma' \in \mathcal{L}_{V_t} \),

\[
M_{\tau}^{\text{act}}(e_\sigma e_{\sigma'}^T) - M_{\tau}^{\text{orth}}(e_\sigma e_{\sigma'}^T) = \sum_{\sigma_2 \in \mathcal{L}, \gamma : \sigma_2 \circ \gamma = \sigma} \frac{1}{|\text{Aut}(U_\gamma)|} \sum_{P \in \mathcal{P}_{\gamma, \tau, \gamma'^T}} N(P) M_{\tau_P}^{\text{orth}}(e_{\sigma_2} e_{\sigma'}^T) + \sum_{\sigma_2' \in \mathcal{L}, \gamma' : \sigma_2' \circ \gamma' = \sigma'} \frac{1}{|\text{Aut}(U_{\gamma'})|} \sum_{P \in \mathcal{P}_{\gamma, \tau, \gamma'^T}} N(P) M_{\tau_P}^{\text{orth}}(e_{\sigma_2'} e_{\sigma'}^T) + \sum_{\sigma_2 \in \mathcal{L}, \gamma : \sigma_2 \circ \gamma = \sigma} \sum_{\sigma_2' \in \mathcal{L}, \gamma' : \sigma_2' \circ \gamma' = \sigma'} \frac{1}{|\text{Aut}(U_\gamma)| \cdot |\text{Aut}(U_{\gamma'})|} \sum_{P \in \mathcal{P}_{\gamma, \tau, \gamma'^T}} N(P) M_{\tau_P}^{\text{orth}}(e_{\sigma_2} e_{\sigma'}^T)
\]
Proof sketch. This lemma follows from the following bijection. Consider the third term

\[
\sum_{\sigma_2 \in \mathcal{L}, \gamma \in \Gamma : \sigma_2 \circ \gamma = \sigma} \frac{1}{|\text{Aut}(U_{\gamma})| \cdot |\text{Aut}(U_{\gamma'})|} \sum_{p \in P_{\gamma, \tau, \gamma'T}} N(P) M_{t_p}^{orth}(e_{\sigma_2} e_{\gamma}^T)
\]

On one side, we have the following data:

1. Ribbons \( R_1, R_2, \) and \( R_3 \) of shapes \( \gamma, \tau, \gamma'T \) such that \( R_1, R_2, R_3 \) are composable but \( R_1 \) and \( R_2 \circ R_3 \) are not properly composable (i.e. \( R_1 \) has an unexpected intersection with \( R_2 \) and/or \( R_3 \)) and \( R_1 \circ R_2 \) and \( R_3 \) are not properly composable (i.e. \( R_3 \) has an unexpected intersection with \( R_1 \) and/or \( R_2 \)).
2. An ordering \( O_{A'} \) on the leftmost minimum vertex separator \( A' \) of \( A_{R_1} \) and \( V_{s} \cup B_{R_1} \) (recall that \( V_{s} \) is the set of vertices which appear more than once in \( V(R_1 \circ R_2 \circ R_3) \)).
3. An ordering \( O_{B'} \) on the rightmost minimum vertex separator \( B' \) of \( V_{s} \cup A_{R_3} \) and \( B_{R_3} \).

On the other side, we have the following data

1. An intersection pattern \( P \in P_{\gamma, \tau, \gamma'T} \) where \( \gamma \) and \( \gamma'T \) are non-trivial.
2. Ribbons \( R'_1, R'_2, R'_3 \) of shapes \( \sigma_2, \tau_p, \sigma'_2 \) which are properly composable
3. A number in \([N(P)]\) describing which possible triple of ribbons resulted in the intersection pattern \( P \) and the ribbon \( R'_2 \).

To see this bijection, note that given the data on the first side, we can recover the ribbons \( R'_1, R'_2, \) and \( R'_3 \) as follows:

1. We decompose \( R_1 \) as \( R_1 = R'_1 \circ R_4 \) where \( B_{R'_1} = A_{R_4} = A' \) with the ordering \( O_{A'} \).
2. We decompose \( R_3 \) as \( R_3 = R_5 \circ R'_3 \) where \( B_{R_5} = A_{R'_3} = B' \) with the ordering \( O_{B'} \).
3. We take \( R'_2 = R_4 \circ R_2 \circ R_5 \).

The intersection pattern \( P \) and the number in \([N(P)]\) can be obtained from \( R_1, R_2, \) and \( R_3 \).

Conversely, with the data on the other side, we can recover the data on the first side as follows:

1. \( R'_2 \) gives an ordering \( O_{A'} \) for \( A' = A_{R'_3} \) and an ordering \( O_{B'} \) for \( B' = B_{R'_2} \).
2. The ribbon \( R'_2, \) intersection pattern \( P, \) and number in \([N(P)]\) allow us to recover \( R_4, R_2, \) and \( R_5 \).
3. We take \( R_1 = R'_1 \circ R_4 \) and \( R_3 = R_5 \circ R'_3 \).

Thus, both sides have the same coefficient for each ribbon.

The analysis for the the first term is the same except that when \( \gamma' \) is trivial, we always take \( \gamma' = Id_{V_s} \). Thus, we always have that \( B' = B_{R'_2} = B_{R_5} \) (with the same ordering) and \( R'_3 = R_3 = Id_{B'} \). Because of this, there is no need to specify \( R_3, R'_3, R_5, \) or an ordering on \( B' \).

Similarly, the analysis for the the second term is the same except that when \( \gamma \) is trivial, we always take \( \gamma = Id_{U_s} \). Thus, we always have that \( A' = A_{R'_2} = A_{R_2} \) (with the same ordering) and \( R'_1 = R_1 = Id_{A'} \). Because of this, there is no need to specify \( R_1, R'_1, R_4, \) or an ordering on \( A' \).

Applying Lemma 8.14 for all \( \sigma \) and \( \sigma' \) simultaneously, we obtain the following corollary.
**Definition 8.15.** For all \( U, V \in \mathcal{I}_{mid} \), given a \( \gamma \in \Gamma_{U} \) and a vector \( v \) indexed by left shapes \( \sigma \in \mathcal{L}_{U} \), define \( v^{-\gamma} \) to be the vector indexed by left shapes \( \sigma_{2} \in \Gamma_{U} \) such that \( v^{-\gamma}(\sigma_{2}) = v(\sigma_{2} \circ \gamma) \) if \( \sigma_{2} \circ \gamma \in \mathcal{L}_{U} \) and \( v^{-\gamma}(\sigma_{2}) = 0 \) otherwise.

**Proposition 8.16.** For all composable \( \gamma_{2}, \gamma_{1} \in \Gamma \) and all vectors \( v \) indexed by left shapes in \( \mathcal{L}_{V_{11}} \), \( (v^{-\gamma_{1}})^{-\gamma_{2}} = v^{-\gamma_{2}\circ\gamma_{1}} \).

**Corollary 8.17.** For all \( \tau \in \mathcal{M}' \), for all left \( \tau \)-vectors \( v \) and all right \( \tau \)-vectors \( w \),

\[
M_{\tau}^{orth}(vw^{T}) = M_{\tau}^{fact}(vw^{T}) - \sum_{\gamma \in \Gamma_{U}, \gamma' \in \Gamma_{V}} \frac{1}{|Aut(U_{\gamma})|} \sum_{P \in \mathcal{P}_{\gamma, \gamma' \gamma}, \gamma' \gamma = \gamma} N(P)M_{\tau P}^{orth}(v^{-\gamma}w^{T})
- \sum_{\gamma' \in \Gamma_{U}, \gamma \in \Gamma_{V}} \frac{1}{|Aut(U_{\gamma'})|} \sum_{P \in \mathcal{P}_{\gamma' \gamma, \gamma', \gamma}} N(P)M_{\tau P}^{orth}(v(w^{-\gamma'})^{T})
- \sum_{\gamma \in \Gamma_{U}, \gamma' \in \Gamma_{V}, \gamma' \gamma = \gamma} \frac{1}{|Aut(U_{\gamma})| \cdot |Aut(U_{\gamma'})|} \sum_{P \in \mathcal{P}_{\gamma, \gamma' \gamma}, \gamma' \gamma = \gamma} N(P)M_{\tau P}^{orth}(v^{-\gamma}w^{-\gamma'})^{T})
\]

Applying Corollary 8.17 iteratively, we obtain the following theorem:

**Definition 8.18.** Given \( \gamma, \gamma' \in \Gamma \cup \{Id_{U} : U \in \mathcal{I}_{mid} \} \) and \( j > 0 \), let \( \Gamma_{\gamma, \gamma', j} \) be the set of all \( \gamma_{1}, \gamma'_{1}, \ldots, \gamma_{j}, \gamma'_{j} \in \Gamma \cup \{Id_{U} : U \in \mathcal{I}_{mid} \} \) such that:

1. \( \gamma_{j}, \ldots, \gamma_{1} \) are composable and \( \gamma_{j} \circ \ldots \circ \gamma_{1} = \gamma \)
2. \( \gamma'_{j}, \ldots, \gamma'_{1} \) are composable and \( \gamma'_{j} \circ \ldots \circ \gamma'_{1} = \gamma' \)
3. For all \( i \in [1, j] \), \( \gamma_{i} \) or \( \gamma'_{i} \) is non-trivial (i.e. \( \gamma_{i} \neq Id_{U_{\gamma_{i}}} \) or \( \gamma'_{i} \neq Id_{U_{\gamma'_{i}}} \)).

**Remark 8.19.** Note that if \( \gamma = Id_{U} \) and \( \gamma' = Id_{V} \) then for all \( j > 0 \), \( \Gamma_{\gamma, \gamma', j} = \emptyset \).

**Theorem 8.20.** For all \( \tau \in \mathcal{M}' \), left \( \tau \)-vectors \( v \), and right \( \tau \)-vectors \( w \),

\[
M_{\tau}^{orth}(vw^{T}) = M_{\tau}^{fact}(vw^{T}) + \sum_{\gamma \in \Gamma_{U}, \gamma' \in \Gamma_{V}, \gamma \circ \gamma' \text{ is non-trivial}} (-1)^{j} \sum_{\gamma_{1}, \gamma'_{1}, \ldots, \gamma_{j}, \gamma'_{j} \in \Gamma_{\gamma, \gamma', j}, \gamma_{j} \circ \ldots \circ \gamma_{1} = \gamma} \prod_{\gamma_{i} \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma_{i}})|} \prod_{\gamma'_{i} \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma'_{i}})|} \sum_{P_{1}, \ldots, P_{j} \in \mathcal{P}_{\gamma, \gamma' \gamma}, \gamma \circ \gamma' = \gamma} N(P_{1}) \cdots N(P_{j}) M_{\tau P_{1} \cdots P_{j}}^{fact}(v^{-\gamma}(w^{-\gamma'})^{T})
\]

where we take \( \tau_{P_{0}} = \tau \).

### 8.4 Bounding the difference between \( M^{fact} \) and \( M^{orth} \)

In this subsection, we bound the difference between \( M_{\tau}^{fact}(H_{\gamma}) \) and \( M_{\tau}^{orth}(H_{\gamma}) \). We recall the following conditions on \( B(\gamma), N(\gamma), \) and \( c(\gamma) \):
1. For all $\tau \in \mathcal{M}'$, $\gamma \in \Gamma_{s,U_\tau} \cup \{Id_{U_\tau}\}$, and $\gamma' \in \Gamma_{s,V_\tau} \cup \{Id_{V_\tau}\}$,

$$
\sum_{j>0} \sum_{\gamma_1, \gamma'_1, \ldots, \gamma_j, \gamma'_j \in \Gamma_{\gamma,\gamma'}} \left( \prod_{i: \gamma_i \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma_i})|} \right) \left( \prod_{i: \gamma'_i \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma'_i})|} \right)
$$

$$
\sum_{P_1, \ldots, P_j \in P_{\gamma_1, \ldots, \gamma_j}} \left( \prod_{i=1}^{j} N(P_i) \right) \leq \frac{N(\gamma)N(\gamma')}{(|Aut(U_{\gamma})|)^{1/2} \gamma \text{ is non-trivial} (|Aut(U_{\gamma'})|)^{1/2} \gamma' \text{ is non-trivial}}
$$

2. For all $\tau \in \mathcal{M}'$, $\gamma \in \Gamma_{s,U_\tau}$, and $\gamma' \in \Gamma_{s,V_\tau}$, for all $P \in P_{\gamma,\gamma'}$, $B_{norm}(\tau P) \leq B(\gamma)B(\gamma')B_{norm}(\tau)$

3. $\forall V \in \mathcal{I}_{mid}, \sum_{\gamma \in \Gamma_{s,V}} \frac{1}{|Aut(U_{\gamma})|} \leq \epsilon' \leq \frac{1}{20}$

With these conditions, we can now bound the difference between $M_{\tau}^{fact}$ and $M_{\tau}^{orth}$.

**Lemma 8.21.** If the norm bounds and the conditions on $B(\gamma), N(\gamma)$, and $c(\gamma)$ hold then for all $\tau \in \mathcal{M}'$, left $\tau$-vectors $v$, and right $\tau$-vectors $w$,

$$
\left( M_{\tau}^{fact}(vw^T) + M_{\tau}^{fact}(wv^T) \right) - \left( M_{\tau}^{orth}(vw^T) + M_{\tau}^{orth}(wv^T) \right) \leq
$$

$$
\epsilon' B_{norm}(\tau) M_{Id_{U_\tau}}^{fact}(vw^T) + 2 \sum_{\gamma \in \Gamma_{s,U_\tau}} \frac{B(\gamma)^2 N(\gamma)^2 B_{norm}(\tau)c(\gamma)}{|Aut(U_{\gamma})|} M_{Id_{U_\tau}}^{fact}(v^-\gamma(w^-\gamma)^T) +
$$

$$
\epsilon' B_{norm}(\tau) M_{Id_{V_\tau}}^{fact}(wv^T) + 2 \sum_{\gamma' \in \Gamma_{s,V_\tau}} \frac{B(\gamma')^2 N(\gamma')^2 B_{norm}(\tau)c(\gamma')}{|Aut(U_{\gamma'})|} M_{Id_{V_\tau}}^{fact}(w^-\gamma'(w^-\gamma')^T)
$$

**Proof.** By Theorem 8.20, taking $\tau_{P_0} = \tau$,

$$
M_{\tau}^{orth}(vw^T) = M_{\tau}^{fact}(vw^T) +
$$

$$
\sum_{\gamma \in \Gamma_{s,U_\tau} \cup \{Id_{U_\tau}\}, \gamma' \in \Gamma_{s,V_\tau} \cup \{Id_{V_\tau}\}, j>0} \sum_{\gamma_1, \gamma'_1, \ldots, \gamma_j, \gamma'_j \in \Gamma_{\gamma,\gamma'}} \left( \prod_{i: \gamma_i \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma_i})|} \right) \left( \prod_{i: \gamma'_i \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma'_i})|} \right)
$$

$$
\sum_{P_1, \ldots, P_j \in P_{\gamma_1, \ldots, \gamma_j}} \left( \prod_{i=1}^{j} N(P_i) \right) M_{\tau P_i}^{fact}(v^-\gamma(w^-\gamma)^T)
$$

Taking the transpose of this equation gives

$$
M_{\tau}^{orth}(wv^T) = M_{\tau}^{fact}(wv^T) +
$$

$$
\sum_{\gamma \in \Gamma_{s,U_\tau} \cup \{Id_{U_\tau}\}, \gamma' \in \Gamma_{s,V_\tau} \cup \{Id_{V_\tau}\}, j>0} \sum_{\gamma_1, \gamma'_1, \ldots, \gamma_j, \gamma'_j \in \Gamma_{\gamma,\gamma'}} \left( \prod_{i: \gamma_i \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma_i})|} \right) \left( \prod_{i: \gamma'_i \text{ is non-trivial}} \frac{1}{|Aut(U_{\gamma'_i})|} \right)
$$

$$
\sum_{P_1, \ldots, P_j \in P_{\gamma_1, \ldots, \gamma_j}} \left( \prod_{i=1}^{j} N(P_i) \right) M_{\tau P_i}^{fact}(w^-\gamma'(v^-\gamma)^T)
$$

69
Now observe that by Lemma 8.5, if the norm bounds hold,

\[ \pm \left( M_{\text{fact}}^f (v^{-\gamma} (w^{-\gamma'} T) + M_{\text{fact}}^f (w^{-\gamma'} (v^{-\gamma})^T) \right) = \]

\[ \pm M_{\text{fact}}^f \left( \sqrt{\frac{N(\gamma) B(\gamma) c(\gamma)}{N(\gamma') B(\gamma') c(\gamma')}} \right) ^{-\gamma} \left( \sqrt{\frac{N(\gamma) B(\gamma) c(\gamma)}{N(\gamma) B(\gamma) c(\gamma')}} \right) ^{-\gamma'} \]

\[ M_{\text{fact}}^f \left( \frac{N(\gamma) B(\gamma) c(\gamma)}{N(\gamma') B(\gamma') c(\gamma')} \right) \leq \]

\[ B_{\text{norm}}(\tau_p) \left( \frac{N(\gamma) B(\gamma) c(\gamma)}{N(\gamma') B(\gamma') c(\gamma')} \right) M_{\text{fact}}^f \left( v^{-\gamma} (v^{-\gamma})^T \right) + \frac{N(\gamma') B(\gamma') c(\gamma')}{N(\gamma) B(\gamma) c(\gamma')} M_{\text{fact}}^f \left( w^{-\gamma'} (w^{-\gamma'})^T \right) \]

Combining these equations,

\[ \left( M_{\text{fact}}^f (vw^T) + M_{\text{fact}}^f (wv^T) \right) - \left( M_{\text{fact}}^f (vw^T) + M_{\text{fact}}^f (wv^T) \right) \leq \]

\[ \sum_{\gamma \in \Gamma_{\gamma,\gamma'}} \sum_{\gamma \in \Gamma_{\gamma,\gamma'}} \sum_{\prod \gamma_j \text{ is non-trivial}} \prod \frac{1}{|\text{Aut}(U_{\gamma_j})|} \prod \frac{1}{|\text{Aut}(U_{\gamma_j})|} \]

\[ \sum_{P_1, \ldots, P_l \in P_{\gamma,\gamma'}} \left( \prod_{i=1}^l N(P_i) \right) \]

\[ \left( \frac{N(\gamma) B(\gamma) c(\gamma)}{N(\gamma') B(\gamma') c(\gamma')} \right) M_{\text{fact}}^f \left( v^{-\gamma} (v^{-\gamma})^T \right) + \frac{N(\gamma') B(\gamma') c(\gamma')}{N(\gamma) B(\gamma) c(\gamma')} M_{\text{fact}}^f \left( w^{-\gamma'} (w^{-\gamma'})^T \right) \]

From the conditions on \( B(\gamma) \) and \( N(\gamma) \),

1. \( B_{\text{norm}}(\tau_p) \leq B(\gamma) B(\gamma') B_{\text{norm}}(\tau) \)
2. 

\[ \sum_{j > 0} \sum_{\gamma_1 \gamma'_1 \ldots \gamma_l \in \Gamma_{\gamma,\gamma'}} \left( \prod_{i=1}^l \frac{1}{|\text{Aut}(U_{\gamma_j})|} \right) \left( \prod_{i=1}^l \frac{1}{|\text{Aut}(U_{\gamma_j})|} \right) \]

\[ \sum_{P_1, \ldots, P_l \in P_{\gamma,\gamma'}} \left( \prod_{i=1}^l N(P_i) \right) \leq \frac{N(\gamma) N(\gamma')}{(|\text{Aut}(U_{\gamma})|)^l \gamma \text{ is non-trivial} (|\text{Aut}(U_{\gamma'})|)^l \gamma' \text{ is non-trivial}} \]

Putting these together,

\[ \left( M_{\text{fact}}^f (vw^T) + M_{\text{fact}}^f (wv^T) \right) - \left( M_{\text{fact}}^f (vw^T) + M_{\text{fact}}^f (wv^T) \right) \leq \]

\[ \sum_{\gamma \in \Gamma_{\gamma,\gamma'}} \sum_{\gamma \in \Gamma_{\gamma,\gamma'}} \sum_{\prod \gamma_j \text{ is non-trivial}} \prod \frac{B(\gamma)^2 N(\gamma)^2 B_{\text{norm}}(\tau) c(\gamma)}{|\text{Aut}(U_{\gamma})|^{l \gamma \text{ is non-trivial}} (|\text{Aut}(U_{\gamma'})|)^{l \gamma' \text{ is non-trivial}} c(\gamma')} \]

\[ M_{\text{fact}}^f \left( v^{-\gamma} (v^{-\gamma})^T \right) + \]

\[ \sum_{\gamma \in \Gamma_{\gamma,\gamma'}} \sum_{\gamma \in \Gamma_{\gamma,\gamma'}} \sum_{\prod \gamma_j \text{ is non-trivial}} \prod \frac{B(\gamma')^2 N(\gamma')^2 B_{\text{norm}}(\tau) c(\gamma')}{(|\text{Aut}(U_{\gamma'})|)^{l \gamma' \text{ is non-trivial}} c(\gamma')} \]

\[ M_{\text{fact}}^f \left( w^{-\gamma'} (w^{-\gamma'})^T \right) \]
Now observe that
\[
\sum_{\gamma \in \Gamma_{\tau,U}, \gamma' \in \Gamma_{\tau,U}, \gamma \neq \gamma': \gamma \text{ or } \gamma' \text{ is non-trivial}} B(\gamma)'N(\gamma)'B_{\text{norm}}(\tau)c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (v^{-\gamma}(v^{-\gamma})^T) \leq
\left( \sum_{\gamma \in \Gamma_{\tau,U}} \frac{1}{|\text{Aut}(U_\gamma)|} \right) B_{\text{norm}}(\tau) M_{id_{U,\gamma}}^{\text{fact}} (v^T) + \sum_{\gamma' \in \Gamma_{\tau,U}} \left( \sum_{\gamma \in \Gamma_{\tau,U}, \gamma \neq \gamma': \gamma \text{ or } \gamma' \text{ is non-trivial}} \right) \frac{1}{|\text{Aut}(U_\gamma)|} B(\gamma)'N(\gamma)'B_{\text{norm}}(\tau)c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (v^{-\gamma}(v^{-\gamma})^T) \leq
\varepsilon' B_{\text{norm}}(\tau) M_{id_{V,\gamma}}^{\text{fact}} (w^T) + 2 \sum_{\gamma' \in \Gamma_{\tau,U}} B(\gamma)'N(\gamma)'B_{\text{norm}}(\tau)c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (w^{-\gamma}(w^{-\gamma})^T)
\]
Following similar logic,
\[
\varepsilon' B_{\text{norm}}(\tau) M_{id_{V,\gamma}}^{\text{fact}} (w^T) + 2 \sum_{\gamma' \in \Gamma_{\tau,U}} B(\gamma)'N(\gamma)'B_{\text{norm}}(\tau)c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (w^{-\gamma}(w^{-\gamma})^T)
\]
Putting everything together,
\[
\left( M_{\tau}^{\text{fact}} (vw^T) + M_{\tau}^{\text{fact}} (vw^T) \right) - \left( M_{\tau}^{\text{orth}} (vw^T) + M_{\tau}^{\text{orth}} (vw^T) \right) \leq
\varepsilon' B_{\text{norm}}(\tau) M_{id_{U,\gamma}}^{\text{fact}} (v^T) + 2 \sum_{\gamma \in \Gamma_{\tau,U}} B(\gamma)'N(\gamma)'B_{\text{norm}}(\tau)c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (v^{-\gamma}(v^{-\gamma})^T) + \varepsilon' B_{\text{norm}}(\tau) M_{id_{V,\gamma}}^{\text{fact}} (w^T) + 2 \sum_{\gamma' \in \Gamma_{\tau,U}} B(\gamma)'N(\gamma)'B_{\text{norm}}(\tau)c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (w^{-\gamma}(w^{-\gamma})^T)
\]
as needed.

Using Lemma 8.21 we have the following corollaries:

**Corollary 8.22.** For all \( U \in \mathcal{I}_{mid} \), if the norm bounds and the conditions on \( B(\gamma) \), \( N(\gamma) \), and \( c(\gamma) \) hold and \( H_{id_U} \geq 0 \) then
\[
M_{id_U}^{\text{fact}} (H_{id_U}) - M_{id_{U}}^{\text{orth}} (H_{id_U}) \leq \varepsilon' M_{id_U}^{\text{fact}} (H_{id_U}) + 2 \sum_{\gamma \in \Gamma_{\tau,U}} B(\gamma)'N(\gamma)'c(\gamma)' M_{id_{U,\gamma}}^{\text{fact}} (H_{id_U})
\]

**Corollary 8.23.** For all \( U \in \mathcal{I}_{mid} \) and all \( \tau \in \mathcal{M}_U \), if the norm bounds and the conditions on \( B(\gamma) \), \( N(\gamma) \), and \( c(\gamma) \) hold and
\[
\left[ \frac{1}{|\text{Aut}(U)| c(\tau)} H_{id_U} \right]^T H_{id_U}^{\tau} B_{\text{norm}}(\tau) H_{id_U}^{\tau} \frac{1}{|\text{Aut}(U)| c(\tau)} H_{id_U} \leq 0
\]
71
then
\[
\left( M^\text{fact}_\tau(H_\tau) + M^\text{fact}_\tau(H_\tau^T) \right) - \left( M^\text{orth}_\tau(H_\tau) + M^\text{orth}_\tau(H_\tau^T) \right) \leq 2\epsilon \frac{1}{|\text{Aut}(U)|c(\tau)} M^\text{fact}_{\text{Id}_U}(H_{\text{Id}_U}) + 4 \sum_{\gamma \in \Gamma_\tau} \frac{B(\gamma)^2 N(\gamma)^2 c(\gamma)}{|\text{Aut}(U_\gamma)| \cdot |\text{Aut}(U)|c(\tau)} M^\text{fact}_{\text{Id}_{U_\gamma}}(H_{\text{Id}_{U_\gamma}}^{-\gamma,\gamma})
\]

8.5 Proof of the Main Theorem

We now prove the following theorem which is a slight modification of Theorem 8.1 and which implies Theorem 8.1.

**Theorem 8.24.** For all \( \epsilon > 0 \) and all \( \epsilon' \in (0, \frac{1}{20}] \), for any moment matrix

\[
\Lambda = \sum_{U \in \mathcal{I}_{\text{mid}}} M^\text{orth}_{\text{Id}_U}(H_{\text{Id}_U}) + \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_U} M^\text{orth}_\tau(H_\tau),
\]

if we have that for all \( \alpha \in \mathcal{M}', ||M_\alpha|| \leq B_{\text{norm}}(\alpha) \) and \( B(\gamma), N(\gamma), \) and \( c(\alpha) \) are functions such that

1. For all \( \tau \in \mathcal{M}', \gamma \in \Gamma_{s,U}, \gamma' \in \Gamma_{s,V}, \) and all intersection patterns \( P \in \mathcal{P}_{\gamma,\tau,\gamma'}, \)

\[
B_{\text{norm}}(\tau_P) \leq B(\gamma)B(\gamma')B_{\text{norm}}(\tau)
\]

2. For all composable \( \gamma_1, \gamma_2, B(\gamma_1)B(\gamma_2) = B(\gamma_1 \circ \gamma_2). \)

3. \( \forall U \in \mathcal{I}_{\text{mid}}, \sum_{\gamma \in \Gamma_{s,U}} \frac{1}{|\text{Aut}(U)|c(\gamma)} < \epsilon' \)

4. \( \forall V \in \mathcal{I}_{\text{mid}}, \sum_{\gamma \in \Gamma_{s,V}} \frac{1}{|\text{Aut}(U_\gamma)|c(\gamma)} < \epsilon' \)

5. \( \forall U \in \mathcal{I}_{\text{mid}}, \sum_{\tau \in \mathcal{M}_U} \frac{1}{|\text{Aut}(U)|c(\tau)} < \epsilon' \)

6. For all \( \tau \in \mathcal{M}', \gamma \in \Gamma_{s,U}, \gamma' \in \Gamma_{s,V}, \) and all \( \gamma' \in \Gamma_{s,V} \cup \{\text{Id}_V\}, \)

\[
\sum_{j \geq 0} \sum_{\gamma_1, \gamma'_{i_1}, \ldots, \gamma'_{i_j} \in \Gamma_{s,V}} \prod_{i=1}^{j} \frac{|\text{Aut}(U_{\gamma_i})|}{|\text{Aut}(U)|c(i)} \prod_{i=2}^{j} \frac{1}{|\text{Aut}(U_{\gamma'_i})|c(i')} = \sum_{\gamma \in \Gamma_{s,c}} \frac{N(\gamma)N(\gamma')}{(|\text{Aut}(U_{\gamma})|)^{1/2} \times \text{non-trivial} (|\text{Aut}(U_{\gamma'})|)^{1/2} \times \text{non-trivial}}
\]

and we have SOS-symmetric coefficient matrices \( \{H'_\gamma : \gamma \in \Gamma\} \) such that the following conditions hold:

1. For all \( U \in \mathcal{I}_{\text{mid}}, H_{\text{Id}_U} \succeq 0 \)

2. For all \( U \in \mathcal{I}_{\text{mid}} \) and \( \tau \in \mathcal{M}_U, \)

\[
\left[ \frac{1}{|\text{Aut}(U)|c(\tau)} H_{\text{Id}_U} \quad \frac{B_{\text{norm}}(\tau)H_{\tau}}{1} \right] \preceq 0
\]

3. For all \( U, V \in \mathcal{I}_{\text{mid}} \) where \( \omega(U) > \omega(V) \) and all \( \gamma \in \Gamma_{U,V}, \)

\[
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{\text{Id}_V}^{-\gamma,\gamma} \preceq H'_\gamma
\]
then

\[ \Lambda \geq \frac{1}{2} \left( \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) \right) - 3 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{d_{I_{d_{U}}}(H_{\gamma}, H_{I_{d_{U}}})}{|\text{Aut}(U)|c(\gamma)} \right) \text{Id}_{\text{sym}} \]

If it is also true that

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) \geq 6 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{d_{I_{d_{U}}}(H_{\gamma}, H_{I_{d_{U}}})}{|\text{Aut}(U)|c(\gamma)} \right) \text{Id}_{\text{sym}} \]

then \( \Lambda \geq 0 \).

\[ \text{Proof.} \] We make the following observations:

1. By Theorem 8.2,

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) + \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_{U}} M_{\tau}(H_{\tau}) \leq (1 - 2\varepsilon') \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) \]

2. By Corollary 8.22,

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} \left( M_{I_{d_{U}}}(H_{I_{d_{U}}}) - M_{I_{d_{U}}}(H_{I_{d_{U}}}) \right) \leq \varepsilon' \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) + 2 \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{M_{I_{d_{U}}}(H_{\gamma})}{c(\gamma)|\text{Aut}(U_{\gamma})|} \]

3. By Corollary 8.23,

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_{U}} \left( M_{\tau}(H_{\tau}) - M_{\tau}(H_{\tau}) \right) \leq \]

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\tau \in \mathcal{M}_{U}} \left( \frac{2\varepsilon'}{|\text{Aut}(U)|c(\tau)} M_{I_{d_{U}}}(H_{I_{d_{U}}}) + 4 \sum_{\gamma \in \Gamma_{U, s}} \frac{B(\gamma)^2 N(\gamma)^2 c(\gamma)}{|\text{Aut}(U_{\gamma})| \cdot |\text{Aut}(U)|c(\tau)} M_{I_{d_{U}}}(H_{\gamma}) \right) \]

\[ 2\varepsilon'^2 \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) + 4\varepsilon' \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{M_{I_{d_{U}}}(H_{\gamma})}{c(\gamma)|\text{Aut}(U_{\gamma})|} \]

4.

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{M_{I_{d_{U}}}(H_{\gamma})}{c(\gamma)|\text{Aut}(U_{\gamma})|} = \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{M_{I_{d_{U}}}(H_{I_{d_{U}}}) + \left( M_{I_{d_{U}}}(H_{\gamma}) - M_{I_{d_{U}}}(H_{I_{d_{U}}}) \right)}{c(\gamma)|\text{Aut}(U_{\gamma})|} \leq \]

\[ \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{M_{I_{d_{U}}}(H_{I_{d_{U}}})}{c(\gamma)|\text{Aut}(U_{\gamma})|} + \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{d_{I_{d_{U}}}(H_{\gamma}, H_{I_{d_{U}}})}{|\text{Aut}(U_{\gamma})|c(\gamma)} \right) \text{Id}_{\text{sym}} \leq \]

\[ \varepsilon' \sum_{U \in \mathcal{I}_{\text{mid}}} M_{I_{d_{U}}}(H_{I_{d_{U}}}) + \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U, s}} \frac{d_{I_{d_{U}}}(H_{\gamma}, H_{I_{d_{U}}})}{|\text{Aut}(U_{\gamma})|c(\gamma)} \right) \text{Id}_{\text{sym}} \]
Putting everything together,

\[
\Lambda = \sum_{U \in \mathcal{I}_{mid}} M_{\text{orth}}(H_{I(U)}) + \sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in \mathcal{M}_{U}} M_{\tau}^{\text{orth}}(H_{\tau}) = \\
\sum_{U \in \mathcal{I}_{mid}} M_{\text{act}}(H_{I(U)}) + \sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in \mathcal{M}_{U}} M_{\tau}^{\text{act}}(H_{\tau}) + \sum_{U \in \mathcal{I}_{mid}} \left( M_{\text{act}}^{\text{fact}}(H_{I(U)}) - M_{\text{orth}}^{\text{fact}}(H_{I(U)}) \right) + \\
\sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in \mathcal{M}_{U}} \left( M_{\tau}^{\text{act}}(H_{\tau}) - M_{\text{orth}}^{\text{act}}(H_{\tau}) \right) \geq \\
(1 - 3\varepsilon' - 2\varepsilon'^2) \sum_{U \in \mathcal{I}_{mid}} M_{\text{act}}(H_{I(U)}) - (2 + 4\varepsilon') \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U}} \frac{M_{I(U)}^{\text{act}}(H_{\gamma}^{\prime})}{|\text{Aut}(U_{\gamma})|} \geq \\
(1 - 5\varepsilon' - 6\varepsilon'^2) \sum_{U \in \mathcal{I}_{mid}} M_{\text{act}}(H_{I(U)}) - (2 + 4\varepsilon') \left( \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U}} \frac{d_{I(U)}^{\text{act}}(H_{\gamma}^{\prime}, H_{I(U)}^{\prime})}{|\text{Aut}(U_{\gamma})|} \right) \text{Id}_{\text{sym}} \geq \\
\frac{1}{2} \sum_{U \in \mathcal{I}_{mid}} M_{\text{act}}(H_{I(U)}) - 3 \left( \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U}} \frac{d_{I(U)}^{\text{act}}(H_{\gamma}^{\prime}, H_{I(U)}^{\prime})}{|\text{Aut}(U_{\gamma})|} \right) \text{Id}_{\text{sym}}
\]

9 Choosing the functions \(B_{\text{norm}}(\alpha), B(\gamma), N(\gamma), \text{and } c(\alpha)\)

In this subsection, we give functions \(B_{\text{norm}}(\alpha), B(\gamma), N(\gamma), \text{and } c(\alpha)\) which satisfy the conditions needed for our machinery.

9.1 Theorem Statements

Recall the following definitions from Section 7.10.

Definition 9.1. We define \(S_{\alpha}\) to be the leftmost minimum vertex separator of \(\alpha\)

Definition 9.2 (Simplified Isolated Vertices). Under our simplifying assumptions, we define

\[
I_{\alpha} = \{ v \in W_{\alpha} : v \text{ is not incident to any edges in } E(\alpha) \}
\]

Theorem 9.3 (Simplified \(B_{\text{norm}}(\alpha), B(\gamma), N(\gamma), \text{and } c(\alpha)\)). Under our simplifying assumptions, for all \(\varepsilon, \varepsilon' > 0\) and all \(D_{V} \in \mathbb{N}\), if we take

1. \(q = 3 \left[ D_{V} \ln(n) + \frac{\ln(\frac{1}{\varepsilon})}{3} + D_{V} \ln(5) + 3D_{V}^{2} \ln(2) \right] \)
2. \(B_{\text{vertex}} = 6D_{V} \sqrt{2eq} \)
3. \(B_{\text{norm}}(\alpha) = B_{\text{vertex}}^{\text{vertex}} \left| V(\alpha) \backslash U_{\alpha} \right| + \left| V(\alpha) \backslash V_{\alpha} \right| n^{-\frac{\omega(\alpha) + \omega(\alpha') - \omega(S_{\alpha})}{2}} \)
4. \(B(\gamma) = B_{\text{vertex}}^{\text{vertex}} \left| V(\gamma) \backslash U_{\gamma} \right| + \left| V(\gamma) \backslash V_{\gamma} \right| n^{-\frac{\omega(\gamma) - \omega(U_{\gamma})}{2}} \)
5. \(N(\gamma) = (3D_{V})^{2} |V(\gamma) \backslash V_{\gamma}| + |V(\gamma) \backslash U_{\gamma}| \)
6. \(c(\alpha) = \frac{5(3D_{V})^{2}|U_{\alpha} \backslash V_{\alpha}| + |V_{\alpha} \backslash U_{\alpha}| + 2E(\alpha)^{2} |V(\alpha) \backslash (U_{\alpha} \cup V_{\alpha})|}{\varepsilon'} \)
then the following conditions hold:

1. With probability at least \((1 - \epsilon)\), \(\forall \alpha \in \mathcal{M}', ||M_\alpha|| \leq B_{\text{norm}}(\alpha)\)
2. For all \(\tau \in \mathcal{M}', \gamma \in \Gamma_{s,U} \cup \{\text{Id}_{U_i}\}, \gamma' \in \Gamma_{s,V_t} \cup \{\text{Id}_{V_t}\}, \) and intersection patterns \(P \in \mathcal{P}_{\gamma',\tau,\gamma},\)

\[
B_{\text{norm}}(\tau_P) \leq B(\gamma)B(\gamma')B_{\text{norm}}(\tau)
\]

3. For all composable \(\gamma_1, \gamma_2, B(\gamma_1)B(\gamma_2) = B(\gamma_1 \circ \gamma_2)\).
4. \(\forall U \in \mathcal{I}_{\text{mid}}, \sum_{\gamma \in \Gamma_{s,U}} \frac{1}{|\text{Aut}(U)|c(\gamma)} < \epsilon'\)
5. \(\forall V \in \mathcal{I}_{\text{mid}}, \sum_{\gamma \in \Gamma_{s,V}} \frac{1}{|\text{Aut}(U)|c(\gamma)} < \epsilon'\)
6. \(\forall U \in \mathcal{I}_{\text{mid}}, \sum_{\tau \in \mathcal{M}_U} \frac{1}{|\text{Aut}(U)|c(\gamma)} < \epsilon'\)
7. For all \(\tau \in \mathcal{M}', \gamma \in \Gamma_{s,U} \cup \{\text{Id}_{U_i}\}, \) and \(\gamma' \in \Gamma_{s,V_t} \cup \{\text{Id}_{V_t}\},\)

\[
\sum_{j > 0} \gamma_1, \gamma_2, \cdots, \gamma_j, \gamma_j' \in \Gamma_{s,U} \ \text{is non-trivial} \prod_{i; \gamma'_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_\gamma)|} \prod_{i; \gamma'_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma'}_i)|} \sum_{P_i, \cdots, P_t \in \mathcal{P}_{\gamma_1,\gamma_2,\cdots,\gamma_j'}} \left( \prod_{i=1}^j N(P_i) \right) \leq \frac{N(\gamma)N(\gamma')}{(|\text{Aut}(U_\gamma)|)^{1/2} \text{is non-trivial} (|\text{Aut}(U_{\gamma'}_i)|)^{1/2} \text{is non-trivial}}
\]

9.1.1 General functions \(B_{\text{norm}}(\alpha), B(\gamma), N(\gamma), \) and \(c(\alpha)\)

Recall the following definitions from Section 7.10.1.

**Definition 9.4** \((S_{a,\text{min}} \text{ and } S_{a,\text{max}})\). Given a shape \(\alpha \in \mathcal{M}', \) define \(S_{a,\text{min}}\) to be the leftmost minimum vertex separator of \(\alpha\) if all edges with multiplicity at least 2 are deleted and define \(S_{a,\text{max}}\) to be the leftmost minimum vertex separator of \(\alpha\) if all edges with multiplicity at least 2 are present.

**Definition 9.5** (General \(I_\alpha\)). Given a shape \(\alpha, \) define \(I_\alpha\) to be the set of vertices in \(V(\alpha) \backslash (U_\alpha \cup V_\alpha)\) such that all edges incident with that vertex have multiplicity at least 2.

**Definition 9.6** \((B_\Omega)\). We take \(B_\Omega(j)\) to be a non-decreasing function such that for all \(j \in \mathbb{N}, E_\Omega[x^j] \leq B_\Omega(j)^j\)

**Definition 9.7**. For all \(i,\) we define \(h^+_i\) to be the polynomial \(h_i\) where we make all of the coefficients have positive sign.

**Lemma 9.8**. If \(\Omega = N(0, 1)\) then we can take \(B_\Omega(j) = \sqrt{j}\) and we have that

**Theorem 9.9** (General \(B_{\text{norm}}(\alpha), B(\gamma), N(\gamma), \) and \(c(\alpha)\)). For all \(\epsilon, \epsilon' > 0 \) and all \(D_V, D_E \in \mathbb{N}, \) if we take

1. \(q = \lceil 3D_V \ln(n) + \ln(\frac{1}{\epsilon}) + (3D_V)k \ln(D_E + 1) + 3D_V \ln(5) \rceil\)
2. \(B_{\text{vertex}} = 6qD_V\)
3. \(B_{\text{edge}}(\epsilon) = 2h^+_{\text{edge}}(B_\Omega(6D_V D_E)) \max_{j \in [0, 3D_V D_E]} \left\{ \left( h^+_i(B_\Omega(2qj)) \right)^{\frac{q}{\max[1,j]}} \right\} \)
4. \(B_{\text{norm}}(\alpha) = 2eB_{\text{vertex}}^{||V(\alpha)\backslash U_\alpha|| + ||V(\alpha)\backslash V_\alpha||} \left( \prod_{e \in E(\alpha)} B_{\text{edge}}(e) \right) n^{|w(V(\alpha)) + w(U_\alpha) - w(\alpha)} \)

75
5. \( B(\gamma) = B_{\text{vertex}}^{\|V(\gamma)\|U_3 + |V(\gamma)\|V_3} \left( \prod_{e \in E(\gamma)} B_{\text{edge}}(e) \right) n^{\frac{|\text{max}(\gamma)\|U_3|}{2}} \)

6. \( N(\gamma) = (3D_V)^{2|V(\gamma)\|V_3 + |V(\gamma)\|U_3|} \)

7. \( c(\alpha) = \frac{5(3\text{max}D_V)^{U_3V_3} + |V_3| + |V_3| + |E(\alpha)| (2\text{max}D_V)^{|V(\gamma)|} (U_3V_3) }{e} \)

then the following conditions hold:

1. With probability at least \((1 - \varepsilon), \forall \alpha \in \mathcal{M'}, ||M_\alpha|| \leq B_{\text{norm}}(\alpha)\)
2. For all \( \tau \in \mathcal{M'}, \gamma_1 \in \Gamma_{\ast, U_3} \cup \{1 \text{Id}_{U_3}\}, \gamma_2 \in \Gamma_{\ast, V_3} \cup \{1 \text{Id}_{V_3}\}, \) and intersection patterns \( P \in \mathcal{P}_{\gamma_1, \gamma_2} \),
\[
B_{\text{norm}}(\tau_P) \leq B(\gamma) B(\gamma') B_{\text{norm}}(\tau)
\]
3. For all composable \( \gamma_1, \gamma_2, B(\gamma_1) B(\gamma_2) = B(\gamma_1 \circ \gamma_2) \).
4. \( \forall U \in \mathcal{I}_{mid}, \sum_{\gamma \in \Gamma_{U_3}} \frac{1}{|\text{Aut}(U)|} < \varepsilon' \)
5. \( \forall V \in \mathcal{I}_{mid}, \sum_{\gamma \in \Gamma_{V_3}} \frac{1}{|\text{Aut}(V)|} < \varepsilon' \)
6. \( \forall U \in \mathcal{I}_{mid}, \sum_{\tau \in \mathcal{M}_U} \frac{1}{|\text{Aut}(U)|} < \varepsilon' \)
7. For all \( \tau \in \mathcal{M'}, \gamma_1 \in \Gamma_{\ast, U_3} \cup \{1 \text{Id}_{U_3}\}, \) and \( \gamma_2 \in \Gamma_{\ast, V_3} \cup \{1 \text{Id}_{V_3}\}, \)
\[
\sum_{j \geq 0} \sum_{\gamma_1, \gamma_1', \ldots, \gamma_1''} \prod_{i: \gamma_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma_i})|} \prod_{i: \gamma_i' \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma_i'})|} \sum_{P_1, \ldots, P_i \in \mathcal{P}_{\gamma_1, \gamma_1', \ldots, \gamma_1''}} \left( \prod_{i=1}^{j} N(P_i) \right) \leq \frac{N(\gamma) N(\gamma')}{(|\text{Aut}(U_{\gamma})|)^{\gamma_1 \text{ is non-trivial}} (|\text{Aut}(U_{\gamma'})|)^{\gamma_1' \text{ is non-trivial}}} \]

**Remark 9.10.** Recall that if \( \Omega = N(0, 1) \) then we may take \( B_{\Omega}(j) = \sqrt{j} \) and we have that
\[
h_j^+(x) \leq \frac{1}{\sqrt{j!}}(x^2 + j)^{\frac{1}{2}} \leq \left( \frac{e}{j}(x^2 + j) \right)^{\frac{1}{2}}
\]
Thus, when \( \Omega = N(0, 1) \) we can take
\[
B_{\text{edge}}(e) = 2 \left( \frac{e}{l_e}(6D_V D_E + l_e) \right)^{l_e} (e(6D_V D_E q + 1))^{l_e} \leq (400D_V^2 D_E^2 q)^{l_e}
\]

### 9.2 Choosing \( B_{\text{norm}}(\alpha) \)

We need matrix norm bounds which hold for all \( \alpha \in \mathcal{M'} \). For convenience, we recall the definition of \( \mathcal{M'} \) below.

**Definition 9.11 (\( \mathcal{M'} \)).** We define \( \mathcal{M'} \) to be the set of all shapes \( \alpha \) such that

1. \( |V(\alpha)| \leq 3D_V \)
2. \( \forall e \in E(\alpha), l_e \leq D_E \)
3. \( \forall e \in E(\alpha) \) have multiplicity at most \( 3D_V \).
To obtain such norm bounds, we start with the norm bounds in the graph matrix norm bound paper. We then modify these bounds as follows:

1. We make the bounds more compatible with the conditions of our machinery. To do this, we upper bound many of the terms in the norm bound by $B_{\text{vertex}}^{\left|V(a)\right|+\left|V(a)\right|} V_a$, where $B_{\text{vertex}}$ is a function of our parameters. In general, we will also need to upper bound some of the terms by $\prod_{e \in E(a)}(B_{\text{edge}}(e))$ where $B_{\text{edge}}(e)$ is a function of $l_e, \Omega$, and our parameters.

2. We generalize the bounds so that they apply to improper shapes as well as proper shapes. Under our simplifying assumptions, all we need to do here is to take isolated vertices into account. In general, we also need to handle multi-edges.

### 9.2.1 Simplified $B_{\text{norm}}(\alpha)$

Under our simplifying assumptions, we start with the following norm bound from the updated graph matrix norm bound paper [AMP20]:

**Theorem 9.12 (Simplified Graph Matrix Norm Bounds).** Under our simplifying assumptions, for all $\varepsilon > 0$ and all proper shapes $\alpha$, taking $c_{\alpha} = \left|V(\alpha)\right| \left(U_\alpha \cup V_\alpha\right) + \left|S_\alpha\right| \left(U_\alpha \cap V_\alpha\right)$,

$$Pr\left(||M_\alpha|| > \left(2\left|V_\alpha \setminus \left(U_\alpha \cap V_\alpha\right)\right|\right)^\frac{|V(\alpha)\setminus\left(U_\alpha \cap V_\alpha\right)|}{n^\frac{u(V(a)) + w(S_\alpha) - w(S_\alpha)}{2}} \right) < \varepsilon$$

where $q = 3 \left[\frac{\ln\left(\frac{2\left|V(\alpha)\right|}{3c_{\alpha}}\right)}{3c_{\alpha}}\right]$.

**Corollary 9.13.** For all shapes $\alpha$ and all $\varepsilon > 0$,

$$Pr\left(||M_\alpha|| > \left(2\left|V_\alpha\right|\sqrt{2eq}\right)^\frac{|V(\alpha)\setminus\left(U_\alpha \cap V_\alpha\right)|}{n^\frac{u(V(a)) + w(S_\alpha) - w(S_\alpha)}{2}} \right) < \varepsilon$$

where $q = 3 \left[\frac{\ln\left(\frac{2\left|V(\alpha)\right|}{3c_{\alpha}}\right)}{3c_{\alpha}}\right]$.

*Proof.* Observe that adding an isolated vertex to $\alpha$ is equivalent to multiplying $M_\alpha$ by $n - \left|V(\alpha)\right|$. Thus, if the bound holds for all proper $\alpha$ then it will hold for improper $\alpha$ as well.

We now make the following observations:

1. $\left|S_\alpha \setminus \left(U_\alpha \cap V_\alpha\right)\right| \leq \left|U_\alpha \setminus V_\alpha\right|$, so $c_{\alpha} = \left|W_\alpha\right| + \left|S_\alpha \setminus \left(U_\alpha \cap V_\alpha\right)\right| \leq \left|V(\alpha)\setminus V_\alpha\right|$. Similarly, $\left|S_\alpha \setminus \left(U_\alpha \cap V_\alpha\right)\right| \leq \left|V_\alpha \setminus U_\alpha\right|$, so $c_{\alpha} \leq \left|V(\alpha)\setminus U_\alpha\right|$. Thus, $c_{\alpha} \leq \frac{\left|V(\alpha)\setminus U_\alpha\right| + \left|V(\alpha)\setminus V_\alpha\right|}{2}$.

2. $\left|V(\alpha) \setminus \left(U_\alpha \cap V_\alpha\right)\right| \leq \left|V(\alpha) \setminus U_\alpha\right| + \left|V(\alpha) \setminus V_\alpha\right|$

Thus, by Theorem 9.12, for all proper shapes $\alpha$ and all $\varepsilon > 0$,

$$Pr\left(||M_\alpha|| > \left(2\left|V_\alpha\right|\sqrt{2eq}\right)^\frac{|V(\alpha)\setminus\left(U_\alpha \cap V_\alpha\right)|}{n^\frac{u(V(a)) + w(S_\alpha) - w(S_\alpha)}{2}} \right) < \varepsilon''$$

where $q = 3 \left[\frac{\ln\left(\frac{w(S_\alpha)}{3c_{\alpha}}\right)}{3c_{\alpha}}\right]$.
Corollary 9.14. For all $z \in \mathbb{N}$ and all $\varepsilon > 0$, taking $\varepsilon'' = \frac{\varepsilon}{5^2 \cdot 2^z}$, with probability at least $1 - \varepsilon$ we have that for all shapes $\alpha$ such that $|V(\alpha)| \leq z$,

$$||M_\alpha|| \leq \left(2 |V_\alpha| \sqrt{2eq} \right)^{|V(\alpha) \setminus U_\alpha| + |V(\alpha) \setminus V_\alpha|} n^{\frac{w(V(\alpha)) - w(S_\alpha)}{2}}$$

where $q = 3 \left\lceil \frac{ln(n)^3 \cdot (nV_\alpha)}{3c_a} \right\rceil$.

Proof. This result can be proved from Corollary 9.13 using a union bound and the following proposition:

Proposition 9.15. Under our simplifying assumptions, for all $z \in \mathbb{N}$, there are at most $5^2 \cdot 2^z$ proper shapes $\alpha$ such that $V(\alpha) \leq z$.

Proof. Observe that we can construct any proper shape $\alpha$ with at most $m$ vertices as follows:

1. Start with $z$ vertices $v_1, \ldots, v_z$.
2. For each vertex $v_i$, choose whether $v_i \in V(\alpha) \setminus U_\alpha \setminus V_\alpha$, $v_i \in U_\alpha \setminus V_\alpha$, $v_i \in U_\alpha \setminus V_\alpha$, or $v_i \notin V(\alpha)$.
3. For each pair of vertices $v_i, v_j \in V(\alpha)$, choose whether or not $(v_i, v_j) \in E(\alpha)$.

Corollary 9.16. For all $D_V \in \mathbb{N}$ and all $\varepsilon > 0$, taking

$$q = 3 \left\lceil \frac{ln(n)^3 \cdot (nD_V)}{3} \right\rceil = 3 \left\lceil D_V ln(n) + \frac{ln(\frac{1}{\varepsilon})}{3} + D_V ln(5) + 3D_V ln(2) \right\rceil,$$

$B_{\text{vertex}} = 6D_V \sqrt{2eq}$, and

$$B_{\text{norm}}(\alpha) = B_{\text{vertex}} \left| V(\alpha) \setminus U_\alpha \right| + \left| V(\alpha) \setminus V_\alpha \right| n^{\frac{w(V(\alpha)) - w(S_\alpha)}{2}},$$

with probability at least $(1 - \varepsilon)$ we have that for all shapes $\alpha \in \mathcal{M}'$, $||M_\alpha|| \leq B_{\text{norm}}(\alpha)$.

Proof. This follows from Corollary 9.14 and the fact that for all $\alpha \in \mathcal{M}'$, $w(S_\alpha) \leq |V(\alpha)| \leq 3D_V$.

9.2.2 General $B_{\text{norm}}(\alpha)$

In general, we start with the following norm bound from the updated graph matrix norm bound paper [AMP20]:

Theorem 9.17 (General Graph Matrix Norm Bounds). For all $\varepsilon > 0$ and all proper shapes $\alpha$, taking $q = \left\lceil \frac{ln(n \cdot w(S_\alpha))}{\varepsilon'} \right\rceil$

$$P \left( ||M_\alpha|| > 2e(2q |V(\alpha)| |V(\alpha) \setminus (U_\alpha \cap V_\alpha)| \prod_{e \in E(\alpha)} h^+_{\varepsilon}(B_{\Omega}(2qI_\varepsilon)) \right) n^{\frac{w(V(\alpha)) - w(S_\alpha)}{2}} < \varepsilon$$

78
Corollary 9.18. For all $\varepsilon > 0$, for all $z, l_{\text{max}}, m \in \mathbb{N}$, taking $\varepsilon' = \frac{\varepsilon}{5'(l_{\text{max}}+1)^3}$, with probability at least $1 - \varepsilon$, for all shapes $\alpha$ such that

1. $|V(\alpha)| \leq z$.
2. All edges in $E(\alpha)$ have label at most $l_{\text{max}}$.
3. All edges in $E(\alpha)$ have multiplicity at most $m$.

\[
||M_{\alpha}|| \leq 2e(2q|V(\alpha)|)|V(\alpha)\setminus U_{\alpha}| + |V(\alpha)\setminus V_{\alpha}| \left( \prod_{e \in E(\alpha)} 2h_{i_e}^+(B_{\Omega}(2m_{\text{max}})) \max_{j \in [0, m_{\text{max}}]} \left\{ \left( h_{i_j}^+(B_{\Omega}(2qj)) \right)^{\frac{l_{\text{max}}}{\max(|J_{\alpha}|)}} \right\} \right)
\]

where $q = \left\lceil \ln \left( \frac{n^{p_{\text{max}}}(S_{\alpha, \text{rand}})}{\varepsilon} \right) \right\rceil$.

Proof. Observe that for each $\alpha$ which has multi-edges, we can write $M_{\alpha} = \sum_i c_i M_{\alpha_i}$, where each $\alpha_i$ has no multiple edges. We first upper bound $\sum_i |c_i|$. Let

Lemma 9.19. For any $a_1, \ldots, a_m \in \mathbb{N} \cup \{0\}$, taking $p_{\text{max}} = \sum_{i=1}^m a_i$ and writing $\prod_{i=1}^m h_{a_i} = \sum_{k=0}^{p_{\text{max}}} c_k h_k$,

\[
\sum_{k=0}^{p_{\text{max}}} |c_k| \leq (p_{\text{max}} + 1) \prod_{i=1}^m h_{a_i}^+(B_{\Omega}(2p_{\text{max}})) \leq \prod_{i=1}^m 2h_{a_i}^+(B_{\Omega}(2p_{\text{max}}))
\]

Proof. Suppose $\prod_{i=1}^m (h_{a_i}(x))^2 = \sum_{k=0}^{2p_{\text{max}}} u_k x^k$ and $\prod_{i=1}^m (h_{a_i}(x))^2 = \sum_{k=0}^{2p_{\text{max}}} v_k x^k$. Then, note that $|u_k| \leq v_k$ and so,

\[
E_{\Omega}[(\prod_{i=1}^m (h_{a_i}(x))^2)] = \sum_{k=0}^{2p_{\text{max}}} u_k E_{\Omega}[x^k] \leq \sum_{k=0}^{2p_{\text{max}}} v_k E_{\Omega}[x^k] \leq \sum_{k=0}^{2p_{\text{max}}} v_k (B_{\Omega}(2p_{\text{max}}))^k = \prod_{i=1}^m (h_{a_i}^+(B_{\Omega}(2p_{\text{max}})))^2
\]

Therefore, using the fact that $h_k$ form an orthonormal basis,

\[
\sum_{k=0}^{p_{\text{max}}} c_k^2 = E_{\Omega}[(\sum_{k=0}^{p_{\text{max}}} c_k h_k(x))^2] = E_{\Omega}[(\prod_{i=1}^m (h_{a_i}(x))^2)] \leq \prod_{i=1}^m (h_{a_i}^+(B_{\Omega}(2p_{\text{max}})))^2
\]

This implies

\[
(\sum_{k=0}^{p_{\text{max}}} |c_k|)^2 \leq (p_{\text{max}} + 1) \prod_{k=0}^{p_{\text{max}}} c_k^2 \leq (p_{\text{max}} + 1) \prod_{i=1}^m (h_{a_i}^+(B_{\Omega}(2p_{\text{max}})))^2
\]

Taking square roots gives the inequality.

Corollary 9.20. For any shape $\alpha$ such that every edge of $\alpha$ has multiplicity at most $m$ and label at most $l_{\text{max}}$, if we write $M_{\alpha} = \sum_i c_i M_{\alpha_i}$ where each $\alpha_i$ has no multi-edges then $\sum_i |c_i| \leq \prod_{e \in E(\alpha)} 2h_{i_e}^+(B_{\Omega}(2m_{\text{max}}))$

The result now follows from Theorem 9.17 and the following observations:
1. \(|V(\alpha) \setminus (U_\alpha \cap V_\alpha) \leq |V(\alpha) \setminus U_\alpha| + |V(\alpha) \setminus V_\alpha|\).
2. For any \(\alpha\), writing \(M_\alpha = \sum_i c_i M_{\alpha_i}\), where each \(\alpha_i\) has no multi-edges, for all \(\alpha_i\),
\[
\omega(V(\alpha_i)) + \omega(I_{\alpha_i}) - \omega(S_{\alpha_i}) \leq \omega(V(\alpha)) + \omega(I_{\alpha}) - \omega(S_{\alpha_{\text{min}}})
\]
3. For any \(a_1, \ldots, a_m \in \mathbb{N} \cup \{0\}\) such that \(\forall i' \in [m], a_{i'} \leq l_{\text{max}}\), for all \(j \in [0, ml_{\text{max}}]\)
\[
h_j^+(B_\Omega(2qj)) \leq \prod_{i'=1}^m \left(h_j^+(B_\Omega(2qj'))\right)^{\frac{d_{i'}^j}{\max_i (d_{i'}^j)}} \leq \prod_{i'=1}^m \left(\max_{j \in [0, ml_{\text{max}}]} \left\{\left(h_j^+(B_\Omega(2qj'))\right)^{\frac{d_{i'}^j}{\max_i (d_{i'}^j)}}\right\}\right)
\]

**Proposition 9.21.** For all \(z, l_{\text{max}} \in \mathbb{N}\), there are at most \(5^z (l_{\text{max}} + 1)^z\) proper shapes \(\alpha\) such that \(|V(\alpha)| \leq z\) and every edge in \(E(\alpha)\).

**Proof.** This can be proved in the same way as before. Observe that we can construct any proper shape \(\alpha\) with at most \(z\) vertices as follows:

1. Start with \(z\) vertices \(v_1, \ldots, v_z\).
2. For each vertex \(v_i\), choose whether \(v_i \in V(\alpha) \setminus U_\alpha \cap V_\alpha\), \(v_i \in U_\alpha \setminus V_\alpha\), \(v_i \in V_\alpha \setminus U_\alpha\), \(v_i \in U_\alpha \cap V_\alpha\), or \(v_i \notin V(\alpha)\).
3. For each \(k\) tuple of vertices in \(V(\alpha)\), choose the label of the hyperedge between these vertices (or 0 if the hyperedge is not in \(E(\alpha)\)).

**Corollary 9.22.** For all \(D_V, D_E \in \mathbb{N}\) and all \(\epsilon > 0\), taking
\[
B_{\text{norm}}(\alpha) = 2e B_{\text{vertex}}^{V(\alpha) \setminus U_\alpha| + |V(\alpha) \setminus V_\alpha|} \left(\prod_{e \in E(\alpha)} B_{\text{edge}}(e)^{\omega(V(\alpha)) + \omega(I_{\alpha}) - \omega(S_{\alpha})}\right)
\]
where

1. \(q = \left\lceil \ln \left(n^{3D_V} \right) \right\rceil = \left[3D_V \ln(n) + \ln\left(\frac{1}{\epsilon}\right) + (3D_V)^4 \ln(D_E + 1) + 3D_V \ln(5)\right]\)
2. \(B_{\text{vertex}} = 6q D_V\)
3. \(B_{\text{edge}}(e) = 2h_j^+(B_\Omega(6D_V D_E)) \max_{j \in [0, 3D_V D_E]} \left\{\left(h_j^+(B_\Omega(2qj))\right)^{\frac{l_j}{\max_i (d_{i'}^j)}}\right\}\)

with probability at least \((1 - \epsilon)\), for all shapes \(\alpha \in \mathcal{M'}\), \(\|M_\alpha\| \leq B_{\text{norm}}(\alpha)\).

**9.3 Choosing \(B(\gamma)\)**

We now describe how to choose the function \(B(\gamma)\). Recall that we want the following conditions to hold:

1. For all \(\gamma, \tau, \gamma'\) and all intersection patterns \(P \in \mathcal{P}_{\gamma, \tau, \gamma'}\),
\[
B_{\text{norm}}(\tau_P) \leq B(\gamma) B(\gamma') B_{\text{norm}}(\tau)
\]

80
2. For all composable $\gamma_1, \gamma_2$, $B(\gamma_1)B(\gamma_2) = B(\gamma_1 \circ \gamma_2)$.

The most important part of choosing $B(\gamma)$ is to make sure that the factors of $n$ are controlled. For this, we use the following intersection tradeoff lemma. Under our simplifying assumptions, this lemma follows from [BHK+16, Lemma 7.12]. We defer the general proof of this lemma to the end of this section.

**Lemma 9.23** (Intersection Tradeoff Lemma). For all $\gamma, \tau, \gamma'$ and all intersection patterns $P \in \mathcal{P}_{\gamma, \tau, \gamma'}$, $w(V(\tau_p)) + w(I_{\gamma_p}) - w(S_{\tau_p, \min}) \leq w(V(\tau)) + w(I_{\gamma}) - w(S_{\tau, \min}) + w(V(\gamma) \setminus U_{\gamma}) + w(V(\gamma') \setminus U_{\gamma'})$

Based on this intersection tradeoff lemma, we can choose the function $B(\gamma)$ as follows.

**Corollary 9.24.** If we take

$$B_{\text{norm}}(a) = C \cdot \frac{B_{\text{vertex}}^{V(a) \setminus U_a + |V(a) \setminus V_a|} \prod_{e \in E(a)} B_{\text{edge}}(e) n^{\frac{w(V(a)) + w(I_a) - w(S_a)}{2}}}{n}$$

for some constant $C > 0$ and take

$$B(\gamma) = B_{\text{vertex}}^{V(\gamma) \setminus U_\gamma + |V(\gamma) \setminus V_\gamma|} \prod_{e \in E(\gamma)} B_{\text{edge}}(e) n^{\frac{w(V(\gamma)) - w(S_\gamma)}{2}}$$

then the following conditions hold:

1. For all $\gamma, \tau, \gamma'$ and all intersection patterns $P \in \mathcal{P}_{\gamma, \tau, \gamma'}$, $B_{\text{norm}}(\tau_p) \leq B(\gamma)B(\gamma')B_{\text{norm}}(\tau)$

2. For all composable $\gamma_1, \gamma_2$, $B(\gamma_1)B(\gamma_2) = B(\gamma_1 \circ \gamma_2)$.

**Proof.** We have that

$$B_{\text{norm}}(\tau_p) = B_{\text{vertex}}^{V(\tau_p) \setminus U_{\tau_p} + |V(\tau_p) \setminus V_{\tau_p}|} \prod_{e \in E(\tau_p)} B_{\text{edge}}(e) n^{\frac{w(V(\tau_p)) + w(I_{\tau_p}) - w(S_{\tau_p})}{2}}$$

and

$$B(\gamma)B(\gamma')B_{\text{norm}}(\tau) = B_{\text{vertex}}^{V(\gamma) \setminus U_\gamma + |V(\gamma) \setminus V_\gamma| + V(\gamma') \setminus U_{\gamma'} + |V(\gamma') \setminus V_{\gamma'}| + V(\tau) \setminus U_{\tau} + |V(\tau) \setminus V_{\tau}|} \prod_{e \in E(\gamma) \cup E(\gamma') \cup E(\tau)} B_{\text{edge}}(e) n^{\frac{w(V(\gamma)) + w(V(\gamma')) + w(V(\tau)) + w(S_\gamma) - w(S_{\gamma'})}{2}}$$

The first condition now follows immediately from the following observations:

1. $|V(\gamma) \setminus U_\gamma| + |V(\gamma) \setminus V_\gamma| + |V(\gamma') \setminus U_{\gamma'}| + |V(\gamma') \setminus V_{\gamma'}| + |V(\tau) \setminus U_{\tau}| + |V(\tau) \setminus V_{\tau}|$

   $= |V(\gamma \circ \tau \circ \gamma'^T) \setminus U_{\gamma \circ \tau \circ \gamma'^T}| + |V(\gamma \circ \tau \circ \gamma'^T) \setminus V_{\gamma \circ \tau \circ \gamma'^T}| \geq |V(\tau_p) \setminus U_{\tau_p}| + |V(\tau_p) \setminus V_{\tau_p}|$

2. $E(\tau_p) = E(\gamma) \cup E(\tau) \cup E(\gamma'^T)$ so $\prod_{e \in E(\tau_p)} B_{\text{edge}}(e) = \prod_{e \in E(\gamma) \cup E(\gamma') \cup E(\tau)} B_{\text{edge}}(e)$.

3. By the intersection tradeoff lemma,

   $w(V(\tau_p)) + w(I_{\tau_p}) - w(S_{\tau_p}) \leq w(V(\tau)) + w(I_\gamma) - w(S_\gamma) + w(V(\gamma) \setminus U_\gamma) + w(V(\gamma') \setminus U_{\gamma'})$

The second condition follows from the form of $B(\gamma)$.
9.4 Choosing \( N(\gamma) \)

To choose \( N(\gamma) \), we use the following lemma:

**Lemma 9.25.** For all \( D_V \in \mathbb{N} \), for all composable \( \gamma, \tau, \gamma'^T \) such that \(|V(\gamma)| \leq D_V\), \(|V(\tau)| \leq D_V\), and \(|V(\gamma')| \leq D_V\),

\[
\sum_{j > 0} \sum_{\gamma_1, \gamma_2, \ldots, \gamma_j \in \Gamma_{\gamma, \gamma'}} \prod_{i : \gamma_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma_i})|} \prod_{i : \gamma'_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma'_i})|} \sum_{P_1, \ldots, P_j : P_i \in P_{\gamma_i, \gamma_{i-1}, \gamma'_i}^{T}} \left( \prod_{i=1}^{j} N(P_i) \right)
\]

\[
\leq (3D_V)^{(2(|V(\gamma)|V_{\gamma}| + |V(\gamma')|V_{\gamma'})\text{ is non-trivial}) + (|V(\gamma)|V_{\gamma}| - |V(\gamma')|V_{\gamma'})\text{ is non-trivial})}
\]

**Proof sketch.** Observe that aside from the orderings (which are canceled out by the \(|\text{Aut}(U_{\gamma_i})|\) and \(|\text{Aut}(U_{\gamma'_i})|\) factors), the intersection patterns \( \{P_i : i \in [j]\} \) are determined by the following data on each vertex \( v \in (V(\gamma) \setminus V_{\gamma_i}) \cup (V(\gamma') \setminus V_{\gamma'_i}) \):

1. The first \( i \in [j] \) such that \( v \in (V(\gamma_i) \setminus V_{\gamma_i}) \cup (V(\gamma'_i) \setminus V_{\gamma'_i}) \). There are at most \( j \) possibilities for this.
2. A vertex \( u \) (if one exists) in \( V(\gamma_{i-1} \circ \cdots \circ (\gamma_1 \circ \tau \circ \gamma_1^T \cdots \circ \gamma_{i-1}^T) \) such that \( u \) and \( v \) are equal. There are at most \( 3D_V \) possibilities for this.

Using these observations and taking \( j_{\text{max}} = |V(\gamma) \setminus V_{\gamma_i}| + |V(\gamma') \setminus V_{\gamma'_i}| \),

\[
\sum_{j > 0} \sum_{\gamma_1, \gamma_2, \ldots, \gamma_j \in \Gamma_{\gamma, \gamma'}} \prod_{i : \gamma_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma_i})|} \prod_{i : \gamma'_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma'_i})|} \sum_{P_1, \ldots, P_j : P_i \in P_{\gamma_i, \gamma_{i-1}, \gamma'_i}^{T}} 1
\]

\[
\leq j_{\text{max}} \sum_{j=1}^{j_{\text{max}}} \frac{(3D_V)^{2(|V(\gamma)|V_{\gamma}| + |V(\gamma')|V_{\gamma'})\text{ is non-trivial}) + (|V(\gamma)|V_{\gamma}| - |V(\gamma')|V_{\gamma'})\text{ is non-trivial})}}{(3D_V)^{2(|V(\gamma)|V_{\gamma}| + |V(\gamma')|V_{\gamma'})\text{ is non-trivial}) + (|V(\gamma)|V_{\gamma}| - |V(\gamma')|V_{\gamma'})\text{ is non-trivial})}}
\]

Now recall that by Lemma, for any \( \gamma_i, \tau_{\gamma_{i-1}}^T \) and any intersection pattern \( P_i \in \mathcal{P}_{\gamma_i, \tau_{\gamma_{i-1}}^T} \),

\[
N(P_i) \leq |V(\tau_{\gamma_i})| |V(\gamma)| + |V(\gamma')| \leq (3D_V)^{|V(\gamma)|V_{\gamma}| + |V(\gamma')|V_{\gamma'}}
\]

Thus, for any \( P_1, \ldots, P_j : P_i \in \mathcal{P}_{\gamma_i, \tau_{\gamma_{i-1}}^T} \), \( \prod_{i=1}^{j} N(P_i) \leq (3D_V)^{|V(\gamma)|V_{\gamma}| + |V(\gamma')|V_{\gamma'}} \). Putting everything together, the result follows.
Corollary 9.26. For all $D_V \in \mathbb{N}$, if we take $N(\gamma) = (3D_V)^{2|V(\gamma)\setminus V_\tau|+|V(\gamma)\setminus U_\tau|}$ then for all composable $\gamma, \tau, \gamma^T$ such that $|V(\gamma)| \leq D_V$, $|V(\tau)| \leq D_V$, and $|V(\gamma')| \leq D_V$,

$$\sum_{f>0} \prod_{\gamma_1, \gamma_2, \ldots, \gamma_f \in \Gamma, \gamma_i \mid \gamma} \frac{1}{|\text{Aut}(U_{\gamma_i})|} \prod_{i: \gamma_i \text{ is non-trivial}} \frac{1}{|\text{Aut}(U_{\gamma_i})|} \sum_{\gamma_1, \gamma_2, \ldots, \gamma_f \in \Gamma, \gamma_i \mid \gamma} \left( \prod_{i=1}^f N(P_i) \right) \leq \frac{N(\gamma)N(\gamma')}{{(|\text{Aut}(U_{\gamma})|)^{1/2}} \text{ is non-trivial} \cdot (|\text{Aut}(U_{\gamma'})|)^{1/2} \text{ is non-trivial}}$$

9.5 Choosing $c(\alpha)$

In this section, we describe how to choose $c(\alpha)$. For simplicity, we first describe how to choose $c(\alpha)$ under our simplifying assumptions. We then describe the minor adjustments that are needed when we have hyperedges and multiple types of vertices.

Lemma 9.27. Under our simplifying assumptions, for all $U \in \mathcal{I}_{\text{mid}}$,

$$\sum_{\alpha: U_\alpha \equiv U, \alpha \text{ is proper and non-trivial}} \frac{1}{|\text{Aut}(U_\alpha \cap V_\alpha)|} |(3D_V)^{|U_\alpha \setminus V_\alpha|+|V_\alpha \setminus U_\alpha|+2|E(\alpha)|}2|V(\alpha)\setminus(U_\alpha \cup V_\alpha)| < 5$$

Proof. In order to choose $\alpha$, it is sufficient to choose the following:

1. The number $j_1$ of vertices in $U_\alpha \setminus V_\alpha$, the number $j_2$ of vertices in $V_\alpha \setminus U_\alpha$, and the number $j_3$ of vertices in $V(\alpha) \setminus (U_\alpha \cup V_\alpha)$.
2. A mapping in $\text{Aut}(U_\alpha \cap V_\alpha)$ determining how the vertices in $U_\alpha \cap V_\alpha$ match up with each other.
3. The position of each vertex $u \in U_\alpha \setminus V_\alpha$ within $U_\alpha$ (there are at most $|U_\alpha| \leq D_V$ choices for this).
4. The position of each vertex $v \in V_\alpha \setminus U_\alpha$ within $V_\alpha$ (there are at most $|U_\alpha| \leq D_V$ choices for this).
5. The number $j_4$ of edges in $E(\alpha)$.
6. The endpoints of each edge in $E(\alpha)$.

This implies that for all $j_1, j_2, j_3, j_4 \geq 0$

$$\sum_{\alpha: U_\alpha \equiv U, \alpha \text{ is proper and non-trivial}} \frac{1}{|\text{Aut}(U_\alpha \cap V_\alpha)|} |(3D_V)^{|U_\alpha \setminus V_\alpha|+|V_\alpha \setminus U_\alpha|+2|E(\alpha)|}2|V(\alpha)\setminus(U_\alpha \cup V_\alpha)| \leq 1$$

Using this, we have that

$$\sum_{\alpha: U_\alpha \equiv U, \alpha \text{ is proper and non-trivial}} \frac{1}{|\text{Aut}(U_\alpha \cap V_\alpha)|} |(3D_V)^{|U_\alpha \setminus V_\alpha|+|V_\alpha \setminus U_\alpha|+2|E(\alpha)|}2|V(\alpha)\setminus(U_\alpha \cup V_\alpha)| \leq \sum_{j_1, j_2, j_3, j_4 \in \mathbb{N}, j_1+j_2+j_3+j_4 \geq 1} \frac{1}{3^{j_1} \cdot 2^{j_2} \cdot 3^{j_3} \cdot 2^{j_4}} \leq 2 \left( \frac{3}{2} \right)^2 \frac{9}{8} - 1 < 5$$
Corollary 9.28. For all \( \varepsilon' > 0 \), if we take

\[
c(\alpha) = \frac{5(3D_V)^{|U_\alpha \setminus V_\alpha| + |V_\alpha \setminus U_\alpha| + 2|E(\alpha)|2|V(\alpha)\setminus (U_\alpha \cup V_\alpha)|}}{\varepsilon'}
\]

then

1. \( \forall U \in I_{\text{mid}}, \sum_{\gamma \in \Gamma_{U,\ast}} \frac{1}{|\text{Aut}(U)|c(\gamma)} < \varepsilon' \)
2. \( \forall V \in I_{\text{mid}}, \sum_{\gamma \in \Gamma_{V,\ast}} \frac{1}{|\text{Aut}(V)|c(\gamma)} < \varepsilon' \)
3. \( \forall U \in I_{\text{mid}}, \sum_{\tau \in M_{U}} \frac{1}{|\text{Aut}(U)|c(\tau)} < \varepsilon' \)

9.5.1 Choosing \( c(\alpha) \) in general*

When we have multiple types of vertices and hyperedges of arity \( k \), Lemma 9.27 can be generalized as follows:

Lemma 9.29. Under our simplifying assumptions, for all \( U \in I_{\text{mid}} \),

\[
\sum_{\alpha: U_\alpha \equiv U, \alpha \text{ is proper and non-trivial}} \frac{1}{|\text{Aut}(U_\alpha \cap V_\alpha)|} \leq \frac{5}{(3D_V t_{\max})^{|U_\alpha \setminus V_\alpha| + |V_\alpha \setminus U_\alpha| + k|E(\alpha)|2|V(\alpha)\setminus (U_\alpha \cup V_\alpha)|}} < 5
\]

Proof sketch. This can be proved in the same way as Lemma 9.27 with the following modifications:

1. In addition to choosing the number of vertices in \( U_\alpha \setminus V_\alpha \), \( V_\alpha \setminus U_\alpha \), and \( V(\alpha) \setminus (U_\alpha \cap V_\alpha) \), we also have to choose the types of these vertices.
2. For each hyperedge, we have to choose \( k \) endpoints rather than 2 endpoints.

Corollary 9.30. For all \( \varepsilon' > 0 \), if we take

\[
c(\alpha) = \frac{5(3t_{\max} D_V)^{|U_\alpha \setminus V_\alpha| + |V_\alpha \setminus U_\alpha| + k|E(\alpha)|2|V(\alpha)\setminus (U_\alpha \cup V_\alpha)|}}{\varepsilon'}
\]

then

1. \( \forall U \in I_{\text{mid}}, \sum_{\gamma \in \Gamma_{U,\ast}} \frac{1}{|\text{Aut}(U)|c(\gamma)} < \varepsilon' \)
2. \( \forall V \in I_{\text{mid}}, \sum_{\gamma \in \Gamma_{V,\ast}} \frac{1}{|\text{Aut}(V)|c(\gamma)} < \varepsilon' \)
3. \( \forall U \in I_{\text{mid}}, \sum_{\tau \in M_{U}} \frac{1}{|\text{Aut}(U)|c(\tau)} < \varepsilon' \)

For technical reasons, we will need a more refined bound when the sum is over all shapes \( \gamma \) of at least a prescribed size.

Lemma 9.31. For all \( \varepsilon' > 0 \), for the same choice of \( c(\alpha) \) as in Corollary 9.30, for any \( U \in I_{\text{mid}} \) and integer \( m \geq 1 \), we have

\[
\sum_{\gamma \in \Gamma_{U,\ast}, |V(\gamma)| \geq |U| + m} \frac{1}{|\text{Aut}(U)|c(\gamma)} \leq \frac{\varepsilon'}{5 \cdot 2^{m-1}}
\]
Proof sketch. The proof is similar to the proof of Corollary 9.30, but we now have the extra condition $j_2 + j_3 \geq m$ in the proof of Lemma 9.27. Then,

$$
\sum_{j_1, j_2, j_3 \in \mathbb{N} \cup \{0\} : j_1 + j_2 + j_3 \geq m} \frac{1}{3^{j_1} + 3^{j_2} + 3^{j_3}} \leq \sum_{j_1, j_2 \in \mathbb{N} \cup \{0\}} \frac{1}{2^{3^{j_1} + 3^{j_2}}} = \frac{27}{16} \cdot 2^m \leq \frac{1}{2^{m-1}}
$$

9.6 Proof of the Generalized Intersection Tradeoff Lemma

We now prove the generalized intersection tradeoff lemma.

Lemma 9.32. For all $\gamma, \tau, \gamma'$ and all intersection patterns $P \in \mathcal{P}_{\gamma, \tau, \gamma}'$,

$$w(V(\tau_P)) + w(I_{\tau_P}) - w(S_{\tau_P, \min}) \leq w(V(\tau)) + w(I_\tau) - w(S_{\tau, \min}) + w(V(\gamma) \setminus U_\tau) + w(V(\gamma') \setminus U_{\gamma'})$$

Proof.

Definition 9.33.

1. We define $I_{LM}$ to be the set of vertices which, after intersections, touch $\gamma$ and $\tau$ but not $\gamma'^T$. In particular, $I_{LM}$ consists of the vertices which result from intersecting a pair of vertices in $V(\gamma) \setminus V_\gamma$ and $V(\tau) \setminus U_\tau \setminus V_\tau$ and the vertices which are in $U_\tau \setminus V_\tau$ and are not intersected with any other vertex.

2. We define $I_{MR}$ to be the set of vertices which, after intersections, touch $\tau$ and $\gamma'^T$ but not $\gamma$. In particular, $I_{MR}$ consists of the vertices which result from intersecting a pair of vertices in $V(\gamma) \setminus U_\tau \setminus V_\tau$ and $V(\gamma'^T) \setminus U_{\gamma'^T}$ and the vertices which are in $V_\tau \setminus U_\tau$ and are not intersected with any other vertex.

3. We define $I_{LR}$ to be the set of vertices which, after intersections, touch $\gamma$ and $\gamma'^T$ but not $\tau$. In particular, $I_{LR}$ consists of the vertices which result from intersecting a pair of vertices in $V(\gamma) \setminus V_\gamma$ and $V(\gamma'^T) \setminus U_{\gamma'^T}$.

4. We define $I_{LMR}$ to be the set of vertices which, after intersections, touch $\gamma$, $\tau$, and $\gamma'^T$. In particular, $I_{LMR}$ consists of the vertices which result from intersecting a triple of vertices in $V(\gamma) \setminus V_\gamma$, $V(\tau) \setminus U_\tau \setminus V_\tau$, and $V(\gamma'^T) \setminus U_{\gamma'^T}$, intersecting a pair of vertices in $V(\gamma) \setminus V_\gamma$ and $V(\gamma'^T) \setminus U_{\gamma'^T}$, and single vertices in $U_\tau \cap V_\tau$.

The main idea is as follows. A priori, any of the vertices in $I_{LM} \cup I_{MR} \cup I_{LR} \cup I_{LMR}$ could become isolated. We handle this by keeping track of the following types of flows:

1. Flows from $U_\gamma$ to $I_{LM} \cup I_{LR} \cup I_{LMR}$
2. Flows from $I_{LR} \cup I_{MR} \cup I_{LMR}$ to $V_{\gamma'^T}$
3. Flows from $I_{LM}$ to $I_{MR}$. For technical reasons, we also view vertices in $I_{LMR}$ as having flow to themselves.

We then observe that flows to and from these vertices prevent these vertices from being isolated and can provide flow from $U_\gamma$ to $V_{\gamma'^T}$, which gives a lower bound on $w(S_{\tau_P})$.

We now implement this idea.
Lemma 9.37. The maximum flow from $U$.

Proof sketch. Observe that if we have a cut $(v, w)$ which is an edge of multiplicity 1 in $E(\alpha)$ (or part of a hyperedge of multiplicity 1 in $E(\alpha)$), we create a directed edge with infinite capacity from $v_{out}$ to $w_{in}$ and we create a directed edge with infinite capacity from $w_{out}$ to $v_{in}$.

3. We define $U_{H_\alpha}$ to be $U_{H_\alpha} = \{u_{in} : u \in U_\alpha\}$ and we define $V_{H_\alpha}$ to be $V_{H_\alpha} = \{v_{out} : v \in V_\alpha\}$

Lemma 9.35. The maximum flow from $U_{H_\alpha}$ to $V_{H_\alpha}$ is equal to the minimum weight of a separator between $U_\alpha$ and $V_\alpha$.

Proof. This can be proved using the max flow min cut theorem.

Definition 9.36 (Modified Flow Graph). Given a shape $\alpha$ together with a set $I_L \subseteq V(\alpha)$ of vertices in $\alpha$ (which will be the vertices in $\alpha$ which are intersected with a vertex to the left of $\alpha$) and a set $I_R \subseteq V(\alpha)$ of vertices in $\alpha$ (which will be the vertices in $\alpha$ which are intersected with a vertex to the right of $\alpha$), we define the modified flow graph $H^{I_L, I_R}_\alpha$ as follows:

1. We start with the flow graph $H_\alpha$
2. For each vertex $u \in I_L$, we delete all of the edges into $u_{in}$ and add $u_{in}$ to $U_{H_\alpha}$
3. For each vertex $v \in I_R$, we delete all of the edges out of $v_{out}$ and add $v_{out}$ to $V_{H_\alpha}$
4. We call the resulting graph $H^{I_L, I_R}_\alpha$ and the resulting sets $U_{H^{I_L, I_R}_\alpha}$ and $V_{H^{I_L, I_R}_\alpha}$

Lemma 9.37. The maximum flow from $U_{H^{I_L, I_R}_\alpha}$ to $V_{H^{I_L, I_R}_\alpha}$ in $H^{I_L, I_R}_\alpha$ is at least as large as the maximum flow from $U_{H_\alpha}$ to $V_{H_\alpha}$ in $H_\alpha$.

Proof sketch. Observe that if we have a cut $C$ in $H^{I_L, I_R}_\alpha$ which separates $U_{H^{I_L, I_R}_\alpha}$ and $V_{H^{I_L, I_R}_\alpha}$ then $C$ separates $U_{H_\alpha}$ and $V_{H_\alpha}$ in $H_\alpha$.

Before the intersections, we have the following flows.

1. We take $F_1$ to be the maximum flow from $U_{\gamma}$ to $V_{\gamma}$ in $\gamma$. Note that $F_1$ has value $w(V_{\gamma})$
2. We take $F_2$ to be the maximum flow from $U_{\tau}$ to $V_{\tau}$ in $\tau$. Note that $F_2$ has value $w(S_{\tau, min})$
3. We take $F_3$ to be the maximum flow from $U_{\gamma^T}$ to $V_{\gamma^T}$ in $\gamma^T$. Note that $F_1$ has value $w(U_{\gamma^T})$

After the intersections, we take the following flows:

1. We take $F'_1$ to be the maximum flow from $U_{H_\gamma^{I_L, I_R, LMR}}$ to $V_{H_\gamma^{I_L, I_R, LMR}}$ in $H^{I_L, I_R, LMR}_\gamma$
2. We take $F'_2$ to be the maximum flow from $U_{H_{\gamma^T}^{I_L, I_R, LMR}}$ to $V_{H_{\gamma^T}^{I_L, I_R, LMR}}$ in $H^{I_L, I_R, LMR, I_M}_\gamma$
3. We take $F'_3$ to be the maximum flow from $U_{H_{\gamma^T}^{I_L, I_R, LMR, \emptyset}}$ to $V_{H_{\gamma^T}^{I_L, I_R, LMR, \emptyset}}$ in $H^{I_L, I_R, LMR, \emptyset}_\gamma$.
Observe that because of how intersection patterns are defined, \( \text{val}(F_1') = w(U_1) \) and \( \text{val}(F_3') = w(V_{\gamma,T}) \). By Lemma 9.37, the value of \( F_2' \) is at least as large as the value of \( F_2 \), so \( \text{val}(F_2') \geq w(S,_{\min}) \).

We now consider \( F_1' + F_2' + F_3' \). As is, this is not a flow, but we can fix this.

**Definition 9.38.** For each vertex \( v \in V(\tau_P) \),

1. We define \( f_{\text{in}}(v) \) to be the flow into \( v \in F_1' + F_2' + F_3' \).
2. We define \( f_{\text{out}}(v) \) to be the flow out of \( v \in F_1' + F_2' + F_3' \).
3. We define \( f_{\text{through}}(v) \) to be the flow from \( v_{\text{in}} \) to \( v_{\text{out}} \) in \( F_1' + F_2' + F_3' \).
4. We define \( f_{\text{imbalance}}(v) \) to be \( f_{\text{imbalance}}(v) = |f_{\text{in}}(v) - f_{\text{out}}(v)| \).
5. We define \( f_{\text{excess}}(v) \) to be \( f_{\text{excess}}(v) = f_{\text{through}}(v) - \max\{f_{\text{in}}(v), f_{\text{out}}(v)\} \).

With this information, we fix the flow \( F_1' + F_2' + F_3' \) as follows. For each vertex \( v \in V(\tau_P) \),

1. If \( f_{\text{in}}(v) > f_{\text{out}}(v) \) then we create a vertex \( v_{\text{supplemental, out}} \) and an edge from \( v_{\text{out}} \) to \( v_{\text{supplemental, out}} \) with capacity \( f_{\text{imbalance}}(v) \) and we route \( f_{\text{imbalance}}(v) \) of flow along this edge. We then add \( v_{\text{supplemental, out}} \) to a set of vertices \( V_{\text{supplemental}} \).
2. If \( f_{\text{in}}(v) < f_{\text{out}}(v) \) then we create a vertex \( v_{\text{supplemental, in}} \) and an edge from \( v_{\text{supplemental, in}} \) to \( v_{\text{in}} \) with capacity \( f_{\text{imbalance}}(v) \) and we route \( f_{\text{imbalance}}(v) \) of flow along this edge. We then add \( v_{\text{supplemental, out}} \) to a set of vertices \( V_{\text{supplemental}} \).
3. We reduce the flow on the edge from \( v_{\text{in}} \) to \( v_{\text{out}} \) by \( f_{\text{excess}}(v) \).

We call the resulting flow \( F' \).

**Proposition 9.39.** \( F' \) is a flow from \( U_{H_{\gamma}^{\text{ILR}}} \cup U_{\text{supplemental}} \) to \( V_{H_{\gamma}^{\text{ILR}}} \cup V_{\text{supplemental}} \) with value \( \text{val}(F') = \text{val}(F_1') + \text{val}(F_2') + \text{val}(F_3') - \sum_{v \in V(\tau)} f_{\text{excess}}(v) \).

**Corollary 9.40.** There exists a flow \( F'' \) from \( U_{H_{\gamma}^{\text{ILR}}} \) to \( V_{H_{\gamma}^{\text{ILR}}} \) with value \( \text{val}(F'') \geq \text{val}(F_1') + \text{val}(F_2') + \text{val}(F_3') - \sum_{v \in V(\tau)} (f_{\text{excess}}(v) + f_{\text{imbalance}}(v)) \).

**Proof.** Consider the minimum cut \( C \) between \( U_{\gamma}^{\text{ILR}} \) and \( V_{\gamma}^{\text{ILR}} \). If we add all of the supplemental edges to \( C \) then this gives a cut \( C' \) between \( U_{\gamma}^{\text{ILR}} \) and \( V_{\gamma}^{\text{ILR}} \) with capacity \( \text{capacity}(C') = \text{capacity}(C) + \sum_{v \in V(\tau)} f_{\text{imbalance}}(v) \geq \text{val}(F') \).

Thus, \( \text{capacity}(C) \geq \text{val}(F') - \sum_{v \in V(\tau)} f_{\text{imbalance}}(v) \) so there exists a flow \( F'' \) from \( U_{\gamma}^{\text{ILR}} \) to \( V_{\gamma}^{\text{ILR}} \) with value

\[
\text{val}(F'') = \text{capacity}(C) \geq \text{val}(F_1') + \text{val}(F_2') + \text{val}(F_3') - \sum_{v \in V(\tau)} (f_{\text{excess}}(v) + f_{\text{imbalance}}(v))
\]

We now make the following observations:

87
Lemma 9.41.

1. For all vertices \( v \not\in I_{LM} \cup I_{MR} \cup I_{LR} \cup I_{LMR} \), \( f_{\text{excess}}(v) = f_{\text{imbalance}}(v) = 0 \) (and these vertices can never be isolated).

2. For all vertices \( v \in I_{LM} \), \( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) \leq w(v) \). Moreover, for all vertices \( v \in I_{LM} \) which are isolated, \( f_{\text{excess}}(v) = f_{\text{imbalance}}(v) = 0 \).

3. For all vertices \( v \in I_{MR} \), \( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) \leq w(v) \). Moreover, for all vertices \( v \in I_{LM} \) which are isolated, \( f_{\text{excess}}(v) = f_{\text{imbalance}}(v) = 0 \).

4. For all vertices \( v \in I_{LR} \), \( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) \leq w(v) \). Moreover, for all vertices \( v \in I_{LM} \) which are isolated, \( f_{\text{excess}}(v) = f_{\text{imbalance}}(v) = 0 \).

5. For all vertices \( v \in I_{LMR} \), \( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) \leq 2w(v) \). Moreover, for all vertices \( v \in I_{LMR} \) which are isolated, \( f_{\text{excess}}(v) = w(v) \) and \( f_{\text{imbalance}}(v) = 0 \).

Proof. For the first statement, observe that for vertices \( v \not\in I_{LM} \cup I_{MR} \cup I_{LR} \cup I_{LMR} \), neither \( v_{in} \) nor \( v_{out} \) is ever a sink or source so the flow into these vertices must equal the flow out of these vertices and thus \( f_{in}(v) = f_{out}(v) = f_{\text{through}}(v) \).

For the second statement, observe that for a vertex \( v \in I_{LM} \),

1. \( F_1' \) will have a flow of \( f_{in}(v) \) into \( v_{in} \) and along the edge from \( v_{in} \) to \( v_{out} \).
2. \( F_2' \) will have a flow of \( f_{out}(v) \) along the edge from \( v_{in} \) to \( v_{out} \) and out of \( v_{out} \).

Thus, \( f_{\text{excess}}(v) = f_{in}(v) + f_{out}(v) - \max\{f_{in}(v), f_{out}(v)\} \). Since \( f_{\text{imbalance}}(v) = \left| f_{in}(v) - f_{out}(v) \right| \), \( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) = f_{in}(v) + f_{out}(v) - \max\{f_{in}(v), f_{out}(v)\} \leq w(v) \).

If \( v \) is isolated then neither \( F_1' \) nor \( F_2' \) can have any flow to \( v_{in} \) or out of \( v_{out} \) so \( f_{in}(v) = f_{\text{through}}(v) = f_{out}(v) = 0 \).

The third and fourth statements can be proved in the same way as the second statement.

For the fifth statement, observe that for a vertex \( v \in I_{LMR} \),

1. \( F_1' \) will have a flow of \( f_{in}(v) \) into \( v_{in} \) and along the edge from \( v_{in} \) to \( v_{out} \).
2. \( F_2' \) will have a flow of \( f_{out}(v) \) along the edge from \( v_{in} \) to \( v_{out} \) and out of \( v_{out} \).
3. \( F_3' \) will have a flow of \( f_{out}(v) \) along the edge from \( v_{in} \) to \( v_{out} \) and out of \( v_{out} \).

Thus, \( f_{\text{excess}}(v) = w(v) + f_{in}(v) + f_{out}(v) - \max\{f_{in}(v), f_{out}(v)\} \). Since \( f_{\text{imbalance}}(v) = \left| f_{in}(v) - f_{out}(v) \right| \), \( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) = w(v) + f_{in}(v) + f_{out}(v) - \max\{f_{in}(v), f_{out}(v)\} \leq 2w(v) \).

If \( v \) is isolated then neither \( F_1' \) nor \( F_2' \) can have any flow to \( v_{in} \) or out of \( v_{out} \) so \( f_{in}(v) = f_{out}(v) = 0 \) and \( f_{\text{through}}(v) = w(v) \).

Putting everything together, we have the following corollary:

Corollary 9.42.

\[
\sum_{v \in V(T_p)} \left( f_{\text{excess}}(v) + f_{\text{imbalance}}(v) \right) \leq w(I_{LM}) + w(I_{LR}) + w(I_{MR}) + 2w(I_{LMR}) - (w(I_{tp}) - w(I_{r}))
\]
Combining this with Corollary 9.40,

\[ w(S_{\tau,\text{min}}) \geq \text{val}(F_1') + \text{val}(F_2') + \text{val}(F_3') - \sum_{v \in V(\tau')} (f_{\text{excess}}(v) + f_{\text{imbalance}}(v)) \]

\[ \geq w(U_\tau) + w(S_{\tau,\text{min}}) + w(V_{\gamma,\tau}) - w(I_{LM}) - w(I_{LR}) - w(I_{MR}) - 2w(I_{LMR}) + (w(I_{\tau}) - w(I_{\tau})) \]

Since \( w(V(\tau')) = w(V(\tau)) + w(V(\gamma)) + w(V(\gamma')) - w(I_{LM}) - w(I_{LR}) - w(I_{MR}) - 2w(I_{LMR}), \)

\[ w(S_{\tau,\text{min}}) \geq w(U_\tau) + w(S_{\tau,\text{min}}) + w(V_{\gamma,\tau}) + w(V(\tau')) - w(V(\tau)) - w(V(\gamma)) - w(V(\gamma')) + (w(I_{\tau}) - w(I_{\tau})) \]

Rearranging this gives

\[ w(V(\tau')) - w(S_{\tau,\text{min}}) + w(I_{\tau}) \leq w(V(\tau)) - w(S_{\tau,\text{min}}) + w(I_{\tau}) + w(V(\gamma) \setminus U_\tau) + w(V(\gamma') \setminus U_{\gamma'}) \]

which is the generalized intersection tradeoff lemma. \( \blacksquare \)

10 Showing Positivity

10.1 General strategy to lower bound \( \sum_{V \in \mathcal{I}_{\text{mid}}} M^{\text{fact}}(H_{Id_V}) \)

In this section, we describe how to show that \( \sum_{V \in \mathcal{I}_{\text{mid}}} M^{\text{fact}}(H_{Id_V}) \geq \delta Id_{\text{Sym}} \) for some \( \delta > 0 \) where \( \delta \) will depend on \( n \) and other parameters. For now, we assume that the indices of \( \Lambda \) are multilinear monomials. We will then describe the adjustments that are needed to handle non-multilinear matrix indices.

We start with a few more definitions.

**Definition 10.1.** For all \( V \in \mathcal{I}_{\text{mid}} \) we define \( Id_{\text{Sym},V} \) to be the matrix such that

1. \( Id_{\text{Sym},V}(A, B) = 1 \) if \( A \) and \( B \) both have index shape \( V \).
2. Otherwise, \( Id_{\text{Sym},V}(A, B) = 0 \).

**Proposition 10.2.** \( Id_{\text{Sym}} = \sum_{V \in \mathcal{I}_{\text{mid}}} Id_{\text{Sym},V} \)

**Definition 10.3.** For all \( V \in \mathcal{I}_{\text{mid}} \) we define \( \lambda_V = |\text{Aut}(V)| H_{Id_V}(Id_V, Id_V) \)

We now describe our strategy for showing \( \sum_{V \in \mathcal{I}_{\text{mid}}} M^{\text{fact}}(H_{Id_V}) \geq \delta Id_{\text{Sym}} \). The idea is as follows. We will consider the index shapes \( V \in \mathcal{I}_{\text{mid}} \) from largest weight to smallest weight and we will show that for each \( V \in \mathcal{I}_{\text{mid}} \), there exists a \( \delta' > 0 \) such that \( \sum_{V \in \mathcal{I}_{\text{mid}}} M^{\text{fact}}(H_{Id_V}) \geq \delta' \sum_{U \in \mathcal{I}_{\text{mid}}; w(U) \geq w(V)} Id_{\text{Sym},U} \).

For the first step, letting \( V_{\text{max}} \) be the maximum weight index shape in \( \mathcal{I}_{\text{mid}}, M^{\text{fact}}(H_{Id_{V_{\text{max}}}}) = \lambda_{V_{\text{max}}} Id_{\text{Sym},V_{\text{max}}} \) because there are no non-trivial left shapes \( \sigma \) such that \( V_{\sigma} = V_{\text{max}} \). For other \( V \in \mathcal{I}_{\text{mid}}, \lambda_V Id_{\text{Sym},V} \) is a part of \( M^{\text{fact}}(H_{Id_V}) \) but \( M^{\text{fact}}(H_{Id_V}) \) will also contain terms of the form \( H_{Id_V}(\sigma, \sigma')M_\sigma M_{\sigma'} \) where \( U_\sigma \neq V \) or \( U_{\sigma'} \neq V \).

We can handle the terms \( H_{Id_V}(\sigma, \sigma')M_\sigma M_{\sigma'} \) where \( U_\sigma \neq V \) and \( U_{\sigma'} \neq V \) by bounding these terms in terms of \( Id_{\text{Sym},U_\sigma} \) and \( Id_{\text{Sym},U_{\sigma'}} \). Since \( w(U_\sigma) > w(V) \) and \( w(U_{\sigma'}) > w(V) \), \( Id_{\text{Sym},U_\sigma} \) and \( Id_{\text{Sym},U_{\sigma'}} \) are already available to us. To handle the terms \( H_{Id_V}(\sigma, \sigma')M_\sigma M_{\sigma'} \) where exactly one of \( U_\sigma \) and \( U_{\sigma'} \) are equal to \( V \), we use the following trick.

89
Definition 10.4. Given $V \in \mathcal{I}_{mid}$, define $H''_{Id_V}$ to be the coefficient matrix such that

1. If $U_{\sigma} = U_{\sigma'} = V$ then $H''_{Id_V}(\sigma, \sigma') = \frac{1}{2}H_{Id_V}(\sigma, \sigma')$
2. If exactly one of $U_{\sigma}$ and $U_{\sigma'}$ are equal to $V$ then $H''_{Id_V}(\sigma, \sigma') = H_{Id_V}(\sigma, \sigma')$
3. If $U_{\sigma} \neq V$ and $U_{\sigma'} \neq V$ then $H''_{Id_V}(\sigma, \sigma') = 2H_{Id_V}(\sigma, \sigma')$

Proposition 10.5. $M_{fact}(H''_{Id_V}) \geq 0$

Proof. Since $H_{Id_V} \succeq 0$, $H''_{Id_V} \succeq 0$ and thus $M_{fact}(H''_{Id_V}) \geq 0$.

Corollary 10.6. For all $V \in \mathcal{I}_{mid}$,

$$M_{fact}(H_{Id_V}) + \sum_{\sigma, \sigma' \in \mathcal{L}_V: U_{\sigma} \neq V, U_{\sigma'} \neq V} H_{Id_V}(\sigma, \sigma') M_{\sigma} M_{\sigma'} \succeq \frac{\lambda_V}{2} Id_{Sym, V}$$

Proof. Observe that

$$M_{fact}(H_{Id_V}) - \frac{\lambda_V}{2} Id_{Sym, V} + \sum_{\sigma, \sigma' \in \mathcal{L}_V: U_{\sigma} \neq V, U_{\sigma'} \neq V} H_{Id_V}(\sigma, \sigma') M_{\sigma} M_{\sigma'} = M_{fact}(H''_{Id_V}) \geq 0$$

We now analyze the terms $\sum_{\sigma, \sigma' \in \mathcal{L}_V: U_{\sigma} \neq V, U_{\sigma'} \neq V} H_{Id_V}(\sigma, \sigma') M_{\sigma} M_{\sigma'}$.

Definition 10.7. Given $U, V \in \mathcal{I}$ with $w(U) > w(V)$, we define $W(U, V)$ to be

$$W(U, V) = \frac{1}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_U: U_{\sigma} = U} \sum_{\sigma' \in \mathcal{L}_V: U_{\sigma'} \neq V} B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_V}(\sigma, \sigma')$$

Lemma 10.8. For all $V \in \mathcal{I}_{mid}$,

$$\sum_{\sigma, \sigma' \in \mathcal{L}_V: U_{\sigma} \neq V, U_{\sigma'} \neq V} H_{Id_V}(\sigma, \sigma') M_{\sigma} M_{\sigma'} \succeq \sum_{U \in \mathcal{I}_{mid}: w(U) > w(V)} W(U, V) Id_{Sym, U}$$

Proof. Observe that for all $\sigma, \sigma' \in \mathcal{L}_V$ such that $U_{\sigma} \neq V$ and $U_{\sigma'} \neq V$, $||M_{\sigma} M_{\sigma'}|| \leq B_{norm}(\sigma)B_{norm}(\sigma')$ and thus

$$\frac{1}{2} (M_{\sigma} M_{\sigma'} + M_{\sigma'} M_{\sigma}) \succeq \frac{1}{2} B_{norm}(\sigma)B_{norm}(\sigma') (M_{Id_{U_{\sigma}}} + M_{Id_{U_{\sigma'}}})$$

Summing this equation over all $\sigma, \sigma' \in \mathcal{L}_V$ such that $U_{\sigma} \neq V$ and $U_{\sigma'} \neq V$,

$$\sum_{\sigma, \sigma' \in \mathcal{L}_V: U_{\sigma} \neq V, U_{\sigma'} \neq V} H_{Id_V}(\sigma, \sigma') M_{\sigma} M_{\sigma'} \succeq \sum_{\sigma, \sigma' \in \mathcal{L}_V: U_{\sigma} \neq V, U_{\sigma'} \neq V} B_{norm}(\sigma)B_{norm}(\sigma') M_{Id_{U_{\sigma}}}$$

$$\succeq \sum_{U \in \mathcal{I}_{mid}: w(U) > w(V)} \sum_{\sigma \in \mathcal{L}_U: U_{\sigma} = U} \sum_{\sigma' \in \mathcal{L}_V: U_{\sigma'} \neq V} B_{norm}(\sigma)B_{norm}(\sigma') H_{Id_V}(\sigma, \sigma') M_{Id_{U_{\sigma}}}

\succeq \sum_{U \in \mathcal{I}_{mid}: w(U) > w(V)} |Aut(U)|W(U, V) M_{Id_{U}}$$

Since all of the coefficient matrices have SOS-symmetry, we can replace $M_{Id_{U}}$ by $\frac{1}{|Aut(U)|} Id_{Sym, U}$ and this completes the proof.
Using this lemma, we can show the following theorem:

**Theorem 10.9.** Let $G$ be the following directed graph:

1. The vertices of $G$ are the index shapes $V \in \mathcal{I}_{\text{mid}}$
2. For each $U, V \in \mathcal{I}_{\text{mid}}$ such that $w(U) > w(V)$, we have an edge $e = (V, U)$ with weight $w(e) = \frac{2w(U,V)}{\lambda_V}$

For all $V \in \mathcal{I}_{\text{mid}}$.

$$Id_{\text{Sym},V} \leq 2 \sum_{U \in \mathcal{I}_{\text{mid}}: w(U) \geq w(V)} \left( \sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \right) \frac{1}{\lambda_U} M^{\text{fact}}(H_{Id_{\text{Sym}}})$$

**Proof sketch.** This can be proved by starting with Corollary 10.6 and iteratively applying Lemma 10.8 and Corollary 10.6.

### 10.1.1 Handling Non-multilinear Matrix Indices

In order to handle non-multilinear matrix indices, we need to make a few adjustments. First, we need to modify the definition of $I_d_{\text{Sym},V}$.

**Definition 10.10.** For all $V \in \mathcal{I}_{\text{mid}}$ we define $I_d_{\text{Sym},V}$ to be the matrix such that

1. $I_d_{\text{Sym},V}(A, B) = 1$ if $A$ and $B$ have the same index shape $U$ and $U$ has the same number of each type of vertex as $V$. Note that $B$ may be a permutation of $A$ and $U$ may have different powers than $V$.
2. Otherwise, $I_d_{\text{Sym},V}(A, B) = 0$.

Observe that with this modified definition, we will still have $Id_{\text{Sym}} = \sum_{V \in \mathcal{I}_{\text{mid}}} Id_{\text{Sym},V}$.

We also need to adjust how we define $\lambda_V$ as there are left shapes $\sigma$ such that $U_\sigma$ and $V_\sigma$ have the same numbers and types of vertices but $U_\sigma$ has different powers.

**Definition 10.11.** Given $V \in \mathcal{I}_{\text{mid}}$, we define $\mathcal{T}_V \subseteq L_V$ to be the set of left shapes $\sigma \in L_V$ such that $U_\sigma$ has the same numbers and types of vertices as $V$ (which automatically implies that $E(\sigma) = \emptyset$).

**Definition 10.12.** Define $I_{d,\text{Sym},V}$ to be the matrix indexed by left shapes $\sigma, \sigma' \in \mathcal{T}_V$ such that $I_{d,\text{Sym},V}(\sigma, \sigma') = \frac{1}{|\text{Aut}(V)|}$ if $U_\sigma \equiv U_{\sigma'}$ and $I_{d,\text{Sym},V}(\sigma, \sigma') = 0$ otherwise.

**Proposition 10.13.** $M^{\text{fact}}(I_{d,\text{Sym},V}) = Id_{\text{Sym},V}$

**Definition 10.14.** Let $H'_{Id_{\text{Sym}}}$ be the matrix $H_{Id_{\text{Sym}}}$ restricted to rows and columns $\sigma, \sigma'$ where $\sigma, \sigma' \in \mathcal{T}_V$. We define $\lambda_V$ to be the largest constant $\lambda$ such that $H'_{Id_{\text{Sym}}} \succeq \lambda I_{d,\text{Sym},V}$.

Finally, whenever we have the condition that $U_\sigma \neq V$, it should instead be the condition that $U_\sigma$ does not have the same number of each type of vertex as $V$. With these adjustments, the same arguments go through.
10.2 Strategy to prove positivity in our applications

Now, we will illustrate the final ingredients needed to show positivity for our applications.

To use Theorem 7.101, we would need to prove a statement of the form: Whenever $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}$,

$$\sum_{U \in \mathcal{I}_{\text{mid}}} M_{Id(U)}^\text{fact}(H_{Id(U)}) \geq 6 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,\sigma}} \frac{d_{Id(U)}(H_{\gamma}, H_{Id(U)})}{|\text{Aut}(U)|} |c(\gamma)| \right) \text{Id}_{\text{sym}}$$

We will sketch the strategy we use to prove this. Let $D_{\text{sos}}$ be the degree of the SoS program.

For the left hand side, we will prove a lower bound of the form: Whenever $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}$,

$$\sum_{U \in \mathcal{I}_{\text{mid}}} M_{Id(U)}^\text{fact}(H_{Id(U)}) \geq \frac{1}{n^{K_1 D_{\text{sos}}^{2}}} \text{Id}_{\text{sym}}$$

for a constant $K_1 > 0$. For this, we use the strategy from Section 10.1. Then, we prove an upper bound on the right hand side of the form

$$\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,\sigma}} \frac{d_{Id(U)}(H_{\gamma}, H_{Id(U)})}{|\text{Aut}(U)|} |c(\gamma)| \leq \frac{n^{K_2 D_{\text{sos}}}}{2^{D_V}}$$

for a constant $K_2 > 0$.

Now, we put these two together. Using the fact that $\text{Id}_{\text{sym}} \succeq 0$, by simply setting $\frac{1}{n^{K_1 D_{\text{sos}}^{2}}} > \frac{n^{K_2 D_{\text{sos}}}}{2^{D_V}}$ which can be obtained by choosing $D_{\text{sos}}$ small enough, we obtain the desired result.

We will also need the following bound that says that that lets us sum over all shapes if we have sufficient decay for each vertex, then, the sum of this decay, over all shapes $\sigma \circ \sigma'$ for $\sigma, \sigma' \in \mathcal{L}'_U$, is bounded.

**Definition 10.15.** For $U \in \mathcal{I}_{\text{mid}}$, let $\mathcal{L}'_U \subset \mathcal{L}_U$ be the set of non-trivial shapes in $\mathcal{L}_U$.

**Lemma 10.16.** Suppose $D_V = n^{C_V}, D_E = n^{C_E}$ for constants $C_V, C_E > 0$, are the truncation parameters for our shapes. For any $U \in \mathcal{I}_{\text{mid}},$

$$\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{1}{D_{\text{sos}}^{D_{\text{sos}} H_{\text{Fe}[V(\sigma \circ \sigma')]}^1}} \leq 1$$

for a constant $F > 0$ that depends only on $C_V, C_E$. In particular, by setting $C_V, C_E$ small enough, we can make this constant arbitrarily small.

**Proof.** For a given $j = |U|$, the number of ways to choose $U$ is at most $t_{\text{max}}^j$. For a given $U \in \mathcal{I}_{\text{mid}}$, we will bound the number of ways to choose $\sigma, \sigma' \in \mathcal{L}'_U$. To choose $\sigma, \sigma' \in \mathcal{L}'_U$, it is sufficient to choose

- The number of vertices $j_1 \geq 1$ (resp. $j_1' \geq 1$) in $U_\sigma \setminus V_\sigma$ (resp. $U_{\sigma'} \setminus V_{\sigma'}$), their types of which there are at most $t_{\text{max}}$, and their powers which have at most $D_{\text{sos}}$ choices.
- The number of vertices $j_2$ (resp. $j_2'$) in $V(\sigma) \setminus (U_\sigma \cup V_\sigma)$ (resp. $V(\sigma') \setminus (U_{\sigma'} \cup V_{\sigma'})$) and also their types, of which there are at most $t_{\text{max}}$.
- The position of each vertex $i$ in $U_\sigma \setminus V_\sigma$ (resp. $U_{\sigma'} \setminus V_{\sigma'}$) within $U_\sigma$ (resp. $U_{\sigma'}$). There are at most $D_V$ choices for each vertex.
- The subset of \( U_\sigma \) (resp. \( U_{\sigma'} \)) that is in \( V_\sigma \) (resp. \( V_{\sigma'} \)) and a mapping in \( Aut(U_\sigma \cap V_\sigma) \) (resp. \( Aut(U_{\sigma'} \cap V_{\sigma'}) \)) that determines the matching between the vertices in \( U_\sigma \cap V_\sigma \) (resp. \( U_{\sigma'} \cap V_{\sigma'} \)).
- The number \( j_3 \) (resp. \( j'_3 \)) of edges in \( E(\sigma) \) (resp. \( E(\sigma') \)). and the \( k \) endpoints of each edge. Each endpoint has at most \( D_V \) choices.

Therefore, for all \( j \geq 0, j_1, j'_1 \geq 1, j_2, j'_2, j_3, j'_3 \geq 0 \), we have

\[
\sum_{U \in I_{mid}} \sum_{\sigma, \sigma' \in I_U} \frac{1}{|Aut(U_\sigma \cap V_\sigma)||Aut(U_{\sigma'} \cap V_{\sigma'})|(2t_{max})^{j_1+2j'_1}(D_V t_{max} D_{sos})^{j_2+2j'_2}(D_V)^{j_3+j'_3}} \leq 1
\]

This implies that

\[
\sum_{U \in I_{mid}} \sum_{\sigma, \sigma' \in I_U} \frac{1}{D_{sos}^{2D_{Fe}V(\sigma \sigma')}} \leq 1
\]

for a constant \( F > 0 \) that only depends on \( C_V, C_E \).

### 11 Planted slightly denser subgraph: Full verification

In this section, we will prove all the required bounds to prove Theorem 1.1.

**Theorem 1.1.** Let \( C_p > 0 \). There exists a constant \( C > 0 \) such that for all sufficiently small constants \( \epsilon > 0 \), if \( k \leq n^{1-\epsilon} \) and \( p = \frac{1}{2} + \frac{n^{-C\mu}}{2} \), then with high probability, the candidate moment matrix \( \Lambda \) given by pseudo-calibration for degree \( n^{\epsilon} \) Sum-of-Squares is PSD.

In particular, we will use Theorem 7.95 where we choose \( \epsilon \) in the theorem, not to be confused with the \( \epsilon \) in Theorem 1.1, to be an arbitrarily small constant.

**Lemma 11.1.** For all \( U \in I_{mid} \) and \( \tau \in M_U \),

\[
\frac{1}{|Aut(U)||\mathit{c}(\tau)|} H_{id_{U,\tau}} \geq 0
\]

**Lemma 11.2.** For all \( U, V \in I_{mid} \) where \( w(U) > w(V) \) and all \( \gamma \in \Gamma_{U,V} \),

\[
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 \nabla H_{id_{U,\gamma}} \leq H_{id_{U,\gamma}}
\]

**Lemma 11.3.** Whenever \( \|M_\alpha\| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in M' \),

\[
\sum_{U \in I_{mid}} M_{id_{U,\gamma}}^\text{fact}(H_{id_{U,\gamma}}) \geq 6 \left( \sum_{U \in I_{mid}} \sum_{\gamma \in \Gamma_{U,\gamma}} \frac{d_{id_{U,\gamma}}(H_{id_{U,\gamma}}, H_{id_{U,\gamma}})}{|Aut(U)||\mathit{c}(\gamma)|} \right) Id_{sym}
\]

**Corollary 11.4.** With constant probability, \( \Lambda \geq 0 \).

**Proof.** This follows by invoking Theorem 7.95 whose conditions follow from Lemma 4.7, Lemma 11.1, Lemma 11.2 and Lemma 11.3. ■
11.1 Proof of Lemma 11.1

Lemma 11.5. Suppose \( k \leq n^{1/2-\varepsilon} \). For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \),

\[
\sqrt{n}|V(\tau)|\frac{|U|}{n}|S(\tau) - 1| 
\leq \frac{1}{n^{C_{p\varepsilon}|E(\tau)|}}
\]

Proof. This result follows by plugging in the value of \( S(\tau) \). Using \( k \leq n^{1/2-\varepsilon} \),

\[
\sqrt{n}|V(\tau)|\frac{|U|}{n}|S(\tau) = \sqrt{n}|V(\tau)|\frac{|U|}{n} \left( k \frac{|V(\tau)|}{|U|} \right) (2\frac{1}{2} + \frac{1}{2n^{C_{p\varepsilon}}}) - 1)^{|E(\tau)|}
\]

\[
\leq \frac{1}{n^{C_{p\varepsilon}|E(\tau)|}}
\]

Corollary 11.6. For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), we have

\[
c(\tau)B_{norm}(\tau)S(\tau) \leq 1
\]

Proof. Since \( \tau \) is a proper middle shape, we have \( w(I_\tau) = 0 \) and \( w(S_\tau) = w(U_\tau) \). This implies

\[
n^{-\frac{w(V(\tau)) + w(U_\tau) - w(S_\tau)}{2}} = \sqrt{n}|V(\tau)|\frac{|U|}{n}
\]

Since \( \tau \) is proper, every vertex \( i \in V(\tau) \setminus U_\tau \) or \( i \in V(\tau) \setminus V_\tau \) has \( \deg^\tau(i) \geq 1 \) and hence, \( |V(\tau) \setminus U_\tau| + |V(\tau) \setminus V_\tau| \leq 4|E(\tau)| \). Also, \( q = n^{O(1)\cdot C_{V}} \). We can set \( C_{V} \) sufficiently small so that, using Lemma 11.5,

\[
c(\tau)B_{norm}(\tau)S(\tau) = 100(3DV_{\frac{|U_\tau|}{|V(\tau)|}} + |V(\tau)|\setminus V_\tau| + 2|E(\tau)|\setminus |U_\tau\cup V_\tau|)
\]

\[
\cdot (6DV_{\frac{\sqrt{2q}}{2q}})|V(\tau)|\setminus U_\tau| + |V(\tau)|\setminus V_\tau| \sqrt{n}|V(\tau)|\frac{|U|}{n}|S(\tau)
\]

\[
\leq n^{O(1)\cdot C_{V}\cdot|E(\tau)|}\cdot \sqrt{n}|V(\tau)|\frac{|U|}{n}|S(\tau)
\]

\[
\leq n^{O(1)\cdot C_{V}\cdot|E(\tau)|}\cdot \frac{1}{n^{C_{p\varepsilon}|E(\tau)|}}
\]

\[
\leq 1
\]

We can now prove Lemma 11.1.

Lemma 11.1. For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \),

\[
\begin{bmatrix}
\frac{1}{|Aut(U)|}c(\tau)^T H_{IdU} & B_{norm}(\tau)H_\tau
\end{bmatrix}
\begin{bmatrix}
\frac{1}{|Aut(U)|}c(\tau) H_{IdU}^T & B_{norm}(\tau)H_\tau
\end{bmatrix} \succeq 0
\]

94
Proof. We have

\[
\begin{bmatrix}
\frac{1}{n	ext{Aut}(U)c(\tau)}H_{id_U} & B_{\text{norm}}(\tau)H_\tau \\
B_{\text{norm}}(\tau)H_\tau^T & \frac{1}{n\text{Aut}(U)c(\tau)}H_{id_U}
\end{bmatrix}
= \begin{bmatrix}
\left(\frac{1}{|\text{Aut}(U)|c(\tau)} - \frac{S(\tau)B_{\text{norm}}(\tau)}{|\text{Aut}(U)|}\right)H_{id_U} & 0 \\
0 & \left(\frac{1}{|\text{Aut}(U)|c(\tau)} - \frac{S(\tau)B_{\text{norm}}(\tau)}{|\text{Aut}(U)|}\right)H_{id_U}
\end{bmatrix}
\]

+ \begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|}H_{id_U} & H_\tau \\
H_\tau^T & \frac{S(\tau)}{|\text{Aut}(U)|}H_{id_U}
\end{bmatrix}
\]

By Lemma 4.9, \( \begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|}H_{id_U} & H_\tau \\
H_\tau^T & \frac{S(\tau)}{|\text{Aut}(U)|}H_{id_U}
\end{bmatrix} \) \geq 0, so the second term above is positive semidefinite. For the first term, by Lemma 4.7, \( H_{id_U} \geq 0 \) and by Corollary 11.6, \( \frac{1}{|\text{Aut}(U)|c(\tau)} - \frac{S(\tau)B_{\text{norm}}(\tau)}{|\text{Aut}(U)|} \geq 0 \), which proves that the first term is also positive semidefinite.

### 11.2 Proof of Lemma 11.2

**Lemma 11.7.** Suppose \( k \leq n^{1/2-\epsilon} \). For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \) and for all \( \gamma \in \Gamma_{U,V} \),

\[
n^{w(V(\gamma)\setminus U_\gamma)}S(\gamma)^2 \leq \frac{1}{n^{Bc(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+|E(\gamma)|)}}
\]

for some constant \( B \) that depends only on \( C_p \). In particular, it is independent of \( C_V \).

**Proof.** Since \( \gamma \) is a left shape, we have \( |U_\gamma| \geq |V_\gamma| \) as \( V_\gamma \) is the unique minimum vertex separator of \( \gamma \) and so, \( n^{w(V(\gamma)\setminus U_\gamma)} = n^{|V(\gamma)| - |U_\gamma|} \leq n^{|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}} \). Also, note that \( 2|V(\gamma)| - |U_\gamma| - |V_\gamma| = |U_\gamma \setminus V_\gamma| + |V_\gamma \setminus U_\gamma| + 2|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| \geq |V(\gamma) \setminus (U_\gamma \cap V_\gamma)| \). Therefore,

\[
n^{w(V(\gamma)\setminus U_\gamma)}S(\gamma)^2 = n^{|V(\gamma)\setminus U_\gamma|} \left( \frac{k}{n} \right)^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} \left( 2\left( \frac{1}{2} + \frac{1}{2n^{c_\epsilon}} \right) - 1 \right)^{2|E(\gamma)|}
\]

\[
\leq n^{|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}} \left( \frac{1}{n^{1/2+\epsilon}} \right)^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} \left( \frac{1}{n^{2c_\epsilon}} \right)^{|E(\gamma)|}
\]

\[
\leq \left( \frac{1}{n^\epsilon} \right)^{2|V(\gamma)| - |U_\gamma| - |V_\gamma|} \left( \frac{1}{n^{2c_\epsilon}} \right)^{|E(\gamma)|}
\]

\[
\leq \frac{1}{n^{Bc(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{\epsilon E(\gamma)\setminus U_\gamma\cap V_\gamma)}}}
\]

for a constant \( B \) that depends only on \( C_p \).

We can now prove Lemma 11.2.

**Lemma 11.2.** For all \( U, V \in \mathcal{I}_{mid} \) where \( w(U) > w(V) \) and all \( \gamma \in \Gamma_{U,V} \),

\[
c(\gamma)^2N(\gamma)^2B(\gamma)^2H_{id_U}^{-\gamma,\gamma} \preceq H'_\gamma
\]

95
Proof. By Lemma 4.11, we have

\[ c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{id_{0}}^{-\gamma} \leq c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H_{\gamma}' \]

Using the same proof as in Lemma 4.7, we can see that \( H_{\gamma}' \geq 0 \). Therefore, it suffices to prove that

\[ c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 1 \]

Since \( U, V \in \mathcal{I}_{mid} \), \( |Aut(U)| = |U|! \), \( |Aut(V)| = |V|! \). Therefore, \( \frac{|Aut(U)|}{|Aut(V)|} = \frac{|U|!}{|V|!} \leq D_{V}^{\mid \mid U \mid \mid \mid V \mid \mid} \). Also, \( q = n^{O(1)} \cdot C_{V} \). Let \( B \) be the constant from Lemma 11.7. We can set \( C_{V} \) sufficiently small so that, using Lemma 11.7,

\[
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 100^2 (3D_{V})^2 |U| |V| + 2 |V(\gamma)\setminus U(\gamma)| + 4 |E(\alpha)| |V(\gamma)\setminus (U(\gamma)\setminus V(\gamma))| \\
\cdot (3D_{V})^4 |V(\gamma)\setminus U(\gamma)| \langle 6D_{V} \sqrt{2eq} \rangle^2 |V(\gamma)\setminus U(\gamma)| + 2 |V(\gamma)\setminus V(\gamma)| \\
\cdot n^{w(V(\gamma)\setminus U(\gamma))} S(\gamma)^2 \cdot D_{V}^{\mid \mid U \mid \mid \mid V \mid \mid} \\
\leq n^{O(1) \cdot C_{V}} \cdot (\mid V(\gamma)\setminus (U(\gamma)\setminus V(\gamma)) \mid + \Sigma_{\in E(\gamma)} l_{e}) \cdot n^{w(V(\gamma)\setminus U(\gamma))} S(\gamma)^2 \\
\leq \frac{1}{n^{B_{c}(\mid V(\gamma)\setminus (U(\gamma)\setminus V(\gamma)) \mid + \Sigma_{\in E(\gamma)} l_{e})}} \\
\leq 1
\]

11.3 Proof of Lemma 11.3

In this section, we will prove Lemma 11.3.

Lemma 11.3. Whenever \( \| M_{\alpha} \| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),

\[
\sum_{U \in \mathcal{I}_{mid}} M_{id_{0}}^{\text{fact}} (H_{id_{0}}) \geq 6 \left( \sum_{U \in \mathcal{I}_{mid}} \sum_{T \subseteq U, \gamma \in \Gamma_{U, T}} d_{id_{0}} (H_{\gamma}', H_{id_{0}}) \right) \text{Id}_{\text{sym}}
\]

We use the strategy from Section 10. We will prove the following lemmas.

Lemma 11.8. Whenever \( \| M_{\alpha} \| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),

\[
\sum_{U \in \mathcal{I}_{mid}} M_{id_{0}}^{\text{fact}} (H_{id_{0}}) \geq \frac{1}{n^{K_{1}D_{\text{sym}}}} \text{Id}_{\text{sym}}
\]

for a constant \( K_{1} > 0 \).

Lemma 11.9.

\[
\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U, T}} d_{id_{0}} (H_{id_{0}}, H_{\gamma}') \leq n^{K_{2}D_{\text{sym}}} \frac{2D_{V}}{2D_{V}}
\]

for a constant \( K_{2} > 0 \).
If we assume these, we can conclude the following.

**Lemma 11.3.** Whenever $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}'$, 

$$
\sum_{U \in I_{\text{mid}}} M_{\text{fact}}^{I_{\text{Id}_U}}(H_{\text{Id}_U}) \geq 6 \left( \sum_{U \in I_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{\text{Id}_U}(H'_{\gamma}, H_{\text{Id}_U})}{|\text{Aut}(U)\,| c(\gamma)} \right) \text{Id}_{\text{sym}}
$$

**Proof.** Let $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}'$. By Lemma 11.8,

$$
\sum_{U \in I_{\text{mid}}} M_{\text{fact}}^{I_{\text{Id}_U}}(H_{\text{Id}_U}) \geq \frac{1}{\eta K_1 n D_{\text{sos}}} \text{Id}_{\text{sym}}
$$

for a constant $K_1 > 0$. By Lemma 11.9,

$$
\sum_{U \in I_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{\text{Id}_U}(H_{\text{Id}_U}, H'_{\gamma})}{|\text{Aut}(U)\,| c(\gamma)} \leq \frac{\eta K_2 n D_{\text{sos}}}{2^D V}
$$

for a constant $K_2 > 0$.

We choose $C_{\text{sos}}$ sufficiently small so that $\frac{1}{\eta K_1 n D_{\text{sos}}} \geq 6 \frac{\eta K_2 n D_{\text{sos}}}{2^D V}$ which can be satisfied by setting $C_{\text{sos}} < K_3 C_V$ for a sufficiently small constant $K_3 > 0$. Then, since $\text{Id}_{\text{sym}} \succeq 0$, using Lemma 11.8 and Lemma 11.9,

$$
\sum_{U \in I_{\text{mid}}} M_{\text{fact}}^{I_{\text{Id}_U}}(H_{\text{Id}_U}) \geq \frac{1}{\eta K_1 n D_{\text{sos}}} \text{Id}_{\text{sym}}
$$

$$
\geq 6 \frac{\eta K_2 n D_{\text{sos}}}{2^D V} \text{Id}_{\text{sym}}
$$

$$
\geq 6 \left( \sum_{U \in I_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{\text{Id}_U}(H'_{\gamma}, H_{\text{Id}_U})}{|\text{Aut}(U)\,| c(\gamma)} \right) \text{Id}_{\text{sym}}
$$

The rest of the section is devoted to proving Lemma 11.8 and Lemma 11.9.

In the proofs of both these lemmas, we will need a bound on $B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{\text{Id}_U}(\sigma, \sigma')$ that is obtained below.

**Lemma 11.10.** Suppose $k \leq n^{1/2-\varepsilon}$. For all $U \in I_{\text{mid}}$ and $\sigma, \sigma' \in \mathcal{L}_U$,

$$
B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{\text{Id}_U}(\sigma, \sigma') \leq \frac{1}{n^{0.5|V(\alpha)|}}
$$
Proof. Let $\alpha = \sigma \circ \sigma'$. Observe that $|V(\sigma)| + |V(\sigma')| = |V(\alpha)| + |U|$. By choosing $C_V$ sufficiently small,

$$B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{Id_U}(\sigma, \sigma') = (6D_V \sqrt{2eq})^{\frac{|V(\sigma)| \cup |V(\sigma')|}{2n}} \cdot (6D_V \sqrt{2eq})^{\frac{|V(\sigma')| \cup |V(\sigma')|}{2n}} \cdot \frac{1}{|\text{Aut}(U)|} \left( \frac{k}{n} \right)^{|V(\alpha)|} \left( 2 \left( \frac{1}{2} + \frac{1}{2nC_p \epsilon} \right) - 1 \right)^{|E(\alpha)|} \leq n^{O(1) - \epsilon C_V - |V(\alpha)| - |U|} \left\| V(\sigma') \right\| - |U| \left\| V(\sigma') \right\| - |U| \left( \frac{k}{n} \right)^{|V(\alpha)|} \frac{1}{nC_{pE}|E(\alpha)|} \leq n^{O(1) - \epsilon C_V - |V(\alpha)|} \cdot \frac{1}{n^{1/2 + \epsilon} |V(\alpha)|} \cdot \frac{1}{nC_{pE}|E(\alpha)|} \leq \frac{1}{n^{0.5\epsilon |V(\alpha)|}}$$

11.3.1 Proof of Lemma 11.8

To prove Lemma 11.8, we will use the strategy from Section 10.1. We will also use the notation from that section. We recall that for $U \in \mathcal{I}_{\text{mid}}, L'_U \subset L_U$ was the set of non-trivial shapes in $L_U$.

Proposition 11.11. For $V \in \mathcal{I}_{\text{mid}}, \lambda_V = \left( \frac{k}{n} \right)^{|V|}$.  

Proof. We have $\lambda_V = |\text{Aut}(V)|H_{Id_V}(Id_V, Id_V) = \left( \frac{k}{n} \right)^{|V|}$. ■

Corollary 11.12. $\lambda_V \geq \frac{1}{n^{O(1)D_{\text{hos}}}}$

Lemma 11.13. For any edge $e = (V, U)$ in $G$, we have

$$w(e) \leq \frac{n^{O(1)D_{\text{hos}}}}{n^{0.1\epsilon |U|}}$$

Proof. Let $e = (V, U)$ be an edge in $G$. Then, $w(U) > w(V)$ and $\bar{w}(e) = \frac{2w(U, V)}{\lambda_V}$. Using Lemma 11.10,
we have

\[
2W(U, V) = \frac{2}{|\text{Aut}(U)|} \sum_{\sigma \in \mathcal{L}_U, \mu_U = U} \sum_{\sigma' \in \mathcal{L}_V, \mu_V \neq V} B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{\text{Id}_U}(\sigma, \sigma')
\]

\[
\leq \frac{2}{|\text{Aut}(U)|} \sum_{\sigma \in \mathcal{L}_U, \mu_U = U} \sum_{\sigma' \in \mathcal{L}_V, \mu_V \neq V} \frac{1}{n^{0.5|V(\sigma \circ \sigma')|}}
\]

\[
\leq \sum_{\sigma, \sigma' \in \mathcal{L}_V} \frac{2}{n^{0.5|V(\sigma \circ \sigma')|}} \frac{D_{\text{sos}}^{D_{\text{sos}}}}{D_{\text{sos}}^{D_{\text{sos}}} n^{F_U|V(\sigma \circ \sigma')|}}
\]

\[
\leq \frac{D_{\text{sos}}^{D_{\text{sos}}}}{n^{0.1|U|}} \sum_{\sigma, \sigma' \in \mathcal{L}_V} \frac{1}{D_{\text{sos}}^{D_{\text{sos}}} n^{F_U|V(\sigma \circ \sigma')|}}
\]

\[
\leq \frac{D_{\text{sos}}^{D_{\text{sos}}}}{n^{0.1|U|}} \leq \lambda V n^{O(1)D_{\text{sos}}}
\]

where we set \( C_V \) small enough so that \( 0.4 \geq F \) and invoked Lemma 10.16. This proves the lemma.

\[\blacksquare\]

**Corollary 11.14.** For any \( U, V \in \mathcal{I}_{\text{mid}} \) such that \( w(U) > w(V) \),

\[
\sum_{P : P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \leq n^{O(1)D_{\text{sos}}^2}
\]

**Proof.** The total number of vertices in \( G \) is at most \( D_{\text{sos}} + 1 \) since each \( U \in \mathcal{I}_{\text{mid}} \) has at most \( D_{\text{sos}} \) vertices. Therefore, for any fixed integer \( j \geq 1 \), the number of paths from \( V \) to \( U \) of length \( j \) is at most \((D_{\text{sos}} + 1)^j\). Take any path \( P \) from \( V \) to \( U \). Suppose it has length \( j \geq 1 \). Note that for all edges \( e = (V', U') \in E(P) \), since \(|U'| \geq 1\), we have

\[
w(e) \leq \frac{n^{O(1)D_{\text{sos}}}}{n^{0.1|U'|}} \leq \frac{n^{O(1)D_{\text{sos}}}}{n^{0.1\epsilon}}
\]

So, \( \prod_{e \in E(P)} w(e) \leq \left( \frac{n^{O(1)D_{\text{sos}}}}{n^{0.1\epsilon}} \right)^j \). Therefore, by setting \( C_{\text{sos}} \) small enough,

\[
\sum_{P : P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \leq \sum_{j=1}^{D_{\text{sos}}} \left( \frac{n^{O(1)D_{\text{sos}}}}{n^{0.1\epsilon}} \right)^j 
\]

\[
\leq n^{O(1)D_{\text{sos}}^2}
\]

\[\blacksquare\]

We can now prove Lemma 11.8.

**Lemma 11.8.** Whenever \( \|M_{\alpha}\| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_U}^{\text{fact}}(H_{\text{Id}_U}) \geq \frac{1}{n^{K_1D_{\text{sos}}^2}} \text{Id}_{\text{sym}}
\]

99
for a constant $K_1 > 0$.

Proof. For all $V \in \mathcal{I}_{\text{mid}}$, we have

$$Id_{\text{Sym},V} \leq 2 \sum_{U \in \mathcal{I}_{\text{mid}} : w(U) \geq w(V)} \left( \sum_{P : P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \right) \frac{1}{\lambda_U} M^{\text{fact}}(H_{Id_U})$$

Summing this over all $V \in \mathcal{I}_{\text{mid}}$, we get

$$Id_{\text{Sym}} \leq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{2}{\lambda_U} \left( \sum_{V \in \mathcal{I}_{\text{mid}} : w(U) \geq w(V)} \sum_{P : P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \right) M^{\text{fact}}(H_{Id_U})$$

For any fixed $U \in \mathcal{I}_{\text{mid}}$ the number of $V \in \mathcal{I}_{\text{mid}}$ such that $w(U) \geq w(V)$ is at most $D_{sos} + 1$. Also, $\lambda_U \geq \frac{1}{d^{\Omega(1)D_{sos}}}$ for all $U \in \mathcal{I}_{\text{mid}}$. Therefore,

$$Id_{\text{Sym}} \leq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{2}{\lambda_U} (D_{sos} + 1) n^{O(1)D_{sos}^2} M^{\text{fact}}(H_{Id_U})$$

$$\leq \sum_{U \in \mathcal{I}_{\text{mid}}} n^{O(1)D_{sos}^2} M^{\text{fact}}(H_{Id_U})$$

where we used the fact that for all $U \in \mathcal{I}_{\text{mid}}$, $M^{\text{fact}}(H_{Id_U}) \geq 0$. \hfill \blacksquare

11.3.2 Proof of Lemma 11.9

We restate the lemma for convenience.

Lemma 11.9.

$$\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,V}} d_{Id_U}(H_{Id_U}, H') \frac{1}{|\text{Aut}(U)| c(\gamma)} \leq \frac{n^{K_2D_{sos}}}{2^{DV}}$$

for a constant $K_2 > 0$.

Proof. We have

$$\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,V}} d_{Id_U}(H_{Id_U}, H') \frac{1}{|\text{Aut}(U)| c(\gamma)} = \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,V}} \frac{1}{|\text{Aut}(U)| c(\gamma)} \sum_{\sigma, \sigma' : \sigma \gamma \in E_{U,V} \land |\sigma| \leq DV \land |\sigma'| \leq DV, |\gamma| > DV \land |\sigma' \gamma| > DV} B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{Id_U,V}(\sigma, \sigma')$$

100
The set of \( \sigma, \sigma' \) that could appear in the above sum must necessarily be non-trivial and hence, \( \sigma, \sigma' \in \mathcal{L}_{U,s}' \). Then,

\[
\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_{U,s}} \frac{d_{Id,U}(H_{Id,U}H'_{U})}{|Aut(U)|c(\gamma)} = \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}_{U,s}'} B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{Id,U}(\sigma, \sigma') \sum_{\gamma \in \Gamma_{U,s} : |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V} \frac{1}{|Aut(U)|c(\gamma)}
\]

For \( \sigma \in \mathcal{L}_{U,s}' \), define \( m_c = D_V + 1 - |V(\sigma)| \geq 1 \). This is precisely set so that for all \( \gamma \in \Gamma_{U,s} \), we have \( |V(\sigma \circ \gamma)| > D_V \) if and only if \( |V(\gamma)| \geq |U| + m_c \). So, for \( \sigma, \sigma' \in \mathcal{L}_{U,s}' \), using Lemma 9.31

\[
\sum_{\gamma \in \Gamma_{U,s} : |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V} \frac{1}{|Aut(U)|c(\gamma)} \leq \frac{1}{2^{\min(m_c,m_c')-1}}
\]

Also, for \( \sigma, \sigma' \in \mathcal{L}_{U,s}' \), we have \( |V(\sigma \circ \sigma')| + \min(m_c, m_c') - 1 \geq D_V \). Therefore,

\[
\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_{U,s}} \frac{d_{Id,U}(H_{Id,U}H'_{U})}{|Aut(U)|c(\gamma)} \leq \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}_{U,s}'} B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{Id,U}(\sigma, \sigma') \frac{1}{2^{\min(m_c,m_c')-1}}
\]

where we used Lemma 11.10. Using \( n^{0.5|V(\sigma \circ \sigma')|} \geq n^{0.1|V(\sigma \circ \sigma')|2|V(\sigma \circ \sigma')|} \),

\[
\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_{U,s}} \frac{d_{Id,U}(H_{Id,U}H'_{U})}{|Aut(U)|c(\gamma)} \leq \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}_{U,s}'} \frac{n^{O(1)D_{sos}}}{n^{0.1|V(\sigma \circ \sigma')|2|V(\sigma \circ \sigma')|2^{\min(m_c,m_c')-1}}}
\]

\[
\leq \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}_{U,s}'} \frac{n^{O(1)D_{sos}}}{n^{0.1|V(\sigma \circ \sigma')|2D_V}}
\]

\[
\leq \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}_{U,s}'} \frac{D_{sos}^{D_{sos}+0.1|V(\sigma \circ \sigma')|2D_V}}{n^{O(1)D_{sos}}}
\]

The final step will be to argue that \( \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}_{U,s}'} \frac{1}{D_{sos}^{D_{sos}+0.1|V(\sigma \circ \sigma')|2D_V}} \leq 1 \) which will complete the proof. But this will follow from Lemma 10.16 if we set \( C_V \) small enough.

\[\blacksquare\]

### 12 Tensor PCA: Full verification

In this section, we will prove all the bounds required to prove Theorem 1.3.

**Theorem 1.3.** Let \( k \geq 2 \) be an integer. There exist constants \( C, C_\Delta > 0 \) such that for all sufficiently small constants \( \varepsilon > 0 \), if \( \lambda \leq n^{\frac{1}{2} - \varepsilon} \) and \( \Delta = n^{-C_\Delta} \), then with high probability, the candidate moment matrix \( \Lambda \) given by pseudo-calibration for degree \( n^{C_{\varepsilon}} \) Sum-of-Squares is PSD.
In particular, we will use Theorem 7.101 where we choose \( \varepsilon \) in the theorem, not to be confused with the \( \varepsilon \) in Theorem 1.3, to be an arbitrarily small constant.

**Lemma 12.1.** For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \),
\[
\left[ \frac{1}{|\text{Aut}(U)|c(\tau)} \right]^{\frac{1}{|\text{Id}_U|}} H_{id_U} \left( \frac{B_{\text{norm}}(\tau)}{H_{\text{Id}_U}^{\frac{1}{|\text{Id}_U|}}} \right) \geq 0
\]

**Lemma 12.2.** For all \( U, V \in \mathcal{I}_{mid} \) where \( \omega(U) > \omega(V) \) and all \( \gamma \in \Gamma_{U,V} \),
\[
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{\text{Id}_U}^{-1} \gamma \leq H_{\gamma}'
\]

**Lemma 12.3.** Whenever \( \|M_\alpha\| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),
\[
\sum_{U \in \mathcal{I}_{mid}} M_{id_U}^{\text{fact}}(H_{id_U}) \geq 6 \left( \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,v}} \frac{d_{id_U}(H_{\gamma}', H_{id_U})}{|\text{Aut}(U)|c(\gamma)} \right) \text{Id}_{\text{sym}}
\]

**Corollary 12.4.** With constant probability, \( \Lambda \geq 0 \).

**Proof.** This follows by invoking Theorem 7.101 whose conditions follow from Lemma 5.10, Lemma 12.1, Lemma 12.2 and Lemma 12.3.

12.1 Proof of Lemma 12.1

**Lemma 12.5.** Suppose \( \lambda \leq n^{\frac{k}{4} - \varepsilon} \). For all \( U \in \mathcal{I}_{mid} \) and \( \tau \in \mathcal{M}_U \), suppose \( \text{deg}^\tau(i) \) is even for all \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \), then
\[
\sqrt{n^{|V(\tau)| - |U_\tau|}} S(\tau) \leq \frac{1}{n^{0.5e} \sum_{e \in E(\tau)} k_e}
\]

**Proof.** Firstly, we claim that \( \sum_{e \in E(\tau)} k_e \geq 2(|V(\tau)| - |U_\tau|) \). For any vertex \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \), \( \text{deg}^\tau(i) \) is even and is not 0, hence, \( \text{deg}^\tau(i) \geq 2 \). Any vertex \( i \in U_\tau \setminus V_\tau \) cannot have \( \text{deg}^\tau(i) = 0 \) otherwise \( U_\tau \setminus \{i\} \) is a vertex separator of strictly smaller weight than \( U_\tau \), which is not possible, hence, \( \text{deg}^\tau(i) \geq 1 \). Therefore,
\[
\sum_{e \in E(\tau)} k_e = \sum_{i \in V(\tau)} \text{deg}^\tau(i)
\]
\[
\geq \sum_{i \in V(\tau) \setminus U_\tau \setminus V_\tau} \text{deg}^\tau(i) + \sum_{i \in U_\tau \setminus V_\tau} \text{deg}^\tau(i) + \sum_{i \in V_\tau \setminus U_\tau} \text{deg}^\tau(i)
\]
\[
\geq 2|V(\tau)| - |U_\tau| + |U_\tau| - |V_\tau| + |V_\tau| - |U_\tau| = 2(|V(\tau)| - |U_\tau|)
\]
By choosing $C_\Delta$ sufficiently small, we have

$$\sqrt{n}^{V(\tau) - |U_\tau|} S(\tau) = \sqrt{n}^{V(\tau) - |U_\tau|} \Delta^{V(\tau) - |U_\tau|} \prod_{e \in E(\tau)} \left( \frac{\lambda}{(\Delta n)^{1/2}} \right)^{I_e}$$

$$\leq \sqrt{n}^{V(\tau) - |U_\tau|} \Delta^{V(\tau) - |U_\tau|} \prod_{e \in E(\tau)} n^{\frac{1}{2} - 0.5d_e I_e}$$

$$= \Delta^{V(\tau) - |U_\tau|} \prod_{e \in E(\tau)} n^{-0.5d_e I_e}$$

$$\leq \frac{1}{n^{0.5\sum_{e \in E(\tau)} I_e}}$$

\[\Box\]

**Corollary 12.6.** For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, we have

$$c(\tau) B_{norm}(\tau) S(\tau) \leq 1$$

**Proof.** Since $\tau$ is a proper middle shape, we have $w(I_\tau) = 0$ and $w(S_{\tau,min}) = w(U_\tau)$. This implies

$$n^{w(V(\tau) - w(I_\tau) - w(S_{\tau,min})} = \sqrt{n}^{V(\tau) - |U_\tau|}$$

If $\deg^+(i)$ is odd for any vertex $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, then $S(\tau) = 0$ and the inequality is true. So, assume $\deg^+(i)$ is even for all $i \in V(\tau) \setminus U_\tau \setminus V_\tau$. As was observed in the proof of Lemma 12.5, every vertex $i \in V(\tau) \setminus U_\tau$ or $i \in V(\tau) \setminus V_\tau$ has $\deg^+(i) \geq 1$ and hence, $|V(\tau) \setminus U_\tau| + |V(\tau) \setminus V_\tau| \leq 4 \sum_{e \in E(\tau)} I_e$.

Also, $|E(\tau)| \leq \sum_{e \in E(\tau)} I_e$ and $q = n^{O(1) - \epsilon(C_V + C_E)}$. We can set $C_V, C_E$ sufficiently small so that, using Lemma 12.5,

$$c(\tau) B_{norm}(\tau) S(\tau) = 100(3D_V)^{|U_\tau| + |V_\tau|} \prod_{e \in E(\tau)} (400D_V^2 D_q^2 q)^{I_e} \sqrt{n}^{V(\tau) - |U_\tau|} S(\tau)$$

$$\leq n^{O(1) - \epsilon(C_V + C_E)} \sum_{e \in E(\tau)} I_e \cdot \frac{1}{n^{0.5\sum_{e \in E(\tau)} I_e}} S(\tau)$$

$$\leq 1$$

\[\Box\]

We can now prove Lemma 12.1.

**Lemma 12.1.** For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,

$$\begin{bmatrix}
\frac{1}{|Aut(U)|c(\tau)} H_{IdU} \\
B_{norm}(\tau) H_{\tau}
\end{bmatrix}^T \begin{bmatrix}
\frac{1}{|Aut(U)|c(\tau)} H_{IdU} \\
B_{norm}(\tau) H_{\tau}
\end{bmatrix} \succeq 0$$

103
Proof. We have
\[
\begin{bmatrix}
\frac{1}{\text{Aut}(U)|c(\tau)|} H_{I_d U} & B_{\text{norm}}(\tau) H_T \\
B_{\text{norm}}(\tau) H_T^T & \frac{1}{|\text{Aut}(U)| c(\tau)} H_{I_d U}
\end{bmatrix}
= \begin{bmatrix}
\left(\frac{1}{|\text{Aut}(U)| c(\tau)} - \frac{S(\tau) B_{\text{norm}}(\tau)}{|\text{Aut}(U)|}\right) H_{I_d U} & 0 \\
0 & \left(\frac{1}{|\text{Aut}(U)| c(\tau)} - \frac{S(\tau) B_{\text{norm}}(\tau)}{|\text{Aut}(U)|}\right) H_{I_d U}
\end{bmatrix}
+ B_{\text{norm}}(\tau) \begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|} H_{I_d U} & H_T \\
H_T^T & \frac{S(\tau)}{|\text{Aut}(U)|} H_{I_d U}
\end{bmatrix}
\]

By Lemma 5.12, \( \begin{bmatrix}
\frac{S(\tau)}{|\text{Aut}(U)|} H_{I_d U} & H_T \\
H_T^T & \frac{S(\tau)}{|\text{Aut}(U)|} H_{I_d U}
\end{bmatrix} \succeq 0 \), so the second term above is positive semidefinite.

For the first term, by Lemma 5.10, \( H_{I_d U} \succeq 0 \) and by Corollary 12.6, \( \frac{1}{|\text{Aut}(U)| c(\tau)} - \frac{S(\tau) B_{\text{norm}}(\tau)}{|\text{Aut}(U)|} \succeq 0 \), which proves that the first term is also positive semidefinite.

12.2 Proof of Lemma 12.2

Lemma 12.7. Suppose \( \lambda \leq n^{\frac{1}{2}-\varepsilon} \). For all \( U, V \in \mathcal{I}_{\text{mid}} \) where \( w(U) > w(V) \) and for all \( \gamma \in \Gamma_{U,V} \),

\[
h^{w(\gamma)\setminus U_{\gamma}} S(\gamma)^2 \leq \frac{1}{n^{B_{\text{C}}(U,V_{\gamma}) + \sum_{e \in E(\gamma)} k_e}}
\]

for some constant \( B \) that depends only on \( C_{\Delta} \). In particular, it is independent of \( C_V \) and \( C_E \).

Proof. Suppose there is a vertex \( i \in V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma} \) such that \( \text{deg}^{\gamma}(i) \) is odd, then \( S(\gamma) = 0 \) and the inequality is true. So, assume \( \text{deg}^{\gamma}(i) \) is even for all vertices \( i \in V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma} \).

We first claim that \( k \sum_{e \in E(\gamma)} l_e \geq |V(\gamma) \setminus U_{\gamma}|. \) Since \( \gamma \) is a left shape, all vertices \( i \in V(\gamma) \setminus U_{\gamma} \) have \( \text{deg}^{\gamma}(i) \geq 1 \). In particular, all vertices \( i \in V_{\gamma} \setminus U_{\gamma} \) have \( \text{deg}^{\gamma}(i) \geq 1 \). Moreover, if \( i \in V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma} \), since \( \text{deg}^{\gamma}(i) \) is even, we must have \( \text{deg}^{\gamma}(i) \geq 2 \).

Let \( S' \) be the set of vertices \( i \in U_{\gamma} \setminus V_{\gamma} \) that have \( \text{deg}^{\gamma}(i) \geq 1 \). Then, note that \( |S'| + |U_{\gamma} \cap V_{\gamma}| \geq |V_{\gamma}| \implies |S'| \geq |V_{\gamma} \setminus U_{\gamma}| \) since otherwise \( S' \cup (U_{\gamma} \cap V_{\gamma}) \) will be a vertex separator of \( \gamma \) of weight strictly less than \( V_{\gamma} \), which is not possible. Then,

\[
\sum_{e \in E(\gamma)} k l_e = \sum_{i \in V(\gamma)} \text{deg}^{\gamma}(i)
\geq \sum_{i \in V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma}} \text{deg}^{\gamma}(i) + \sum_{i \in U_{\gamma} \setminus V_{\gamma}} \text{deg}^{\gamma}(i) + \sum_{i \in V_{\gamma} \setminus U_{\gamma}} \text{deg}^{\gamma}(i)
\geq 2|V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma}| + |S'| + |V_{\gamma} \setminus U_{\gamma}|
\geq 2|V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma}| + 2|V_{\gamma} \setminus U_{\gamma}|
= 2|V(\gamma) \setminus U_{\gamma}|
\]

Finally, note that \( 2|V(\gamma)| - |U_{\gamma}| - |V_{\gamma}| = |U_{\gamma} \setminus V_{\gamma}| + |V_{\gamma} \setminus U_{\gamma}| + 2|V(\gamma) \setminus U_{\gamma} \setminus V_{\gamma}| \geq |V(\gamma) \setminus U_{\gamma}| \)
Lemma 12.2. By choosing $C_\Delta$ sufficiently small, we have

$$n^{\omega(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 = n^{V(\gamma) \setminus U_\gamma} \Delta^2 [V(\gamma) \setminus |U_\gamma| \setminus |V_\gamma|] \prod_{e \in E(\gamma)} \left( \frac{\lambda^2}{(\Delta n)^k} \right)^{l_e} \leq n^{V(\gamma) \setminus U_\gamma} \Delta^2 [V(\gamma) \setminus |U_\gamma| \setminus |V_\gamma|] \prod_{e \in E(\gamma)} n^{-\frac{1}{2} + \varepsilon} l_e \leq \Delta^2 [V(\gamma) \setminus |U_\gamma| \setminus |V_\gamma|] \prod_{e \in E(\gamma)} n^{-\varepsilon l_e} \leq \frac{1}{n^{B\varepsilon([V(\gamma) \setminus (U_\gamma \cap V_\gamma)] + \sum_{e \in E(\gamma)} l_e)}}$$

for a constant $B$ that depends only on $C_\Delta$. 

Remark 12.8. In the above bounds, note that there is a decay of $n^{B\varepsilon}$ for each vertex in $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$. One of the main technical reasons for introducing the slack parameter $C_\Delta$ in the planted distribution was to introduce this decay, which is needed in the current machinery.

We can now prove Lemma 12.2.

Lemma 12.2. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{id \gamma}^{-\gamma} \leq H'_{\gamma}$$

Proof. By Lemma 5.14, we have

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{id \gamma}^{-\gamma} \leq c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H'_{\gamma}$$

Using the same proof as in Lemma 5.10, we can see that $H'_{\gamma} \geq 0$. Therefore, it suffices to prove that

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 1$$

Since $U, V \in \mathcal{I}_{mid}$, $|Aut(U)| = |U|!. |Aut(V)| = |V|!$. Therefore, $\frac{|Aut(U)|}{|Aut(V)|} = \frac{|U|!}{|V|!} \leq D_{V}^{[U_\gamma \setminus V_\gamma]}$. Also, $|E(\gamma)| \leq \sum_{e \in E(\gamma)} l_e$ and $q = n^{O(1) \cdot (C_V + C_E)}$. Let $B$ be the constant from Lemma 12.7. We can set $C_V, C_E$ sufficiently small so that, using Lemma 12.7,

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 100^2 (3D_V)^2 |U_\gamma \setminus V_\gamma| + 2 |V_\gamma \setminus U_\gamma| + 2k |E(a)| 4 |V(\gamma) \setminus (U_\gamma \cup V_\gamma)|$$

$$\cdot (3D_V)^{4 |V(\gamma) \setminus V_\gamma| + 2 |V(\gamma) \setminus U_\gamma|} (6q D_V)^2 |V(\gamma) \setminus U_\gamma| + 2 |V(\gamma) \setminus V_\gamma| \prod_{e \in E(\gamma)} (400D_V^2 D_E^2 q)^{2l_e}$$

$$\cdot n^{\omega(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 \cdot D_{V}^{[U_\gamma \setminus V_\gamma]} \leq n^{O(1) \cdot (C_V + C_E) \cdot (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)} \cdot n^{\omega(V(\gamma) \setminus U_\gamma)} S(\gamma)^2$$

$$\leq n^{O(1) \cdot (C_V + C_E) \cdot (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)} \cdot \frac{1}{n^{B\varepsilon([V(\gamma) \setminus (U_\gamma \cap V_\gamma)] + \sum_{e \in E(\gamma)} l_e)}} \leq 1$$

105
12.3 Proof of Lemma 12.3

In this section, we will prove Lemma 12.3 using the strategy sketched in Section 10.

**Lemma 12.3.** Whenever \( \|M_\alpha\| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_U}^{\text{fact}} (H_{\text{Id}_U}) \geq 6 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{Id}_U}(H_{\gamma}', H_{\text{Id}_U})}{|\text{Aut}(U)| c(\gamma)} \right) \text{Id}_{\text{sym}}
\]

In particular, we prove the following lemmas.

**Lemma 12.9.** Whenever \( \|M_\alpha\| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_U}^{\text{fact}} (H_{\text{Id}_U}) \geq \frac{\Delta^{2D_{\text{sos}}}}{n^{D_{\text{sos}}} \text{Id}_{\text{sym}}}
\]

**Lemma 12.10.**

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{Id}_U}(H_{\gamma}', H_{\text{Id}_U})}{|\text{Aut}(U)| c(\gamma)} \leq \frac{1}{\Delta^{2D_{\text{sos}} 2D_V}}
\]

Assuming these, we can conclude the following.

**Lemma 12.3.** Whenever \( \|M_\alpha\| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \),

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_U}^{\text{fact}} (H_{\text{Id}_U}) \geq 6 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{Id}_U}(H_{\gamma}', H_{\text{Id}_U})}{|\text{Aut}(U)| c(\gamma)} \right) \text{Id}_{\text{sym}}
\]

**Proof.** We choose \( C_{\text{sos}} \) sufficiently small so that \( \frac{\Delta^{2D_{\text{sos}}}}{n^{D_{\text{sos}}}} \geq \frac{6}{\Delta^{2D_{\text{sos}} 2D_V}} \) which is satisfied by setting \( C_{\text{sos}} < 0.5 C_V \). Then, since \( \text{Id}_{\text{sym}} \geq 0 \), using Lemma 12.9 and Lemma 12.10,

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{Id}_U}^{\text{fact}} (H_{\text{Id}_U}) \geq \frac{\Delta^{2D_{\text{sos}}}}{n^{D_{\text{sos}}} \text{Id}_{\text{sym}}} \geq \frac{6}{\Delta^{2D_{\text{sos}} 2D_V}} \text{Id}_{\text{sym}} \geq 6 \left( \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{Id}_U}(H_{\gamma}', H_{\text{Id}_U})}{|\text{Aut}(U)| c(\gamma)} \right) \text{Id}_{\text{sym}}
\]

The rest of the section is devoted to proving Lemma 12.9 and Lemma 12.10.

In the proofs of both these lemmas, we will need a bound on \( B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{\text{Id}_U}(\sigma, \sigma') \) that is obtained below.

**Lemma 12.11.** Suppose \( \lambda = n^{\frac{k}{4} - \epsilon} \). For all \( U \in \mathcal{I}_{\text{mid}} \) and \( \sigma, \sigma' \in \mathcal{L}_U \),

\[
B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{\text{Id}_U}(\sigma, \sigma') \leq \frac{1}{n^{0.5 \epsilon \mathcal{C}_A |V(\sigma \circ \sigma')| \Delta^{D_{\text{sos}} \eta |U|}}
\]
Proof. Suppose there is a vertex \( i \in V(\sigma) \setminus V_\sigma \) such that \( \deg^\sigma(i) + \deg^{\sigma'}(i) \) is odd, then \( H_{I_{\mid U}}(\sigma, \sigma') = 0 \) and the inequality is true. So, assume that \( \deg^\sigma(i) + \deg^{\sigma'}(i) \) is even for all \( i \in V(\sigma) \setminus V_\sigma \). Similarly, assume that \( \deg^\sigma(i) + \deg^{\sigma'}(i) \) is even for all \( i \in V(\sigma') \setminus V_\sigma' \). Also, if \( \rho_\sigma \neq \rho_{\sigma'} \), we will have \( H_{I_{\mid U}}(\sigma, \sigma') = 0 \) and we’d be done. So, assume \( \rho_\sigma = \rho_{\sigma'} \).

Let \( \alpha = \sigma \circ \sigma' \). We will first prove that \( \sum_{e \in E(\alpha)} kl_e + 2\deg(\alpha) \geq 2|V(\alpha)| + 2|U| \). Firstly, note that all vertices \( i \in V(\alpha) \setminus (U_\alpha \cup V_\alpha) \) have \( \deg^\alpha(i) \) to be even and nonzero, and hence at least 2. Moreover, in both the sets \( U_\alpha \setminus (U_\alpha \cap V_\alpha) \) and \( V_\alpha \setminus (U_\alpha \cap V_\alpha) \), there are at least \( |U| - |U_\alpha \cap V_\alpha| \) vertices of degree at least 1, because \( U \) is a minimum vertex separator. Also, note that \( \deg(\alpha) \geq |U_\alpha| + |V_\alpha| \). This implies that

\[
\sum_{e \in E(\alpha)} kl_e + 2\deg(\alpha) \geq 2|V(\alpha)| \setminus (U_\alpha \cup V_\alpha) \) + 2(|U| - |U_\alpha \cap V_\alpha|) + 2(|U_\alpha \cap V_\alpha|) + 2(|U_\alpha \cup V_\alpha| + |U_\alpha \cap V_\alpha|)
\]

\[
= 2|V(\alpha)| + 2|U|
\]

where we used the fact that \( U_\alpha \cap V_\alpha \subseteq U \). Finally, by choosing \( C_V, C_E \) sufficiently small,

\[
B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{I_{\mid U}}(\sigma, \sigma') = 2e(6qD_V)|V(\sigma)\setminus U_\sigma|+|V(\sigma)\setminus V_\sigma| \prod_{e \in E(\sigma)} (400D_V^2D_E^2)^{\frac{n^{\deg(\alpha)} - n(\deg(\alpha))}{2}}
\]

\[
:: 2e(6qD_V)|V(\sigma)\setminus U_\sigma|+|V(\sigma)\setminus V_\sigma| \prod_{e \in E(\sigma)} (400D_V^2D_E^2)^{\frac{n^{\deg(\alpha)} - n(\deg(\alpha))}{2}}
\]

\[
\cdot \frac{1}{|\text{Aut}(U)|} \Delta |V_\alpha| \left( \frac{1}{\Delta n} \right)^{\deg(\alpha)} \prod_{e \in E(\alpha)} \left( \frac{\lambda}{\Delta(n)^{\frac{\lambda}{2}}} \right)^{l_e}
\]

\[
\leq n^{O(1)} \frac{1}{C_V + C_E} \cdot n^{\sum_{e \in E(\alpha)} l_e} \Delta |V(\alpha)| \left( \frac{1}{\Delta} \right)^{\deg(\alpha)} \prod_{e \in E(\alpha)} n^{\left( \frac{l_e}{\Delta} - 0.5 \right)}
\]

\[
\leq n^{O(1)} \frac{1}{C_V + C_E} \cdot n^{\sum_{e \in E(\alpha)} l_e} \Delta |V(\alpha)| \left( \frac{1}{\Delta} \right)^{\deg(\alpha)} \prod_{e \in E(\alpha)} n^{\left( \frac{l_e}{\Delta} - 0.5 \right)}
\]

\[
\leq \frac{1}{n^{0.5 \Delta C_V |V(\alpha)| \sum_{e \in E(\alpha)} l_e}} n^{\Delta |V(\alpha)| + |U| - \deg(\alpha) - \sum_{e \in E(\alpha)} kl_e} \cdot \frac{1}{\Delta D_{\text{sos}} |U|} \sqrt{n}|V(\alpha)| + |U| - \deg(\alpha) - \frac{1}{2} \sum_{e \in E(\alpha)} kl_e
\]

where we used the facts \( \Delta \leq 1, \deg(\alpha) \leq 2D_{\text{sos}} \).

12.3.1 Proof of Lemma 12.9

To prove Lemma 12.9, we will use the strategy from Section 10.1. We will also use the notation from that section. We recall that for \( U \in I_{\text{mid}}, L'_U \subseteq L_U \) was the set of non-trivial shapes in \( L_U \).

Proposition 12.12. For \( V \in I_{\text{mid}} \), \( \lambda_V = \frac{1}{n^2} \).

Proof. We have \( \lambda_V = |\text{Aut}(V)|H_{I_{\mid V}}(I_{\mid V}, I_{\mid V}) = \Delta |V| \left( \frac{1}{\sqrt{n}} \right)^{2 |V|} = \frac{1}{n^2} \).
Lemma 12.13. For any edge \( e = (V, U) \) in \( G \), we have

\[
w(e) \leq \frac{1}{n^{0.1C_{\Delta}|U|\Delta 2D_{sos}}}
\]

Proof. Let \( e = (V, U) \) be an edge in \( G \). Then, \( w(U) > w(V) \) and

\[
w(e) = \frac{2W(U, V)}{\lambda_V}.
\]

Using Lemma 12.11, we have

\[
2W(U, V) = \frac{2}{|Aut(U)|} \sum_{\sigma \in L_V, U = \sigma'j \neq V} \sum_{\sigma' \in L_V, U, \sigma' \neq V} B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{\text{Id}_U}(\sigma, \sigma')
\]

\[
\leq \frac{2}{|Aut(U)|} \sum_{\sigma \in L_V, U = \sigma'j \neq V} \sum_{\sigma' \in L_V, U, \sigma' \neq V} \frac{1}{n^{0.5C_{\Delta}|V(\sigma\sigma')|\Delta D_{sos}n|V|}}
\]

\[
\leq \frac{1}{n^{0.1C_{\Delta}|U|\Delta 2D_{sos}}n|V|} \sum_{\sigma', U, \sigma' \neq V} \sum_{\sigma' \in L_V} \frac{1}{n^{2D_{sos}n|V(\sigma\sigma')|}}
\]

\[
\leq \frac{1}{n^{0.1C_{\Delta}|U|\Delta 2D_{sos}}\lambda_V}
\]

where we set \( C_V, C_E \) small enough so that \( 0.4C_{\Delta} \geq F \) and invoked Lemma 10.16. This proves the lemma.

Corollary 12.14. For any \( U, V \in I_{mid} \) such that \( w(U) > w(V) \),

\[
\sum_{P: P \text{ is a path from } V \to U \text{ in } G \in E(P)} w(e) \leq \frac{1}{2D_{sos}}
\]

Proof. The total number of vertices in \( G \) is at most \( D_{sos} + 1 \) since each \( U \in I_{mid} \) has at most \( D_{sos} \) vertices. Therefore, for any fixed integer \( j \geq 1 \), the number of paths from \( V \) to \( U \) of length \( j \) is at most \( (D_{sos} + 1)^j \).

Take any path \( P \) from \( V \) to \( U \). Suppose it has length \( j \geq 1 \). Note that for all edges \( e = (V', U') \) in \( E(P) \), since \( |U'| \geq 1 \), we have

\[
w(e) \leq \frac{1}{n^{0.1C_{\Delta}|U'|\Delta 2D_{sos}}} \leq \frac{1}{n^{0.1C_{\Delta}\Delta 2D_{sos}}}
\]

So, \( \prod_{e \in E(P)} w(e) \leq \left( \frac{1}{n^{0.1C_{\Delta}\Delta 2D_{sos}}} \right)^j \). Therefore, by setting \( C_{sos} \) small enough,

\[
\sum_{P: P \text{ is a path from } V \to U \text{ in } G \in E(P)} \prod_{e \in E(P)} w(e) \leq \sum_{j=1}^{D_{sos}} (D_{sos} + 1)^j \left( \frac{1}{n^{0.1C_{\Delta}\Delta 2D_{sos}}} \right)^j \leq \frac{1}{2D_{sos}\Delta 2D_{sos}}
\]
We can now prove Lemma 12.9.

**Lemma 12.9.** Whenever $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}'$,

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{fact}}^{I_U}(H_{I_U}) \geq \frac{\Delta^{2D_{\text{sos}}}}{H_{\text{sos}}} \text{Id}_{\text{sym}}
\]

**Proof.** For all $V \in \mathcal{I}_{\text{mid}}$, we have

\[
\text{Id}_{\text{Sym},V} \preceq 2 \sum_{U \in \mathcal{I}_{\text{mid}}, w(U) \geq w(V)} \left( \sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in P} w(e) \right) \frac{1}{\lambda_U} M_{\text{fact}}^{I_U}(H_{I_U})
\]

Summing this over all $V \in \mathcal{I}_{\text{mid}}$, we get

\[
\text{Id}_{\text{Sym}} \preceq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{2}{\lambda_U} \left( \sum_{V \in \mathcal{I}_{\text{mid}}, w(U) \geq w(V)} \frac{1}{2D_{\text{sos}} \Delta^{2D_{\text{sos}}}} \right) M_{\text{fact}}^{I_U}(H_{I_U})
\]

For any fixed $U \in \mathcal{I}_{\text{mid}}$, the number of $V \in \mathcal{I}_{\text{mid}}$ such that $w(U) \geq w(V)$ is at most $D_{\text{sos}}$. Therefore,

\[
\text{Id}_{\text{Sym}} \preceq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{1}{\lambda_U \Delta^{2D_{\text{sos}}}} M_{\text{fact}}^{I_U}(H_{I_U}) = \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{1}{\Delta^{2D_{\text{sos}}}} n_{|U|} M_{\text{fact}}^{I_U}(H_{I_U}) \preceq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{1}{\Delta^{2D_{\text{sos}}}} n_{\text{sos}} M_{\text{fact}}^{I_U}(H_{I_U})
\]

where we used the fact that for all $U \in \mathcal{I}_{\text{mid}}$, we have $|U| \leq D_{\text{sos}}$ and $M_{\text{fact}}^{I_U}(H_{I_U}) \geq 0$. \[\blacksquare\]

### 12.3.2 Proof of Lemma 12.10

We restate the lemma for convenience.

**Lemma 12.10.**

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U}} d_{I_{U}}(H_{I_{U}}, H'_{I_{U}}) \leq \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U}} \frac{1}{|\text{Aut}(U)|c(\gamma)^{-1}} \sum_{\sigma, \sigma' \in \mathcal{U}_{\gamma}, |V(\sigma)| \leq D_{V}, |V(\sigma')| \leq D_{V}, |V(\sigma \gamma)| > D_{V}} B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{I_{U}}(\sigma, \sigma')
\]

**Proof.** We have

\[
\sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U}} d_{I_{U}}(H_{I_{U}}, H'_{I_{U}}) \leq \sum_{U \in \mathcal{I}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U}} \frac{1}{|\text{Aut}(U)|c(\gamma)^{-1}} \sum_{\sigma, \sigma' \in \mathcal{U}_{\gamma}, |V(\sigma)| \leq D_{V}, |V(\sigma')| \leq D_{V}, |V(\sigma \gamma)| > D_{V}} B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{I_{U}}(\sigma, \sigma')
\]
The set of $\sigma, \sigma'$ that could appear in the above sum must necessarily be non-trivial and hence, $\sigma, \sigma' \in \mathcal{L}'_{U}$. Then,

$$\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_{U,s}} \frac{d_{Id,U}(H_{Id,U}, H'_{U})}{|Aut(U)| c(\gamma)} = \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_{U}} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id,U}(\sigma, \sigma') \sum_{\gamma \in \Gamma_{U,s}, |V(\sigma \gamma)| > D_{V} or |V(\sigma' \gamma)| > D_{V}} \frac{1}{|Aut(U)| c(\gamma)}$$

For $\sigma \in \mathcal{L}'_{U}$, define $m_\sigma = D_{V} + 1 - |V(\sigma)| \geq 1$. This is precisely set so that for all $\gamma \in \Gamma_{U,s}$, we have $|V(\sigma \gamma)| > D_{V}$ if and only if $|V(\gamma)| \geq |U| + m_\sigma$. So, for $\sigma, \sigma' \in \mathcal{L}'_{U}$, using Lemma 9.31

$$\sum_{\gamma \in \Gamma_{U,s}, |V(\sigma \gamma)| > D_{V} or |V(\sigma' \gamma)| > D_{V}} \frac{1}{|Aut(U)| c(\gamma)} = \sum_{\gamma \in \Gamma_{U,s}, |V(\gamma)| \geq |U| + \min(m_\sigma, m_{\sigma'})} \frac{1}{|Aut(U)| c(\gamma)} \leq \frac{1}{2\min(m_\sigma, m_{\sigma'}) - 1}$$

Also, for $\sigma, \sigma' \in \mathcal{L}'_{U}$, we have $|V(\sigma \circ \sigma')| + \min(m_\sigma, m_{\sigma'}) - 1 \geq D_{V}$. Therefore,

$$\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_{U,s}} \frac{d_{Id,U}(H_{Id,U}, H'_{U})}{|Aut(U)| c(\gamma)} \leq \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_{U}} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id,U}(\sigma, \sigma') \frac{1}{2\min(m_\sigma, m_{\sigma'}) - 1}$$

where we used Lemma 12.11. Using $n^{0.5 C_{A}} |V(\sigma \circ \sigma')| \geq n^{0.1 C_{A}} |V(\sigma \circ \sigma')| 2 |V(\sigma \circ \sigma')|$

$$\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_{U,s}} \frac{d_{Id,U}(H_{Id,U}, H'_{U})}{|Aut(U)| c(\gamma)} \leq \sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_{U}} \frac{n^{0.1 C_{A}} |V(\sigma \circ \sigma')| \Delta_{D_{sos}} N_{U} 2 |V(\sigma \circ \sigma')| 2 \min(m_\sigma, m_{\sigma'}) - 1}{D_{sos} 2 |V(\sigma \circ \sigma')| 2 \min(m_\sigma, m_{\sigma'}) - 1}$$

where we set $C_{sos}$ small enough so that $D_{sos} = n^{C_{sos}} \leq n^{0.1 C_{A}} = \frac{1}{\Delta}$. The final step will be to argue that $\sum_{U \in \mathcal{L}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_{U}} \frac{1}{D_{sos} N_{U} 0.1 C_{A} |V(\sigma \circ \sigma')| 2 D_{sos} 2 D_{V}} \leq 1$ which will complete the proof. But this will follow from Lemma 10.16 if we set $C_V, C_E$ small enough.

### 13 Sparse PCA: Full verification

In this section, we will prove all the bounds required to prove Theorem 1.5.
Theorem 1.5. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$, $m \leq \frac{d^{2-\varepsilon}}{\lambda^x}$, and there exists a constant $A$ such that $0 < A < \frac{1}{4}$, $d^{4A} \leq k \leq d^{1-A\varepsilon}$, and $\frac{\lambda A}{\sqrt{k}} \leq d^{-A\varepsilon}$, then with high probability, the candidate moment matrix $\Lambda$ given by pseudo-calibration for degree $d^{C\varepsilon}$ Sum-of-Squares is PSD.

In particular, we will use Theorem 7.101 where we choose $\varepsilon$ in the theorem, not to be confused with the $\varepsilon$ in Theorem 1.5, to be an arbitrarily small constant.

Definition 13.1. Define $n = \max(d, m)$.

Remark 13.2. The above definition conforms with the notation used in Theorem 7.101. So, we can use the bounds as stated there.

Lemma 13.3. For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,

$$
\left[ \frac{1}{|\text{Aut}(U)|} \text{c}(\tau) H_{idU} \quad \frac{1}{|\text{Aut}(U)|} \text{c}(\tau) H_{idU}^\text{T} \right] \geq 0
$$

Lemma 13.4. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U, V}$,

$$
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{idV}^{-1} \gamma \preceq H_{idV}^\gamma
$$

Lemma 13.5. Whenever $\|M_\alpha\| \leq B_{\text{norm}}(\alpha)$ for all $\alpha \in \mathcal{M}^\prime$,

$$
\sum_{U \in \mathcal{I}_{mid}} M_{\text{idU}}^{\text{act}} (H_{idU}) \geq 6 \left( \sum_{U \in \mathcal{I}_{mid}} \sum_{\tau \in \mathcal{M}_U} \frac{d_{idU}(H_{\gamma}^{\tau}, H_{idU})}{|\text{Aut}(U)| c(\gamma)} \right) I_{\text{sym}}
$$

Corollary 13.6. With high probability, $\Lambda \succeq 0$.

Proof. This follows by invoking Theorem 7.101 whose conditions follow from Lemma 6.11, Lemma 13.3, Lemma 13.4, and Lemma 13.5.

13.1 Proof of Lemma 13.3

Lemma 13.7. Suppose $0 < A < \frac{1}{4}$ is a constant such that $\frac{\lambda A}{\sqrt{k}} \leq d^{-A\varepsilon}$ and $\frac{1}{\sqrt{k}} \leq d^{-2A}$. For all $m$ such that $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$, $m \leq \frac{d^{2-\varepsilon}}{\lambda^x}$, for all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, suppose $\deg(\tau)$ is even for all $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, then

$$
\sqrt{d} |\tau|_1 |U_\tau| \sqrt{m} |\tau|_2 |U_\tau|_2 S(\tau) \leq \prod_{j \in V_{\tau}(\tau) \setminus U_{\tau} \setminus V_{\tau}} (\deg(j) - 1)!! \cdot \frac{1}{d^{A\varepsilon} \sum_{E \in E(\tau)} l_e}
$$

Proof. Let $r_1 = |\tau|_1 - |U_\tau|_1, r_2 = |\tau|_2 - |U_\tau|_2$. Since $\Delta \leq 1$, it suffices to prove

$$
E := \sqrt{d} r_1 \sqrt{m} r_2 \left( \frac{k}{d} \right)^{r_1} \left( \frac{\lambda}{\sqrt{k}} \right)^{\sum_{E \in E(\tau)} l_e} \leq \frac{1}{d^{A\varepsilon} \sum_{E \in E(\tau)} l_e}
$$

We will need the following claim.
Claim 13.8. $\sum_{e \in E(\tau)} l_e \geq 2 \max(r_1, r_2)$.

Proof. We will first prove $\sum_{e \in E(\tau)} l_e \geq 2r_1$. For any vertex $i \in V_1(\tau) \setminus U_\tau \setminus V_\tau$, $\deg^+(i)$ is even and is not 0, hence, $\deg^+(i) \geq 2$. Any vertex $i \in U_\tau \setminus V_\tau$ cannot have $\deg^+(i) = 0$ otherwise $U_\tau \setminus \{i\}$ is a vertex separator of strictly smaller weight than $U_\tau$, which is not possible, hence, $\deg^+(i) \geq 1$. Similarly, for $i \in V_\tau \setminus U_\tau$, $\deg^+(i) \geq 1$. Also, since $H_\tau$ is bipartite, we have $\sum_{i \in V_1(\tau)} \deg^+(i) = \sum_{i \in V_2(\tau)} \deg^+(i) = \sum_{e \in E(\tau)} l_e$. Consider

$$\sum_{e \in E(\tau)} l_e = \sum_{i \in V_1(\tau)} \deg^+(i) \geq \sum_{i \in V_1(\tau) \setminus U_\tau \setminus V_\tau} \deg^+(i) + \sum_{i \in (U_\tau)_1 \setminus V_\tau} \deg^+(i) + \sum_{i \in (V_\tau)_1 \setminus U_\tau} \deg^+(i) \geq 2|V_1(\tau) \setminus U_\tau \setminus V_\tau| + |(U_\tau)_1 \setminus V_\tau| + |(V_\tau)_1 \setminus U_\tau| = 2r_1$$

We can similarly prove $\sum_{e \in E(\tau)} l_e \geq 2r_2$. 

To illustrate the main idea, we will start by proving the weaker bound $E \leq 1$. Observe that our assumptions imply $m \leq \frac{d}{\lambda^2}$, $m \leq \frac{k^2}{\lambda}$ and also, $E \leq \sqrt{d} \sqrt{m^2} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2 \max(r_1, r_2)}$ where we used the fact that $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-Ae} \leq 1$.

Claim 13.9. For integers $r_1, r_2 \geq 0$, if $m \leq \frac{d}{\lambda^2}$ and $m \leq \frac{k^2}{\lambda}$, then,

$$\sqrt{d} \sqrt{m^2} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2 \max(r_1, r_2)} \leq 1$$

Proof. We will consider the cases $r_1 \geq r_2$ and $r_1 < r_2$ separately. If $r_1 \geq r_2$, we have

$$\sqrt{d} \sqrt{m^2} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1} \leq \sqrt{d} \sqrt{m^2} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1} = \left(\frac{\lambda}{\sqrt{d}}\right)^{r_1-r_2} \leq \left(\frac{1}{\sqrt{m}}\right)^{r_1-r_2} \leq 1$$

And if $r_1 < r_2$, we have

$$\sqrt{d} \sqrt{m^2} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} = \sqrt{d} \sqrt{m^2} \sqrt{m^{r_2-r_1}} \sqrt{m^{r_1}} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} \leq \sqrt{d} \sqrt{m^2} \left(\frac{k}{\lambda}\right)^{r_2-r_1} \left(\frac{\sqrt{d}}{\lambda}\right)^{r_1} \left(\frac{k}{\lambda}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} = 1$$

112
For the desired bounds, we mimic this argument while carefully keeping track of factors of $d^e$.

**Claim 13.10.** For integers $r_1, r_2 \geq 0$ and an integer $r \geq 2 \max(r_1, r_2)$, if $m \leq d^{1-\varepsilon}$ and $m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, then,

$$\sqrt{d}^{r_1} \sqrt{m}^{r_2} \left( \frac{k}{d} \right)^{r_1} \left( \frac{\lambda}{\sqrt{k}} \right)^{2r_1} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_1} \left( \frac{1}{d^{Ae}} \right)^r \leq \left( \frac{1}{d^{Ae}} \right)^r$$

**Proof.** If $r_1 \geq r_2$,

$$E = \sqrt{d}^{r_1} \sqrt{m}^{r_2} \left( \frac{k}{d} \right)^{r_1} \left( \frac{\lambda}{\sqrt{k}} \right)^{2r_1} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_1} \left( \frac{1}{d^{Ae}} \right)^r \leq \sqrt{d}^{r_1} \left( \frac{\lambda}{\sqrt{d^{1-\varepsilon}}} \right)^{r_1} \left( \frac{\sqrt{d})}{\lambda} \right)^{2r_1} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_1} \left( \frac{1}{d^{Ae}} \right)^r$$

$$= \left( \frac{\lambda}{\sqrt{d^{1-\varepsilon}}} \right)^{r_1-r_2} \left( \frac{1}{\sqrt{d}} \right)^{r_1} \sqrt{d} \sqrt{m} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_1} \left( \frac{1}{d^{Ae}} \right)^r \leq \left( \frac{\sqrt{m}}{\sqrt{d}} \right)^{r_1} \left( \frac{1}{\sqrt{d}} \right)^{r_1} \left( \frac{1}{d^{2Ae}} \right) \left( \frac{1}{d^{Ae}} \right)^r \leq \left( \frac{1}{d^{Ae}} \right)^r$$

And if $r_1 < r_2$,

$$E = \sqrt{d}^{r_1} \sqrt{m}^{r_2-r_1} \sqrt{m}^{r_1} \left( \frac{k}{d} \right)^{r_1} \left( \frac{\lambda}{\sqrt{k}} \right)^{2r_2} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_2} \left( \frac{1}{d^{Ae}} \right)^r \leq \sqrt{d}^{r_1} \left( \frac{\sqrt{d^{1-\varepsilon}}}{\lambda} \right)^{r_2-r_1} \left( \frac{\sqrt{d^{1-\varepsilon}}}{\lambda} \right)^{r_1} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{2r_2} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_2} \left( \frac{1}{d^{Ae}} \right)^r$$

$$= \left( \frac{\sqrt{d^{1-\varepsilon}}}{\lambda} \right)^{r_2} \left( \frac{1}{\sqrt{d}} \right)^{r_2} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_2} \left( \frac{1}{d^{Ae}} \right)^r \leq \left( \frac{\sqrt{d^{1-\varepsilon}}}{\lambda} \right)^{r_2} \left( \frac{\sqrt{\lambda}}{\sqrt{k}} \right)^{r-2r_2} \left( \frac{1}{d^{Ae}} \right)^r \leq \left( \frac{1}{d^{2Ae}} \right)^{r_2} \left( \frac{1}{d^{Ae}} \right)^r \leq \left( \frac{1}{d^{Ae}} \right)^r \sum_{e \in E(e)} l_e$$

\[113\]
The result follows by setting \( r = \sum_{e \in E(\tau)} l_e \) in the above claim.

**Corollary 13.11.** For all \( U \in \mathcal{I}_{\text{mid}} \) and \( \tau \in \mathcal{M}_U \), we have

\[
c(\tau)B_{\text{norm}}(\tau)S(\tau)R(\tau) \leq 1
\]

**Proof.** First, note that if \( \deg^\tau(i) \) is odd for any vertex \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \), then \( S(\tau) = 0 \) and the inequality is true. So, assume that \( \deg^\tau(i) \) is even for all \( i \in V(\tau) \setminus U_\tau \setminus V_\tau \).

Since \( \tau \) is a proper middle shape, we have \( w(I_\tau) = 0 \) and \( w(S_{\tau,\min}) = w(U_\tau) \). This implies

\[
n \frac{w(V(\tau) + w(I_\tau) - w(S_{\tau,\min})}{\sqrt{d} |V(\tau)| - |U_\tau|} \leq \sqrt{m} \frac{1}{|\tau|_{L_2}} \mid \tau \mid_{L_2} \mid \tau \mid_{L_2}
\]

As was observed in the proof of **Lemma 13.7**, every vertex \( i \in V(\tau) \setminus U_\tau \) or \( i \in V(\tau) \setminus V_\tau \) has \( \deg^\tau(i) \geq 1 \) and hence, \( |V(\tau) \setminus U_\tau| + |V(\tau) \setminus V_\tau| \leq 4 \sum_{e \in E(\tau)} l_e \). Also, \( q = d^{O(1)} \epsilon (C_V + C_E) \). We can set \( C_V, C_E \) sufficiently small so that

\[
c(\tau)B_{\text{norm}}(\tau)S(\tau)R(\tau) = 100(6D_V)^{|U_\tau|} |V_\tau| + |V_\tau| + 2|E(\tau)| |V(\tau)| \setminus (U_\tau \cup V_\tau) |E(\tau) | |V(\tau)| \sum_{\tau \in E(\tau)} \sum_{j \in V(\tau) \setminus V_\tau \setminus U_\tau} \sum_{|V(\tau)| \setminus V_\tau \setminus U_\tau} \deg^\tau(j)
\]

\[
\leq d^{O(1)} \epsilon (C_V + C_E) \sum_{\tau \in E(\tau)} l_e \cdot \prod_{j \in V(\tau) \setminus V_\tau \setminus U_\tau} (deg^\tau(j) - 1)! \cdot \frac{1}{d^{4 \epsilon} \sum_{\tau \in E(\tau)} l_e}
\]

\[
\leq 1
\]

We can now prove **Lemma 13.3**.

**Lemma 13.3.** For all \( U \in \mathcal{I}_{\text{mid}} \) and \( \tau \in \mathcal{M}_U \),

\[
\begin{bmatrix}
\frac{1}{|\text{Aut}(U)\mid c(\tau)} H_{IdU} & B_{\text{norm}}(\tau)H_\tau & \frac{1}{|\text{Aut}(U)\mid c(\tau)} H_{IdU}
\end{bmatrix} \leq 0
\]

**Proof.** We have

\[
\begin{bmatrix}
\frac{1}{|\text{Aut}(U)\mid c(\tau)} H_{IdU} & B_{\text{norm}}(\tau)H_\tau & \frac{1}{|\text{Aut}(U)\mid c(\tau)} H_{IdU}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\left( \frac{1}{|\text{Aut}(U)\mid c(\tau)} - \frac{S(\tau)R(\tau)B_{\text{norm}}(\tau)}{|\text{Aut}(U)|} \right) H_{IdU} & 0 & \left( \frac{1}{|\text{Aut}(U)\mid c(\tau)} - \frac{S(\tau)R(\tau)B_{\text{norm}}(\tau)}{|\text{Aut}(U)|} \right) H_{IdU}
\end{bmatrix}
\]

\[
+ B_{\text{norm}}(\tau) \begin{bmatrix}
\frac{S(\tau)R(\tau)}{|\text{Aut}(U)|} H_{IdU} & H_\tau & \frac{S(\tau)R(\tau)}{|\text{Aut}(U)|} H_{IdU}
\end{bmatrix}
\]
By Lemma 6.15, \[
\begin{bmatrix}
\frac{S(\tau)R(\tau)}{|\text{Aut}(U)|}H_{id_U} & H_{\tau} \\
H_{\tau}^T & \frac{S(\tau)R(\tau)}{|\text{Aut}(U)|}H_{id_U}
\end{bmatrix} \succeq 0,
\]
so the second term above is positive semidefinite.

For the first term, by Lemma 6.11, \(H_{id_U} \succeq 0\) and by Corollary 13.11, \(\frac{1}{|\text{Aut}(U)|c(\tau)} - \frac{S(\tau)R(\tau)B_{\text{term}}(\tau)}{|\text{Aut}(U)|} \succeq 0\), which proves that the first term is also positive semidefinite.

13.2 Proof of Lemma 13.4

Lemma 13.12. Suppose \(0 < A < \frac{1}{4}\) is a constant such that \(\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-Ae}, \frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-2A}\) and \(\frac{k}{\sqrt{\lambda}} \leq d^{-Ae}\). For all \(m\) such that \(m \leq \frac{d^{1-e}}{\lambda}, m \leq \frac{k^2 - \epsilon}{\lambda}e\), for all \(U, V \in \mathcal{I}_{\text{mid}}\) where \(w(U) > w(V)\) and for all \(\gamma \in \Gamma_{U,V}\),

\[
h^w(V(\gamma) \setminus U) \gamma S(\gamma) \leq \left( \prod_{j \in V(\gamma) \setminus U \setminus V(\gamma)} \text{deg}^\gamma(j) - 1 \right)! \left( \frac{1}{d^{Be(|V(\gamma) \setminus (U \cup V(\gamma))| + \sum_{e \in E(\gamma)} l_e)}} \right)^2
\]

for some constant \(B > 0\) that depends only on \(C_\Delta\). In particular, it is independent of \(C_V\) and \(C_E\).

Proof. Suppose there is a vertex \(i \in V(\gamma) \setminus U \setminus V(\gamma)\) such that \(\text{deg}^\gamma(i)\) is odd, then \(S(\gamma) = 0\) and the inequality is true. So, assume \(\text{deg}^\gamma(i)\) is even for all vertices \(i \in V(\gamma) \setminus U \setminus V(\gamma)\). We have \(n^w(V(\gamma) \setminus U) = d^{\gamma_1 - |U_1|} m^{\gamma_2 - |U_2|} \). Plugging in \(S(\gamma)\), we get that we have to prove

\[
E := d^{\gamma_1 - |U_1|} m^{\gamma_2 - |U_2|} \left( \frac{k}{d} \right)^{2|\gamma_1 - |U_1|| - |V_1|} \Delta^{2|\gamma_2 - |U_2|| - |V_2|} \prod_{e \in E(\gamma)} \lambda_e \leq \frac{1}{d^{Be(|V(\gamma) \setminus (U \cup V(\gamma))| + \sum_{e \in E(\gamma)} l_e)}}
\]

Let \(S'\) be the set of vertices \(i \in U \setminus V(\gamma)\) that have \(\text{deg}^\gamma(i) \geq 1\). Let \(e, f\) be the number of type 1 vertices and the number of type 2 vertices in \(S'\) respectively. Observe that \(S' \cup (U \setminus V(\gamma))\) is a vertex separator of \(\gamma\).

Let \(g = |V(\gamma) \setminus U_1|\) (resp. \(h = |V(\gamma) \setminus U_2|\)) be the number of type 1 vertices (resp. type 2 vertices) in \(V(\gamma) \setminus U_\gamma\).

We first claim that \(d^e m_f \geq d^g m_h\). To see this, note that the vertex separator \(S' \cup (U \setminus V(\gamma))\) has weight \(\sqrt{d^{e + |U \cap V(\gamma)|}} \sqrt{m^{f + |U \cap V(\gamma)|}}\). On the other hand, \(V(\gamma)\) has weight \(\sqrt{d^{e + |U \cap V(\gamma)|}} \sqrt{m^{f + |U \cap V(\gamma)|}}\). Since \(\gamma\) is a left shape, \(V(\gamma)\) is the unique minimum vertex separator and hence, \(\sqrt{d^{e + |U \cap V(\gamma)|}} \sqrt{m^{f + |U \cap V(\gamma)|}} \geq \sqrt{d^{e + |U \cap V(\gamma)|}} \sqrt{m^{f + |U \cap V(\gamma)|}}\), which implies \(d^e m_f \geq d^g m_h\).

Let \(p = |V(\gamma) \setminus (U \cup V(\gamma))|\) (resp. \(q = |V(\gamma) \setminus (U \cup V(\gamma))|\)) be the number of type 1 vertices (resp. type 2 vertices) in \(V(\gamma) \setminus (U \cup V(\gamma))\).

To illustrate the main idea, we will first prove the weaker inequality \(E \leq 1\). Since \(\Delta \leq 1\), it suffices to prove

\[
d^{\gamma_1 - |U_1|} m^{\gamma_2 - |U_2|} \left( \frac{k}{d} \right)^{2|\gamma_1 - |U_1|| - |V_1|} \prod_{e \in E(\gamma)} \lambda_e \leq 1
\]

We have

\[
d^{\gamma_1 - |U_1|} m^{\gamma_2 - |U_2|} \leq d^{p + s} m^{q + h} \leq n^{p + s - \frac{\epsilon}{\lambda}} m^{q + h}
\]

115
since $d^e m^f \geq d^f m^h$. Also, $2|\gamma|_1 - |U_\gamma|_1 - |V_\gamma|_1 = 2p + e + g$. So, it suffices to prove

$$n^{p+\frac{c+g}{2}} m^{q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2p+e+g} \prod_{e \in E(\gamma)} \left(\frac{\lambda}{k}\right)^{l_e} \leq 1$$

We will need the following claim.

**Claim 13.9.** $\sum_{e \in E(\gamma)} l_e \geq \max(2p + e + g, 2q + f + h)$

**Proof.** Since $H_\gamma$ is bipartite, we have $\sum_{e \in E(\gamma)} l_e = \sum_{i \in V_1(\gamma)} \deg^\gamma(i) = \sum_{i \in V_2(\gamma)} \deg^\gamma(i)$. Observe that all vertices $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$ have $\deg^\gamma(i)$ nonzero and even, and hence, $\deg^\gamma(i) \geq 2$. Then,

$$\sum_{e \in E(\gamma)} l_e = \sum_{i \in V_1(\gamma)} \deg^\gamma(i) \geq \sum_{i \in V_1(\gamma) \setminus U_\gamma \setminus V_\gamma} \deg^\gamma(i) + \sum_{i \in (U_\gamma)_1 \setminus V_\gamma} \deg^\gamma(i) + \sum_{i \in (V_\gamma)_1 \setminus U_\gamma} \deg^\gamma(i) \geq 2p + e + g$$

Similarly,

$$\sum_{e \in E(\gamma)} l_e = \sum_{i \in V_2(\gamma)} \deg^\gamma(i) \geq \sum_{i \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} \deg^\gamma(i) + \sum_{i \in (U_\gamma)_2 \setminus V_\gamma} \deg^\gamma(i) + \sum_{i \in (V_\gamma)_2 \setminus U_\gamma} \deg^\gamma(i) \geq 2q + f + h$$

Therefore, $\sum_{e \in E(\gamma)} l_e \geq \max(2p + e + g, 2q + f + h)$. □

Now, let $r_1 = p + \frac{c+g}{2}$, $r_2 = q + \frac{f+h}{2}$. Then, $\sum_{e \in E(\gamma)} l_e \geq 2 \max(r_1, r_2)$ and we wish to prove

$$d^e m^f \left(\frac{k}{d}\right)^{2r_1} \left(\frac{\lambda}{k}\right)^{2\max(r_1, r_2)} \leq 1$$

This expression simply follows by squaring Claim 13.9.

Now, to prove that $E \leq \frac{1}{d^{m(\gamma_1 - |U_\gamma|_1 - |V_\gamma|_1)} d^{m(\gamma_2 - |U_\gamma|_2 - |V_\gamma|_2)} \prod_{e \in E(\gamma)} \lambda^l_e \prod_{k \leq e} \leq \frac{1}{d^{m(\gamma_1 - |U_\gamma|_1 - |V_\gamma|_1)} d^{m(\gamma_2 - |U_\gamma|_2 - |V_\gamma|_2)} \prod_{e \in E(\gamma)} \lambda^l_e \prod_{k \leq e}}{d^{B_m(|V(\gamma)| - (U_\gamma \cap V_\gamma)) + \sum_{e \in E(\gamma)} l_e}}$

The idea is that the $d^{Be}$ decay for the edges are obtained from the stronger assumption on $m$, namely $m \geq \frac{d_{\gamma, \gamma}^c}{\Delta^2}$, $m \leq \frac{k_{\gamma, \gamma}}{\Delta^2}$. And the $d^{Be}$ decay for the type 1 vertices of $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$ are obtained both from the stronger assumption on $m$ as well as the factors of $\frac{k}{\Delta}$, the latter especially useful for the degree 0 vertices. Finally, the $d^{Be}$ decay for the type 2 vertices of $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$ are obtained from the factors of $\Delta$.  

116
Indeed, note that for a constant $B$ that depends on $C_\Delta$, $\Delta^2|\gamma|_2-|U_\gamma|_2-|V_\gamma|_2 \leq d^{B(eV(\gamma)\setminus(U_\gamma\cap V_\gamma))_2}$. So, we would be done if we prove
\[
d^{p+\frac{c+g}{2}m^q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1} \frac{\Sigma_{e \in E(\gamma)} l_e}{d^{Be(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)_1+\Sigma_{e \in E(\gamma)} l_e)}} \leq \frac{1}{d^{Be(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)_1+\Sigma_{e \in E(\gamma)} l_e)}}
\]

Let $c_0$ be the number of type 1 vertices $i$ in $V(\gamma)\setminus(U_\gamma\cap V_\gamma)$ such that $\deg^{\gamma}(i) = 0$. Since they have degree 0, they must be in $(U_\gamma)_1 \setminus V_\gamma$. Also, we have $2|\gamma|_1 - |U_\gamma|_1 - |V_\gamma|_1 = 2p + e + g + c_0$ and hence, $(\frac{k}{d})^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1} = (\frac{k}{d})^{2p+e+g+c_0}$. For these degree 0 vertices, we have that the factors of $\frac{k}{d}$ offer a decay of $\frac{1}{e^{Be}}$. Therefore, it suffices to prove
\[
d^{p+\frac{c+g}{2}m^q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2p+e+g} \frac{\Sigma_{e \in E(\gamma)} l_e}{d^{Be(p+q+e+f+g+h)+\Sigma_{e \in E(\gamma)} l_e}} \leq \frac{1}{d^{Be\Sigma_{e \in E(\gamma)} l_e}}
\]
for a constant $B > 0$. Observe that $p + q + e + f + g + h \leq 2(\Sigma_{e \in E(\gamma)} l_e)$. Therefore, using the notation $r_1 = p + \frac{c+g}{2}, r_2 = q + \frac{f+h}{2}$, it suffices to prove
\[
d^{r_1}m^{r_2} \left(\frac{k}{d}\right)^{2r_1} \frac{\Sigma_{e \in E(\gamma)} l_e}{d^{Be\Sigma_{e \in E(\gamma)} l_e}} \leq \frac{1}{d^{Be\Sigma_{e \in E(\gamma)} l_e}}
\]
for a constant $B > 0$. But this follows by squaring Claim 13.10 where we set $r = \Sigma_{e \in E(\gamma)} l_e$. 

**Remark 13.14.** In the above bounds, note that there is a decay of $d^{Be}$ for each vertex in $V(\gamma)\setminus(U_\gamma\cap V_\gamma)$. One of the main technical reasons for introducing the slack parameter $C_\Delta$ in the planted distribution was to introduce this decay, which is needed in the current machinery.

We can now prove Lemma 13.4.

**Lemma 13.4.** For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,

\[
c(\gamma)^2N(\gamma)^2B(\gamma)^2H_{ldV}^{\gamma,\gamma} \leq H'_\gamma
\]

**Proof.** By Lemma 6.17, we have
\[
c(\gamma)^2N(\gamma)^2B(\gamma)^2H_{ldV}^{\gamma,\gamma} \leq c(\gamma)^2N(\gamma)^2B(\gamma)^2S(\gamma)^2R(\gamma)^2\frac{|\text{Aut}(U)|}{|\text{Aut}(V)|}H'_\gamma
\]

Using the same proof as in Lemma 6.11, we can see that $H'_\gamma \succeq 0$. Therefore, it suffices to prove that
\[
c(\gamma)^2N(\gamma)^2B(\gamma)^2S(\gamma)^2R(\gamma)^2\frac{|\text{Aut}(U)|}{|\text{Aut}(V)|} \leq 1
\]

Since $U, V \in \mathcal{I}_{mid}$, $\text{Aut}(U) = |U_1|!|U_2|!$, $\text{Aut}(V) = |V_1|!|V_2|!$. Therefore, $\frac{|\text{Aut}(U)|}{|\text{Aut}(V)|} = \frac{|U_1|!|U_2|!}{|V_1|!|V_2|!} \leq D_V^{|U_\gamma\setminus V_\gamma|}$. Also, $|E(\gamma)| \leq \Sigma_{e \in E(\gamma)} l_e$ and $d = 2^{O(1)}-\varepsilon(C_V+C_\varepsilon)$. Note that $R(\gamma)^2 = (C_{\text{disc}}\sqrt{D_E}2^{\Sigma_{e \in E(\gamma)} l_e})^{2\deg(\gamma)} \leq d^{O(1)}\varepsilon(C_V+C_\varepsilon)\Sigma_{e \in E(\gamma)} l_e$. 

117
Let $B$ be the constant from Lemma 13.12. We can set $C_V, C_E$ sufficiently small so that, using Lemma 13.12,

$$
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{\text{Aut}(U)}{\text{Aut}(V)} \leq 100^2 (6D_V)^2 |U_\gamma \setminus V_\gamma| + 2 |V_\gamma \setminus U_\gamma| + |E(\gamma)| \sum_{e \in E(\gamma)} |V(\gamma)\setminus (U_\gamma \cup V_\gamma)| \cdot (3D_V)^4 |V(\gamma)| |V_\gamma| + 2 |V(\gamma)| |U_\gamma| (6qD_V)^2 |V(\gamma)| |U_\gamma| + 2 |V(\gamma)| |V_\gamma| \prod_{e \in E(\gamma)} (400D_V^2 D_E^2 q)^{2l_e}
$$

$$
\cdot n^\omega(V(\gamma)|U_\gamma|) S(\gamma)^2 d^O(1) e C_E \sum_{e \in E(\gamma)} l_e \cdot D_V^{U_\gamma \setminus V_\gamma}
\leq d^O(1) e (C_V + C_E) \cdot (|V(\gamma)| |U_\gamma \cap V_\gamma| + \sum_{e \in E(\gamma)} l_e) \cdot n^\omega(V(\gamma)|U_\gamma|) S(\gamma)^2
\leq d^O(1) e (C_V + C_E) \cdot (|V(\gamma)| |U_\gamma \cap V_\gamma| + \sum_{e \in E(\gamma)} l_e) \cdot \frac{1}{d^B(\gamma)|V(\gamma)| |U_\gamma \cap V_\gamma| + \sum_{e \in E(\gamma)} l_e)}
\leq 1
$$

\[\blacksquare\]

13.3 Proof of Lemma 13.5

In this section, we will prove Lemma 13.5 using the strategy sketched in Section 10.

**Lemma 13.5.** Whenever $\|M_\alpha\| \leq B_\text{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,

$$
\sum_{U \in \mathcal{L}_{mid}} M_{I_dU}^{\text{fact}}(H_{I_dU}) \geq 6 \left( \sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_U} \frac{d_{I_dU}(H'_\gamma, H_{I_dU})}{|\text{Aut}(U)| c(\gamma)} \right) Id_{sym}
$$

In particular, we prove the following lemmas.

**Lemma 13.15.** Whenever $\|M_\alpha\| \leq B_\text{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,

$$
\sum_{U \in \mathcal{L}_{mid}} M_{I_dU}^{\text{fact}}(H_{I_dU}) \geq \frac{1}{dK_1 D_{los}^2} Id_{sym}
$$

for a constant $K_1 > 0$ that can depend on $C_\Delta$.

**Lemma 13.16.**

$$
\sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_U} \frac{d_{I_dU}(H_{I_dU}, H'_\gamma)}{|\text{Aut}(U)| c(\gamma)} \leq \frac{dK_2 D_{los}}{2D_V}
$$

for a constant $K_2 > 0$ that can depend on $C_\Delta$.

If we assume the above lemmas, we can prove Lemma 13.5.

**Lemma 13.5.** Whenever $\|M_\alpha\| \leq B_\text{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,

$$
\sum_{U \in \mathcal{L}_{mid}} M_{I_dU}^{\text{fact}}(H_{I_dU}) \geq 6 \left( \sum_{U \in \mathcal{L}_{mid}} \sum_{\gamma \in \Gamma_U} \frac{d_{I_dU}(H'_\gamma, H_{I_dU})}{|\text{Aut}(U)| c(\gamma)} \right) Id_{sym}
$$

118
Proof. Let \( \| M_\alpha \| \leq B_{\text{norm}}(\alpha) \) for all \( \alpha \in \mathcal{M}' \). By Lemma 13.15,

\[
\sum_{U \in \mathcal{L}_{\text{mid}}} M_{Id_U}^{\text{fact}}(H_{Id_U}) \geq \frac{1}{d^k D_{\text{sos}}^2} Id_{\text{sym}}
\]

for a constant \( K_1 > 0 \). By Lemma 13.16,

\[
\sum_{U \in \mathcal{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,\ast}} \frac{d_{Id_U}(H_{Id_U}, H'_{\gamma})}{|\text{Aut}(U)|c(\gamma)} \leq \frac{d D_{\text{sos}}^2}{2 D_U}
\]

for a constant \( K_2 > 0 \).

We choose \( C_{\text{sos}} \) sufficiently small so that \( \frac{1}{d^k D_{\text{sos}}^2} \geq 6 \frac{d K_2 D_{\text{sos}}}{2 D_U} \) which can be satisfied by setting \( C_{\text{sos}} < K_3 C_V \) for a sufficiently small constant \( K_3 > 0 \). Then, since \( Id_{\text{sos}} \succeq 0 \), using Lemma 13.15 and Lemma 13.16,

\[
\sum_{U \in \mathcal{L}_{\text{mid}}} M_{Id_U}^{\text{fact}}(H_{Id_U}) \geq \frac{1}{d^k D_{\text{sos}}^2} Id_{\text{sym}}
\]

\[
\geq 6 \frac{d K_2 D_{\text{sos}}}{2 D_U} Id_{\text{sym}}
\]

\[
\geq 6 \left( \sum_{U \in \mathcal{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U,\ast}} \frac{d_{Id_U}(H'_{\gamma}, H_{Id_U})}{|\text{Aut}(U)|c(\gamma)} \right) Id_{\text{sym}}
\]

In the rest of the section, we will prove Lemma 13.15 and Lemma 13.16.

To begin with, we will need a bound on \( B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{Id_U}(\sigma, \sigma') \).

Lemma 13.17. Suppose \( 0 < A < \frac{1}{4} \) is a constant such that \( \frac{\sqrt{A}}{\sqrt{k}} \leq d^{-A} \) and \( \frac{1}{\sqrt{k}} \leq d^{-2A} \). Suppose \( m \) is such that \( m \leq \frac{d_{V\sigma}}{2^A}, m \leq \frac{k_{V\sigma}}{2^A} \). For all \( U \in \mathcal{L}_{\text{mid}} \) and \( \sigma, \sigma' \in \mathcal{L}_U \),

\[
B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{Id_U}(\sigma, \sigma') \leq \frac{d^{O(1)} D_{\text{sos}}}{d^{0.5 A / |V(\sigma')|}}
\]

Proof. Suppose there is a vertex \( i \in V(\sigma) \setminus V_\sigma \) such that \( d_{\text{g}}(i) + d_{\text{g}}(\sigma) \) is odd, then \( H_{Id_U}(\sigma, \sigma') = 0 \) and the inequality is true. So, assume that \( d_{\text{g}}(i) + d_{\text{g}}(\sigma) \) is even for all \( i \in V(\sigma) \setminus V_\sigma \). Similarly, assume that \( d_{\text{g}}(i) + d_{\text{g}}(\sigma) \) is even for all \( i \in V(\sigma') \setminus V_{\sigma'} \). Also, if \( \rho_\sigma \neq \rho_{\sigma'} \), we will have \( H_{Id_U}(\sigma, \sigma') = 0 \) and we would be done. So, assume \( \rho_\sigma = \rho_{\sigma'} \).

Let there be \( e \) (resp. \( f \)) vertices of type 1 (resp. type 2) in \( V(\sigma) \setminus U_\sigma \setminus V_{\sigma} \). Then,

\[
n^{\frac{|V(\sigma) - \text{ver}(U)|}{2}} = \sqrt{d}^{\frac{|V(\sigma)| - |U_1|}{2}} \frac{|V(\sigma)| - |U_1|}{\sqrt{|V(\sigma)| - |U_1|}}
\]

\[
= \sqrt{d}^{\frac{|U_1|}{2}} \frac{|U_1|}{\sqrt{|U_1|}} \sqrt{d^e \frac{\sqrt{m}}{m^f}}
\]

\[
\leq d^{O(1)} D_{\text{sos}} \sqrt{d} \frac{\sqrt{m}}{m^f}
\]

where we used the fact that \( |U_\sigma| \leq D_{\text{sos}} \).
Let there be $g$ (resp. $h$) vertices of type 1 (resp. type 2) in $V(\sigma') \setminus U_{\sigma'} \setminus V_{\sigma'}$. Then, similarly, $n^{\frac{w(V(\sigma') - w(U))}{2}} \leq d^{O(1)D_{bos}} \sqrt{d^2 \sqrt{m^h}}$.

Let $\alpha = \sigma \circ \sigma'$. Since all vertices in $V(\alpha) \setminus U_{\alpha} \setminus V_{\alpha}$ have degree at least 2, we have $\sum_{e \in E(\alpha)} l_e \geq \sum_{i \in V_1(\alpha) \setminus U_{\alpha} \setminus V_{\alpha}} deg^a(i) \geq 2(e + g)$. Similarly, $\sum_{e \in E(\alpha)} l_e \geq 2(f + h)$. Therefore, by setting $r_1 = e + g, r_2 = f + h$ in Claim 13.10, we have

$$\sqrt{d} \cdot \sqrt{m^{f+h}} \left( \frac{k}{d} \right)^{{\epsilon+g}} \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda^k}}{\sqrt{k}} \leq \frac{1}{d^{dA_\alpha \sum_{e \in E(\alpha)} l_e}}$$

Also, $(\frac{k}{d})^{|\alpha|_1} \leq (\frac{k}{d})^{{\epsilon+g}}$ and $\prod_{j \in V_2(\alpha)} (deg^a(j) - 1)!! \leq d^{C_V \sum_{e \in E(\alpha)} l_e}$. Therefore,

$$n^{\frac{w(V(\sigma') - w(U))}{2}} n^{\frac{w(V(\sigma') - w(U))}{2}} H_{I_{dU}}(\sigma, \sigma')$$

$$\leq d^{O(1)D_{bos}} \sqrt{d} \sqrt{m^{f+h}} d^{O(1)D_{bos}} \sqrt{d^2 \sqrt{m^h}} \cdot \frac{1}{|Aut(U)|} \left( \frac{1}{\sqrt{k}} \right)^{deg(a)} \left( \frac{k}{d} \right)^{|\alpha|_1} \Delta^{|\alpha|_2} \prod_{j \in V_2(\alpha)} (deg^a(j) - 1)!! \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda^k}}{\sqrt{k}}$$

$$\leq d^{O(1)D_{bos}} d^{C_V \sum_{e \in E(\alpha)} l_e} \sqrt{d} \sqrt{m^{f+h}} \left( \frac{k}{d} \right)^{{\epsilon+g}} \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda^k}}{\sqrt{k}}$$

$$\leq d^{O(1)D_{bos}} d^{C_V \sum_{e \in E(\alpha)} l_e} \frac{1}{d^{dA_\alpha \sum_{e \in E(\alpha)} l_e}}$$

Now, observe that since all vertices in $V(\alpha) \setminus U_{\alpha} \setminus V_{\alpha}$ have degree at least 1, $|V(\alpha)| \leq 2D_{bos} + 2 \sum_{e \in E(\alpha)} l_e$. So, by setting $C_V, C_E$ sufficiently small,

$$B_{norm}(\sigma)B_{norm}(\sigma')H_{I_{dU}}(\sigma, \sigma') = 2e(6qD_V)^{|V(\sigma') \setminus U_{\sigma'}| + |V(\sigma) \setminus V_{\sigma}|} \prod_{e \in E(\sigma)} (400D_V^2 D_E^2 q)^l n \frac{w(V(\sigma') - w(U))}{2}$$

$$\cdot 2e(6qD_V)^{|V(\sigma') \setminus U_{\sigma'}| + |V(\sigma) \setminus V_{\sigma}|} \prod_{e \in E(\sigma')} (400D_V^2 D_E^2 q)^l n \frac{w(V(\sigma') - w(U))}{2}$$

$$\cdot H_{I_{dU}}(\sigma, \sigma')$$

$$\leq d^{O(1)}d^{C_V + C_E} - (|V(\alpha)| + \sum_{e \in E(\alpha)} l_e) d^{O(1)D_{bos}} d^{C_V \sum_{e \in E(\alpha)} l_e} \frac{1}{d^{dA_\alpha \sum_{e \in E(\alpha)} l_e}}$$

$$\leq d^{O(1)D_{bos}} d^{0.5A_\alpha |V(\alpha)|}$$

### 13.3.1 Proof of Lemma 13.15

To prove Lemma 13.15, we will use the strategy from Section 10.1. We will also use the notation from that section. We recall that for $U \in \mathcal{I}_{mid}$, $L_{U}^\prime \subset L_U$ was the set of non-trivial shapes in $L_U$.  

120
Proposition 13.18. For \( V \in \mathcal{I}_{\text{mid}} \),
\[
\lambda_V = \frac{\Delta|V|_2}{d|V|_1 k|V|_2}
\]

Proof. We have \( \lambda_V = |\text{Aut}(V)| H_{Id_V}(\text{Id}_V, \text{Id}_V) = \left( \frac{1}{\sqrt{k}} \right)^{2|V|} \left( \frac{k}{\pi} \right)^{|V|_1} \Delta|V|_2 = \frac{\Delta|V|_2}{d|V|_1 k|V|_2}. \)

Corollary 13.19. \( \lambda_V \geq \frac{1}{d^{O(1)} D_{\text{sos}}} \)

Lemma 13.20. For any edge \( e = (V, U) \) in \( G \), we have
\[
\omega(e) \leq \frac{d^{O(1)} D_{\text{sos}}}{d^{0.1Ae[V(\sigma \circ \sigma')]}|V(\sigma \circ \sigma')|}
\]

Proof. Let \( e = (V, U) \) be an edge in \( G \). Then, \( \omega(U) > \omega(V) \) and \( \omega(e) = \frac{2W(U, V)}{\lambda_V} \). Using Lemma 13.17, we have
\[
2W(U, V) = \frac{2}{|\text{Aut}(U)|} \sum_{\sigma \in L_V, U_0 = U} \sum_{\sigma' \in L_V, U_0 \neq V} B_{\text{norm}}(\sigma) B_{\text{norm}}(\sigma') H_{Id_U}(\sigma, \sigma')
\]
\[
\leq \frac{2}{|\text{Aut}(U)|} \sum_{\sigma \in L_V, U_0 = U} \sum_{\sigma' \in L_V, U_0 \neq V} \frac{d^{O(1)} D_{\text{sos}}}{d^{0.1Ae[V(\sigma \circ \sigma')]}}
\]
\[
\leq \sum_{\sigma, \sigma' \in L_V} \frac{d^{O(1)} D_{\text{sos}}}{d^{0.1Ae[V(\sigma \circ \sigma')]}}
\]
\[
\leq \frac{d^{O(1)} D_{\text{sos}}}{d^{0.1Ae[V(\sigma \circ \sigma')]}} \sum_{\sigma, \sigma' \in L_V} \frac{1}{D_{\text{sos}} d^{2Ae[V(\sigma \circ \sigma')]}}
\]
\[
\leq \frac{\lambda_V d^{O(1)} D_{\text{sos}}}{d^{0.1Ae[V(\sigma \circ \sigma')]}}
\]

where we set \( C_V, C_E \) small enough and invoked Lemma 10.16. Rearranging proves the lemma.

Corollary 13.21. For any \( U, V \in \mathcal{I}_{\text{mid}} \) such that \( \omega(U) > \omega(V) \),
\[
\sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} \omega(e) \leq d^{O(1)} D_{\text{sos}}^2
\]

Proof. The total number of vertices in \( G \) is at most \((D_{\text{sos}} + 1)^2\) since each \( U \in \mathcal{I}_{\text{mid}} \) has at most 2 index shape pieces corresponding to each type and each index shape piece has at most \( D_{\text{sos}} \) vertices. Therefore, for any fixed integer \( j \geq 1 \), the number of paths from \( V \) to \( U \) of length \( j \) is at most \((D_{\text{sos}} + 1)^{2j}\). Take any
path $P$ from $V$ to $U$. Suppose it has length $j \geq 1$. Note that for all edges $e = (V', U')$ in $E(P)$, since $|U'| \geq 1$, we have
\[ w(e) \leq \frac{d^{O(1)}}{d^{0.1A} |U'|} \leq \frac{d^{O(1)}}{d^{0.1A}}. \]
So, \( \prod_{e \in E(P)} w(e) \leq \left( \frac{d^{O(1)}}{d^{0.1A}} \right)^j \). Therefore, by setting $C_{sos}$ small enough,
\[
\sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \leq \left( d^{O(1)} D_{sos} + 1 \right)^2 \left( \frac{d^{O(1)}}{d^{0.1A}} \right)^j \leq d^{O(1)} D_{sos}^2.
\]

We can now prove Lemma 13.15.

**Lemma 13.15.** Whenever \( \| M_\alpha \| \leq B_{\text{norm}}(\alpha) \) for all $\alpha \in \mathcal{M}'$,
\[
\sum_{U \in \mathcal{I}_{\text{mid}}} M_{\text{fact}}^{\text{Id}_U} (H_{\text{Id}_U}) \geq \frac{1}{d^{K_1 D_{sos}^2}} \text{Id}_{\text{sym}}
\]
for a constant $K_1 > 0$ that can depend on $C_\Delta$.

**Proof.** For all $V \in \mathcal{I}_{\text{mid}}$, we have
\[
\text{Id}_{\text{sym}, V} \leq 2 \sum_{U \in \mathcal{I}_{\text{mid}}, w(U) \geq w(V)} \left( \sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \right) \frac{1}{\lambda_U} M_{\text{fact}}^{\text{Id}_U} (H_{\text{Id}_U})
\]
Summing this over all $V \in \mathcal{I}_{\text{mid}}$, we get
\[
\text{Id}_{\text{sym}} \leq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{2}{\lambda_U} \left( \sum_{V \in \mathcal{I}_{\text{mid}}, w(U) \geq w(V)} \sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \right) M_{\text{fact}}^{\text{Id}_U} (H_{\text{Id}_U})
\]
For any fixed $U \in \mathcal{I}_{\text{mid}}$, the number of $V \in \mathcal{I}_{\text{mid}}$ such that $w(U) \geq w(V)$ is at most $(D_{sos} + 1)^2$. Also, $\lambda_U \geq \frac{1}{d^{O(1) D_{sos}^2}}$ for all $U \in \mathcal{I}_{\text{mid}}$. Therefore,
\[
\text{Id}_{\text{sym}} \leq \sum_{U \in \mathcal{I}_{\text{mid}}} \frac{2}{\lambda_U} (D_{sos} + 1)^2 d^{O(1) D_{sos}^2} M_{\text{fact}}^{\text{Id}_U} (H_{\text{Id}_U})
\]
where we used the fact that for all $U \in \mathcal{I}_{\text{mid}}$, $M_{\text{fact}}^{\text{Id}_U} (H_{\text{Id}_U}) \geq 0$.\qed
13.3.2 Proof of Lemma 13.16

We restate the lemma for convenience.

**Lemma 13.16.**

\[ \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{id}_U}(H_{\text{id}_U}, H'_{\gamma})}{|\text{Aut}(U)|c(\gamma)} \leq \frac{d_{\text{Kos}}K_2}{2D_V} \]

for a constant \( K_2 > 0 \) that can depend on \( C_\Delta \).

**Proof.** We have

\[ \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{id}_U}(H_{\text{id}_U}, H'_{\gamma})}{|\text{Aut}(U)|c(\gamma)} = \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{1}{|\text{Aut}(U)|c(\gamma)} \sum_{\gamma' \in \mc{L}_{U^*}(\gamma \circ \sigma)} B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma')H_{\text{id}_U}(\sigma, \sigma') \]

The set of \( \sigma, \sigma' \) that could appear in the above sum must necessarily be non-trivial and hence, \( \sigma, \sigma' \in \mc{L}_U \). Then,

\[ \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{id}_U}(H_{\text{id}_U}, H'_{\gamma})}{|\text{Aut}(U)|c(\gamma)} \]

\[ = \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\sigma, \sigma' \in \mc{L}_U} B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma') \sum_{\gamma \in \Gamma_{U^*}, \gamma' \in \mc{L}_U(\gamma \circ \sigma), |V(\gamma)| \geq |U| + m_\sigma} \frac{1}{|\text{Aut}(U)|c(\gamma)} \]

For \( \sigma \in \mc{L}_U \), define \( m_\sigma = D_V + 1 - |V(\sigma)| \geq 1 \). This is precisely set so that for all \( \gamma \in \Gamma_{U^*} \), we have \( |V(\gamma \circ \sigma)| > D_V \) if and only if \( |V(\gamma)| \geq |U| + m_\sigma \). So, for \( \sigma, \sigma' \in \mc{L}_U \), using Lemma 9.31,

\[ \sum_{\gamma \in \Gamma_{U^*}, |V(\gamma)| \geq |\gamma'| + m_{\sigma'} \} \frac{1}{|\text{Aut}(U)|c(\gamma)} \leq \frac{1}{2\min(m_{\sigma}, m_{\sigma'})} \]

Also, for \( \sigma, \sigma' \in \mc{L}_U \), we have \( |V(\gamma \circ \sigma')| + \min(m_{\sigma}, m_{\sigma'}) \geq D_V \). Therefore,

\[ \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\gamma \in \Gamma_{U^*}} \frac{d_{\text{id}_U}(H_{\text{id}_U}, H'_{\gamma})}{|\text{Aut}(U)|c(\gamma)} \leq \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\sigma, \sigma' \in \mc{L}_U} B_{\text{norm}}(\sigma)B_{\text{norm}}(\sigma') \frac{1}{2\min(m_{\sigma}, m_{\sigma'})} \]

\[ \leq \sum_{U \in \mc{L}_{\text{mid}}} \sum_{\sigma, \sigma' \in \mc{L}_U} \frac{d_{O(1)D_{\text{los}}}}{d_{O(1)D_{\text{los}}}^2} \frac{1}{2\min(m_{\sigma}, m_{\sigma'})} \]
where we used Lemma 13.17. Using $d^{0.5|V(\sigma \circ \sigma')|} \geq d^{0.1|V(\sigma \circ \sigma')|} 2|V(\sigma \circ \sigma')|$, 

$$
\sum_{U \in \mathcal{L}_0} \sum_{\gamma \in \Gamma(U)} \frac{d_{dij}(H_{dij}, H'_{dij})}{|Aut(U)|^2 (\gamma)} \leq \sum_{U \in \mathcal{L}_0} \sum_{\sigma, \sigma' \in \mathcal{L}'_0} d^{O(1) D_{sos}} ^{0.1|V(\sigma \circ \sigma')|} 2|V(\sigma \circ \sigma')|^2 \min(m_{\sigma}, m_{\sigma'}) - 1
$$

$$
\leq \sum_{U \in \mathcal{L}_0} \sum_{\sigma, \sigma' \in \mathcal{L}'_0} d^{O(1) D_{sos}} ^{0.1|V(\sigma \circ \sigma')|} 2D_{V}
$$

$$
\leq \sum_{U \in \mathcal{L}_0} \sum_{\sigma, \sigma' \in \mathcal{L}'_0} D_{sos} d^{O(1) D_{sos}} ^{0.1|V(\sigma \circ \sigma')|} 2D_{V}
$$

The final step will be to argue that $\sum_{U \in \mathcal{L}_0} \sum_{\sigma, \sigma' \in \mathcal{L}'_0} \frac{1}{D_{sos} d^{0.1|V(\sigma \circ \sigma')|} 2D_{V}} \leq 1$ which will complete the proof. But this will follow from Lemma 10.16 if we set $C_V, C_E$ small enough.

Acknowledgements

We thank Sam Hopkins, Pravesh Kothari, Prasad Raghavendra, Tselil Schramm, and David Steurer for helpful discussions. We also thank Sam Hopkins and Pravesh Kothari for assistance in drafting the informal description of the machinery (Section 3).

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.

[AMS11] Genevera I Allen and Mirjana Maletić-Savatić. Sparse non-negative generalized pca with applications to metabolomics. Bioinformatics, 27(21):3029–3035, 2011.

[ARV04] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut. 2004.

[AW08] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In 2008 IEEE international symposium on information theory, pages 2454–2458. IEEE, 2008.

[BAP+05] Jinho Baik, Gérard Ben Arous, Sandrine Péché, et al. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. The Annals of Probability, 33(5):1643–1697, 2005.

[BB19] Matthew Brennan and Guy Bresler. Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. arXiv preprint arXiv:1902.07380, 2019.

[BBH+12] Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. CoRR, abs/1205.4484, 2012.

[BBH+20] Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. arXiv preprint arXiv:2009.06107, 2020.
[BCC⁺¹⁰] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an o(n^1/4) approximation for densest k-subgraph. In Proceedings of the forty-second ACM symposium on Theory of computing, pages 201–210, 2010. 4

[BCG⁺¹²] Aditya Bhaskara, Moses Charikar, Venkatesan Guruswami, Aravindan Vijayaraghavan, and Yuan Zhou. Polynomial integrality gaps for strong sdp relaxations of densest k-subgraph. In Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms, pages 388–405. SIAM, 2012. 4

[BGG⁺¹⁷] Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. Weak decoupling, polynomial folds and approximate optimization over the sphere. In 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), pages 1008–1019. IEEE, 2017. 5

[BGL16] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. Sum-of-squares certificates for maxima of random tensors on the sphere. arXiv preprint arXiv:1605.00903, 2016. 5

[BR⁺¹³b] Quentin Berthet, Philippe Rigollet, et al. Optimal detection of sparse principal components in high dimension. The Annals of Statistics, 41(4):1780–1815, 2013. 6

[BV09] S Charles Brubaker and Santosh S Vempala. Random tensors and planted cliques. In Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pages 406–419. Springer, 2009. 5

[CK09] Hyonho Chun and Sündüz Keleş. Expression quantitative trait loci mapping with multivariate sparse partial least squares regression. Genetics, 182(1):79–90, 2009. 5

[dKNS20] Tommaso d’Orsi, Pravesh K. Kothari, Gleb Novikov, and David Steurer. Sparse pca: Algorithms, adversarial perturbations and certificates. In 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), 2020. 6
Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), pages 73–84. IEEE, 2017.

Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. The Journal of Machine Learning Research, 17(1):4913–4953, 2016.

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

Alan Frieze and Ravi Kannan. A new approach to the planted clique problem. In IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2008.

Uriel Feige, David Peleg, and Guy Kortsarz. The dense k-subgraph problem. Algorithmica, 29(3):410–421, 2001.

Vitaly Feldman, Will Perkins, and Santosh Vempala. On the complexity of random satisfiability problems with planted solutions. SIAM Journal on Computing, 47(4):1294–1338, 2018.

 Uriel Feige, Michael Seltser, et al. On the densest k-subgraph problem. Citeseer, 1997.

Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for sherrington-kirkpatrick via planted affine planes. In 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), 2020.

Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. Computational Complexity, 10(2):139–154, 2001.

Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. Theor. Comput. Sci., 259(1-2):613–622, 2001.

Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In FOCS, pages 482–491, 2011.

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.

David Gamarnik and Ilias Zadik. The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. arXiv preprint arXiv:1904.07174, 2019.

Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tseltl Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), pages 720–731. IEEE, 2017.

Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method. 2018.

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.
A Proof that the Leftmost and Rightmost Minimum Vertex Separators are Well-defined

In this section, we give a general proof that the leftmost and rightmost minimum vertex separators are well-defined.

**Lemma A.1.** For any two distinct vertex separators $S_1$ and $S_2$ of $\alpha$, there exist vertex separators $S_L$ and $S_R$ of $\alpha$ such that:

1. $S_L$ is a vertex separator of $U_\alpha$ and $S_1$ and a vertex separator of $U_\alpha$ and $S_2$.
2. $S_R$ is a vertex separator of $S_1$ and $V_\alpha$ and a vertex separator of $S_2$ and $V_\alpha$.
3. $w(S_L) + w(S_R) \leq w(S_1) + w(S_2)$

**Proof.** Take $S_L$ to be the set of vertices $v \in V(\alpha) \cap (S_1 \cup S_2)$ such that there is a path from $U_\alpha$ to $v$ which doesn’t intersect $S_1 \cup S_2$ before reaching $v$. Similarly, take $S_R$ to be the set of vertices $v \in V(\alpha) \cap (S_1 \cup S_2)$ such that there is a path from $V_\alpha$ to $v$ which doesn’t intersect $S_1 \cup S_2$ before reaching $v$.  

128
Now observe that \( S_L \) is a vertex separator between \( U_a \) and \( S_1 \). To see this, note that for any path \( P \) from \( U_a \) to a vertex \( v \in S_1 \), either \( P \) intersects \( S_L \) before reaching \( v \) or \( P \) does not intersect \( S_L \) before reaching \( v \). In the latter case, \( v \in S_L \). Thus, in either case, \( P \) intersects \( S_L \). Following similar logic, \( S_L \) is also a vertex separator between \( U_a \) and \( S_2 \), \( S_R \) is a vertex separator between \( S_1 \) and \( V_a \), and \( S_R \) is also a vertex separator between \( S_2 \) and \( V_a \).

To show that \( w(S_L) + w(S_R) \leq w(S_1) + w(S_2) \), observe that \( w(S_L) + w(S_R) = w(S_R \cup S_R) + w(S_L \cap S_R) \) and \( w(S_1) + w(S_2) = w(S_1 \cup S_2) + w(S_1 \cap S_2) \). Thus, to show that \( w(S_L) + w(S_R) \leq w(S_1) + w(S_2) \), it is sufficient to show that

1. \( S_L \cup S_R \subseteq S_1 \cup S_2 \)
2. \( S_L \cap S_R \subseteq S_1 \cap S_2 \)

For the first statement, note that by definition any vertex in \( S_L \cup S_R \) must be in \( S_1 \cup S_2 \). For the second statement, note that if \( v \in S_L \cap S_R \) then there is a path from \( U_a \) to \( v \) which does not intersect any other vertices in \( S_1 \cup S_2 \) and there is a path from \( v \) to \( V_a \) which does not intersect any other vertices in \( S_1 \cup S_2 \). Combining these paths, we obtain a path \( P \) from \( U_a \) to \( V_a \) such that \( v \) is the only vertex in \( P \) which is in \( S_1 \cup S_2 \). This implies that \( v \in S_1 \cap S_2 \) as otherwise either \( S_1 \) or \( S_2 \) would not be a vertex separator between \( U_a \) and \( V_a \).

**Corollary A.2.** The leftmost and rightmost minimum vertex separators between \( U_a \) and \( V_a \) are well-defined.

**Proof.** Assume that there is no minimum leftmost vertex separator. If so, then there exists a minimum vertex separator \( S_1 \) between \( U_a \) and \( V_a \) such that

1. There does not exist a minimum vertex separator \( S' \) of \( \alpha \) such that \( S' \) is also a minimum vertex separator of \( U_a \) and \( S_1 \) (otherwise we would take \( S' \) rather than \( S \))
2. There exists a minimum vertex separator \( S_2 \) of \( \alpha \) such that \( S' \) is not a minimum vertex separator of \( U_a \) and \( S_2 \) (as otherwise \( S_1 \) would be the leftmost minimum vertex separator)

Now let \( S_L \) and \( S_R \) be the vertex separators of \( \alpha \) obtained by applying Lemma A.1 to \( S_1 \) and \( S_2 \). Since \( S_1 \) and \( S_2 \) are minimum vertex separators of \( \alpha \), we must have that \( w(S_L) = w(S_R) = w(S_1) = w(S_2) \). Since \( S_L \) is a vertex separator of \( U_a \) and \( S_2 \), \( S_L \neq S_1 \). However, \( S_L \) is a vertex separator of \( U_a \) and \( S_1 \), which contradicts our choice of \( S_1 \).

Thus, there must be a leftmost minimum vertex separator of \( \alpha \). Following similar logic, there must be a rightmost minimum vertex separator of \( \alpha \) as well.

**B Proofs with Canonical Maps**

In this section, we give alternative proofs of Lemmas 7.78 and 8.14 using canonical maps.

**Definition B.1** (Canonical Maps). For each shape \( \alpha \) and each ribbon \( R \) of shape \( \alpha \), we arbitrarily choose a canonical map \( \varphi_R : V(\alpha) \to V(R) \) such that \( \varphi_R(H_\alpha) = H_R \), \( \varphi_R(U_\alpha) = A_R \), and \( \varphi_R(V_\alpha) = B_R \). Note that there are \( |\text{Aut}(\alpha)| \) possible choices for this map.
B.1 Proof of Lemma 7.78

Lemma B.2.

\[ M^{\sigma_0}_{\tau}(H) = \sum_{\sigma \in \text{Row}(H), \sigma' \in \text{Col}(H)} H(\sigma, \sigma') |\text{Decomp}(\sigma, \tau, \sigma^T)| M_{\sigma \circ \sigma' \circ \tau} \]

Proof. Observe that there is a bijection between ribbons \( R \) with shape \( \sigma \circ \tau \circ \sigma^T \) together with an element \( \pi \in \text{Decomp}(\sigma, \tau, \sigma') \) and triples of ribbons \( (R_1, R_2, R_3) \) such that

1. \( R_1, R_2, R_3 \) have shapes \( \sigma, \tau, \) and \( \sigma^T \), respectively.
2. \( V(R_1) \cap V(R_2) = A_{R_2} = B_{R_1}, V(R_2) \cap V(R_3) = A_{R_3} = B_{R_2}, \) and \( V(R_1) \cap V(R_3) = A_{R_2} \cap B_{R_2} \)

To see this, note that given such ribbons \( R_1, R_2, R_3 \), the ribbon \( R = R_1 \circ R_2 \circ R_3 \) has shape \( \sigma \circ \tau \circ \sigma^T \). Further note that we have two bijective maps from \( V(\sigma \circ \tau \circ \sigma^T) \) to \( V(R) \). The first map is \( \varphi_R \). The second map is \( \varphi_{R_1} \circ \varphi_{R_2} \circ \varphi_{R_3} \). Using this, we can take \( \pi = \varphi_R^{-1}(\varphi_{R_1} \circ \varphi_{R_2} \circ \varphi_{R_3}) \).

Conversely, given a ribbon \( R \) of shape \( \sigma \circ \tau \circ \sigma^T \) and an element \( \pi \in \text{Decomp}(\sigma, \tau, \sigma') \), let \( R_1 = \varphi_R(\pi(\sigma)) \), \( R_2 = \varphi_R(\pi(\tau)) \), and \( R_3 = \varphi_R(\pi(\sigma')) \). Note that this is well defined because for any element \( \pi' \in \text{Aut}(\sigma) \times \text{Aut}(\tau) \times \text{Aut}(\sigma^T) \), \( \varphi_R(\pi(\sigma')) = \varphi_R(\pi(\sigma')) \). Similarly, \( \varphi_R(\pi(\tau')) = \varphi_R(\pi(\tau')) \) and \( \varphi_R(\pi(\sigma')) = \varphi_R(\pi(\sigma')) \).

To confirm that this is bijection, we have to show that these two maps are inverses of each other. Given \( R_1, R_2, \) and \( R_3 \), applying these two maps gives us ribbons \( R'_1 = \varphi_R^{-1}(\varphi_{R_1} \circ \varphi_{R_2} \circ \varphi_{R_3})(H_{\sigma}) = R_1, \)
\( R'_2 = \varphi_R^{-1}(\varphi_{R_1} \circ \varphi_{R_2} \circ \varphi_{R_3})(H_{\tau}) = R_2, \) and \( R'_3 = \varphi_R^{-1}(\varphi_{R_1} \circ \varphi_{R_2} \circ \varphi_{R_3})(H_{\sigma^T}) = R_3 \). Conversely, given \( R \) and an element \( \pi \in \text{Decomp}(\sigma, \tau, \sigma') \) (which we represent by an element \( \pi \in \text{Aut}(\sigma \circ \tau \circ \sigma^T) \)), applying these two maps gives us the ribbon \( R' = \varphi_R(\pi(\sigma)) \circ \varphi_R(\pi(\tau)) \circ \varphi_R(\pi(\sigma^T)) = \varphi_R\pi(\sigma \circ \tau \circ \sigma^T) = R \)

and gives us the map \( \varphi_R^{-1}(\varphi_{\varphi_R(\pi(\sigma))} \circ \varphi_{\varphi_R(\pi(\tau))} \circ \varphi_{\varphi_R(\pi(\sigma^T))}) \).

Now observe that both \( \varphi_R \pi \) and \( \varphi_{\varphi_R(\pi(\sigma))} \) give bijective maps from \( \sigma \) to the ribbon \( \varphi_R \pi(\sigma) \) so \( \varphi_R^{-1}(\varphi_{\varphi_R(\pi(\sigma))}) \varphi_R \pi \in \text{Aut}(\sigma) \). Following similar logic for \( \tau \) and \( \sigma^T \), in \( \text{Decomp}(\sigma, \tau, \sigma') \) this map is equivalent to \( \varphi_R^{-1}(\varphi_R \pi) = \pi \).

B.2 Proof of Lemma 8.14

Definition B.3 (Rigorous definition of intersection patterns). We define an intersection pattern \( P \) on composable shapes \( \gamma, \tau, \gamma^T \) to consist of the shape \( \gamma \circ \tau \circ \gamma^T \) together with a non-empty set of constraint edges \( E(P) \) on \( V(\gamma \circ \tau \circ \gamma^T) \) such that:

1. For all vertices \( u, v, w \in V(\gamma \circ \tau \circ \gamma^T) \), if \((u, v), (v, w) \in E(P)\) then \((u, w) \in E(P)\)
2. \( E(P) \) does not contain a path between two vertices of \( \gamma \), two vertices of \( \tau \), or two vertices of \( \gamma^T \). This ensures that when we consider \( \gamma, \tau, \gamma^T \) individually, their vertices are distinct.

130
3. Defining \( V_s(\gamma) \subseteq V(\gamma) \) to be the vertices of \( \gamma \) which are incident to an edge in \( E(P) \), \( U_\gamma \) is the unique minimum-weight vertex separator between \( U_\gamma \) and \( V_s(\gamma) \cup V_\gamma \).

4. Similarly, defining \( V_s(\gamma^T) \subseteq V(\gamma^T) \) to be the vertices of \( \gamma^T \) which are incident to an edge in \( E(P) \), \( V_{\gamma^T} \) is the unique minimum-weight vertex separator between \( V_s(\gamma^T) \cup U_{\gamma^T} \) and \( V_{U_{\gamma^T}} \).

5.* All edges in \( E(P) \) are between vertices of the same type.

**Definition B.4.** We say that two intersection patterns \( P, P' \) on shapes \( \gamma, \tau, \gamma^T \) are equivalent (which we write as \( P \equiv P' \)) if there is an automorphism \( \pi \in \text{Aut}(\gamma) \times \text{Aut}(\tau) \times \text{Aut}(\gamma^T) \) such that \( \pi(P) = P' \) (i.e. if \( E(P) \) and \( E(P') \) are the constraint edges for \( P \) and \( P' \) respectively then \( \pi(E(P)) = E(P') \)).

**Definition B.5.** Given composable shapes \( \gamma, \tau, \gamma^T \), we define \( \mathcal{P}_{\gamma, \tau, \gamma^T} \) to be the set of all possible intersection patterns \( P \) on \( \gamma, \tau, \gamma^T \) (up to equivalence).

**Definition B.6.** Given composable (but not properly composable) ribbons \( R_1, R_2, R_3 \) of shapes \( \gamma, \tau, \gamma', \) we define the intersection pattern \( P \in \mathcal{P}_{\gamma, \tau, \gamma'} \) induced by \( R_1, R_2, R_3 \) as follows:

1. Take the canonical maps \( \varphi_{R_1} : V(\gamma) \to V(R_1), \varphi_{R_2} : V(\tau) \to V(R_2), \) and \( \varphi_{R_3} : V(\gamma^T) \to V(R_3) \).
2. Given vertices \( u \in V(\gamma) \) and \( v \in V(\tau) \), add a constraint edge between \( u \) and \( v \) if and only if \( \varphi_{R_1}(u) = \varphi_{R_2}(v) \). Similarly, given vertices \( u \in V(\gamma) \) and \( w \in V(\gamma^T) \), add a constraint edge between \( u \) and \( w \) if and only if \( \varphi_{R_3}(u) = \varphi_{R_3}(w) \) and given vertices \( v \in V(\tau) \) and \( w \in V(\gamma^T) \), add a constraint edge between \( v \) and \( w \) if and only if \( \varphi_{R_2}(v) = \varphi_{R_3}(w) \).

**Definition B.7.** Given an intersection pattern \( P \in \mathcal{P}_{\gamma, \tau, \gamma^T} \), we define \( V(\gamma \circ \tau \circ \gamma^T) / E(P) \) to be \( V(\gamma \circ \tau \circ \gamma^T) \) where all of the edges in \( E(P) \) are contracted (i.e. if \( (u, v) \in E(P) \) then \( u = v \) and \( u = v \) only appears once).

**Definition B.8.** Given an intersection pattern \( P \in \mathcal{P}_{\gamma, \tau, \gamma^T} \), we define \( \tau_P \) to be the shape such that:

1. \( V(H_{\tau_P}) = V(\gamma \circ \tau \circ \gamma^T) / E(P) \)
2. \( E(H_{\tau_P}) = E(\gamma) \cup E(\tau) \cup E(\gamma^T) \)
3. \( U_{\tau_P} = U_\gamma \)
4. \( V_{\tau_P} = V_{\gamma^T} \)

**Definition B.9.** Given an intersection pattern \( P \in \mathcal{P}_{\gamma, \tau, \gamma^T} \), we make the following definitions:

1. We define \( \text{Aut}(P) = \{ \pi \in \text{Aut}(\gamma \circ \tau \circ \gamma^T) : \pi(E(P)) = E(P) \} \)
2. We define \( \text{Aut}_{\text{pieces}}(P) = \{ \pi \in \text{Aut}(U_\gamma) \times \text{Aut}(\tau) \times \text{Aut}(\gamma^T) : \pi(E(P)) = E(P) \} \)
3. We define \( N(P) = |\text{Aut}(P) / \text{Aut}_{\text{pieces}}(P)| \)

131
Lemma B.10. For all composable $\sigma$, $\tau$, and $\sigma T'$ (including improper $\tau$),

$$M^\text{fact}_T(e_\sigma e_{T'}^T) - M^\text{orth}_T(e_\sigma e_{T'}^T) = \sum_{\gamma \in \text{non-trivial}} \frac{1}{|\text{Aut}(U_\gamma)|} \sum_{p \in P_{\gamma,\tau',\gamma T'}} N(P)M^\text{orth}_{T_p}(e_\sigma e_{T'}^T)$$

$$+ \sum_{\gamma \in \text{non-trivial}} \frac{1}{|\text{Aut}(U_\gamma)|} \sum_{p \in P_{\gamma,\tau',\gamma T'}} N(P)M^\text{orth}_{T_p}(e_\sigma e_{T'}^T)$$

$$+ \sum_{\gamma \in \text{non-trivial}} \frac{1}{|\text{Aut}(U_\gamma)| \cdot |\text{Aut}(U_{T'})|} \sum_{p \in P_{\gamma,\tau',\gamma T'}} N(P)M^\text{orth}_{T_p}(e_\sigma e_{T'}^T)$$

Proof. This lemma follows from the following bijection. Consider the third term

$$\sum_{\gamma \in \text{non-trivial}} \frac{1}{|\text{Aut}(U_\gamma)| \cdot |\text{Aut}(U_{T'})|} \sum_{p \in P_{\gamma,\tau',\gamma T'}} N(P)M^\text{orth}_{T_p}(e_\sigma e_{T'}^T)$$

On one side, we have the following data:

1. Ribbons $R_1$, $R_2$, and $R_3$ such that
   a. $R_1$, $R_2$, $R_3$ have shapes $\sigma$, $\tau$, and $\sigma T'$, respectively.
   b. $A_{R_2} = B_{R_2}$ and $A_{R_3} = B_{R_3}$
   c. $(V(R_1) \cup V(R_2)) \cap V(R_3) \neq A_{R_3}$ and $(V(R_2) \cup V(R_3)) \cap V(R_1) \neq B_{R_1}$

2. An ordering $O_\sigma'$ on the leftmost minimum vertex separator $S'$ between $A_{R_3}$ and $V_* \cup B_{R_1}$.

3. An ordering $O_{T'}$ on the rightmost minimum vertex separator $S'$ between $V_* \cup A_{R_3}$ and $B_{R_3}$.

On the other side, we have the following data:

1. An intersection pattern $P \in \mathcal{P}_{\gamma,\tau,\gamma T'}$ where $\gamma$ and $\gamma T'$ are non-trivial.

2. Ribbons $R_1', R_2', R_3'$ of shapes $\sigma_2, \tau_2, \sigma_2 T'$ such that $V(R_1') \cap V(R_2') = A_{R_2} = B_{R_2}$, $V(R_2') \cap V(R_3') = B_{R_2} = B_{R_3}$, and $V(R_1') \cap V(R_3') = A_{R_2} \cap B_{R_2}$

3. An element $\pi \in \text{Aut}(P) / \text{Aut}_\text{pieces}(P)$

To see this bijection, given $R_1, R_2, R_3$, we again implement our strategy for analyzing intersection terms. Recall that $V_*$ is the set of vertices in $V(R_1) \cup V(R_2) \cup V(R_3)$ which have an unexpected equality with another vertex, $S'$ is the leftmost minimum vertex separator between $A_{R_3}$ and $B_{R_1} \cup V_*$, and $T'$ is the rightmost minimum vertex separator between $A_{R_3} \cup V_*$ and $B_{R_3}$.

1. Decompose $R_1$ as $R_1 = R_1' \circ R_4$ where $R_1'$ is the part of $R_1$ between $A_{R_4}$ and $(S', O_{S'})$ and $R_4$ is the part of $R_1$ between $(S', O_{S'})$ and $B_{R_4} = A_{R_4}$. Decompose $R_3$ as $R_3' \cup R_3''$ where $R_5$ is the part of $R_3$ between $A_{R_3}$ and $(T', O_{T'})$ and $R_3''$ is the part of $R_3$ between $(T', O_{T'})$ and $B_{R_3}$

2. Take the intersection pattern $P$ and the ribbon $R_2'$ induced by $R_4$, $R_2$, and $R_5$.

3. Observe that we have two bijective maps from $V(\gamma \circ \tau \circ \gamma T') / E(P)$ to $V(R_4) \cup V(R_2) \cup V(R_3)$. The first map is $\varphi_{R_4} \circ \varphi_{R_2} \circ \varphi_{R_5}$ and the second map is $\varphi_{R_2}^{-1}$. We take $\pi = \varphi_{R_2}^{-1}(\varphi_{R_4} \circ \varphi_{R_2} \circ \varphi_{R_5})$. 

132
Conversely, given an intersection pattern \( P \in \mathcal{P}_{\gamma, \tau, \gamma^T} \), \( R_1', R_2', R_3' \), and an element \( \pi \in \text{Aut}(P) / \text{Aut}_{\text{pieces}}(P) \):

1. Take \( R_4 = \varphi_{R_2'} \pi(V(\gamma)) \), \( R_2 = \varphi_{R_2'} \pi(V(\tau)) \), and \( R_5 = \varphi_{R_2'} \pi(V(\gamma^T)) \).
2. Take \( R_1 = R_1' \cup R_4 \) and take \( R_3 = R_5 \cup R_3' \).
3. Take \( O_S \) and \( O_T \) based on \( B_{R_1'} = A_{R_4} \) and \( B_{R_5} = A_{R_3'} \).

To confirm that this is a bijection, we need to show that these maps are inverses of each other.

If we apply the first map and then the second, we obtain the following:

1. We obtain the ribbons
   
   (a) \( R_1'' = R_1' \circ \varphi_{R_2'} \pi(V(\gamma)) \)
   
   (b) \( R_2'' = \varphi_{R_2'} \pi(V(\tau)) \)
   
   (c) \( R_3'' = \varphi_{R_2'} \pi(V(\gamma^T)) \circ R_3' \)

   where
   
   (a) \( R_1' \) is the part of \( R_1 \) between \( A_{R_4} \) and \( (S', O_{S'}) \) where \( S' \) is the minimum vertex separator between \( A_{R_4} \) and \( V_s \cup B_{R_1} \).
   
   (b) \( R_4 \) is the part of \( R_1 \) between \( (S', O_{S'}) \) and \( B_{R_1} \).
   
   (c) \( R_2' \) is the ribbon of shape \( \tau_P \) induced (along with the intersection pattern \( P \)) by \( R_1, R_2, \) and \( R_3 \).
   
   (d) \( R_3 \) is the part of \( R_3 \) between \( A_{R_3} \) and \( (T', O_{T'}) \).
   
   (e) \( R_3' \) is the part of \( R_3 \) between \( (T', O_{T'}) \) and \( B_{R_3} \).

   This implies that \( R_1'' = R_1' \circ R_4 = R_1, R_2'' = R_2, \) and \( R_3'' = R_3 \circ R_3' = R_3 \). Since the second map leaves \( R_1' \) and \( R_3' \) unchanged, we recover the orderings \( O_S \) and \( O_T \) as well.

Conversely, if we apply the second map, we have that \( R_1 = R_1' \circ \varphi_{R_2'} \pi(V(\gamma)), R_2 = \varphi_{R_2'} \pi(V(\tau)), \) and \( R_3 = \varphi_{R_2'} \pi(V(\gamma^T)) \circ R_3' \) and we have the orderings \( O_S \) and \( O_T \) corresponding to \( B_{R_1'} \) and \( A_{R_3'} \) respectively. If we apply the first map:

1. \( R_1' \) and \( R_3' \) are preserved.
2. \( R_2'' \) and \( P'' \) are the ribbon and intersection pattern induced by the ribbons \( \varphi_{R_2'} \pi(\gamma), \varphi_{R_2'} \pi(\tau), \) and \( \varphi_{R_2'} \pi(\gamma^T) \). To see that \( R_2'' = R_2' \), observe that

   \[
   R_2'' = \varphi_{R_2'} \pi(V(\gamma)) \circ \varphi_{R_2'} \pi(V(\tau)) \circ \varphi_{R_2'} \pi(V(\gamma^T)) = \varphi_{R_2'} \pi(\gamma \circ \tau \circ \gamma^T) = \varphi_{R_2'} \pi(\gamma^T) = R_2'
   \]

   To see that \( P'' = P \), observe that:

   (a) We have two bijective maps from \( V(\gamma) \) to \( V(\varphi_{R_2'} \pi(\gamma)) \). These two maps are \( \varphi_{R_2'} \pi \) and \( \varphi_{R_2'} \pi(\gamma) \).

   (b) We have two bijective maps from \( V(\tau) \) to \( V(\varphi_{R_2'} \pi(\tau)) \). These two maps are \( \varphi_{R_2'} \pi \) and \( \varphi_{R_2'} \pi(\tau) \).

   (c) We have two bijective maps from \( V(\gamma^T) \) to \( V(\varphi_{R_2'} \pi(\gamma^T)) \). These two maps are \( \varphi_{R_2'} \pi \) and \( \varphi_{R_2'} \pi(\gamma^T) \).

   (d) For \( P'' \), the constraint edges are

   \[
   \left( \varphi_{\varphi_{R_2} \pi(\gamma)}^{-1} \varphi_{R_2} \pi \circ \varphi_{\varphi_{R_2} \pi(\tau)}^{-1} \varphi_{R_2} \pi \circ \varphi_{\varphi_{R_2} \pi(\gamma^T)}^{-1} \varphi_{R_2} \pi \right)(E(P))
   \]

133
3. We have that

\[ \pi'' = \varphi_{R_2^*}^{-1}(\varphi_{R_2^*}\pi(V(\gamma)) \circ \varphi_{R_2^*}\pi(V(\tau)) \circ \varphi_{R_2^*}\pi(V(\gamma^T))) \]

To see that \( \pi'' \equiv \pi \), note that

\[ \pi = \pi'' \left( \varphi_{R_2^*}^{-1}(\varphi_{R_2^*}\pi(V(\gamma)) \circ \varphi_{R_2^*}\pi(V(\tau)) \circ \varphi_{R_2^*}\pi(V(\gamma^T))) \varphi_{R_2^*}\pi \right) \]

The analysis for the first term is the same except that when \( \gamma' \) is trivial, we always take \( \gamma' \) to be the identity so \( T = V(V_\tau) = V(U_{\omega'\tau}) \) and the ordering \( O_T \) is given by \( V_\tau = U_{\omega'\tau} \). Similarly, the analysis for the second term is the same except that when \( \gamma \) is trivial, we always take \( \gamma \) to be the identity so \( S = V(V_\sigma) = V(U_\tau) \) and the ordering \( O_S \) is given by \( V_\sigma = U_\tau \).