THE UNIVERSITY OF CHICAGO

NONLINEAR RANDOM MATRICES AND APPLICATIONS TO THE SUM OF SQUARES HIERARCHY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTORATE OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

BY
GOUTHAM RAJENDRAN

CHICAGO, ILLINOIS

Dedication Text

Epigraph Text

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

ABSTRACT

We develop new tools in the theory of nonlinear random matrices and apply them to study the performance of the Sum-of-Squares (SoS) hierarchy on average-case problems.

The SoS hierarchy is a powerful optimization technique that has achieved tremendous success for various problems in combinatorial optimization, robust statistics and machine learning. It's a family of convex relaxations that lets us smoothly tradeoff running time for approximation guarantees. In recent works, it's been shown to be extremely useful to recover structure in high dimensional noisy data. It also remains our best approach towards refuting the notorious Unique Games Conjecture.

In this work, we analyze the performance of the SoS hierarchy on fundamental problems stemming from statistics, theoretical computer science and statistical physics. In particular, we show subexponential-time SoS lower bounds for the problems of the Sherrington-Kirkpatrick Hamiltonian, Planted Slightly Denser Subgraph, Tensor Principal Components Analysis and Sparse Principal Components Analysis. These SoS lower bounds involve analyzing large random matrices, wherein lie our main contributions. These results offer strong evidence for the truth of and insight into the low-degree likelihood ratio hypothesis, an important conjecture that predicts the power of bounded-time algorithms for hypothesis testing.

We also develop general-purpose tools for analyzing the behavior of random matrices which are functions of independent random variables. Towards this, we build on and generalize the matrix variant of the Efron-Stein inequalities. In particular, our general theorem on matrix concentration recovers various results that have appeared in the literature. We expect these random matrix theory ideas to have other significant applications.

CHAPTER 1

INTRODUCTION

Algorithm design, mathematical optimization and computational complexity are close-knit fields of computer science that have largely developed in parallel in the beginning. In recent decades, there has been an explosion of research in these fields that often borrowed ideas from the other ones, and there is no longer a discernible wall separating them. Indeed, these fields of computer science can now be construed as trying to achieve the same goal - Which problems are easy and which are hard?

Early researchers have mainly focused on search problems. Given an input, the objective is to search for a desired hidden structure. Often, this can be equivalently restated as the problem of optimizing an appropriate objective function under various constraints.

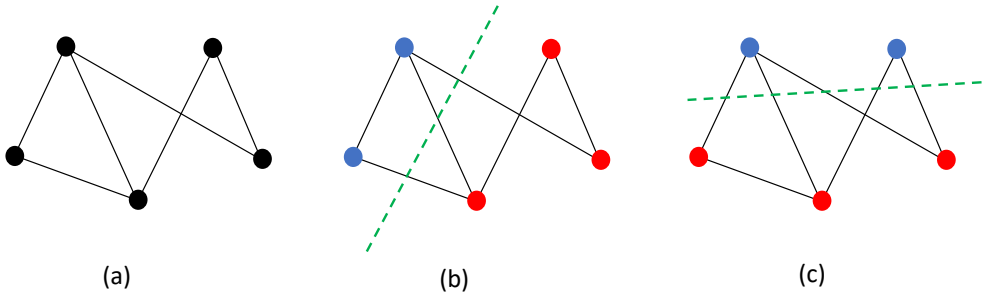

Figure 1.1: An example graph on 5 vertices and two possible cuts.

For example, consider the *Maximum Cut* problem, where the input is a graph and the goal is to partition the set of vertices into two subsets that maximizes the number of edges with endpoints in different parts. If we take for instance the graph (a) in Fig. 1.1, two possible partitions are shown in (b) and (c) where blue colored vertices form a part and red colored vertices form a part. Then, the partition in (b) cuts 3 edges and the partition in (c) cuts 5 edges, namely the edges intersecting the green line. It's easy to see via a simple parity argument that we cannot do better than 5 edges.

In the search problem formulation, we would like our search algorithm to output a partition that cuts the maximum possible number of edges. And in the optimization problem formulation, we would like our optimization algorithm to output the maximum value correctly.

Another formulation of computational problems are decision problems. Given an input, the objective is to decide whether there exists a hidden structure or if the objective value satisfies some properties, with the restriction that the algorithm can only return a boolean output - for example, true or false; or yes or no. In the above example of maximum cut, the decision problem perspective could be to ask if the maximum cut in the given graph contains at least 0.6 (say) fraction of the total number of edges.

These types of problems are all intimately related and in many cases, essentially boil down to the search for algorithms. For practicality, we require various properties like efficiency, accuracy, etc. This has led to the development of a rich theory of computability, complexity theory and optimization. In this dissertation, we will also consider the viewpoints of related types of problems, namely certification problems and hypothesis testing. As we will see, these other formulations are related to the former and to each other but it's not clear how deep the connections go, and trying to understand this is an important pursuit in theoretical computer science. That said, underlying all these formulations is the goal of searching for efficient algorithms to detect and extract structure from data, or arguing that no such exist unless we're willing to compromise on other things like efficiency or accuracy.

1.1 Certification problems

As opposed to search or decision problems, certification problems, given an input, ask for a bound on the objective value, that holds true with probability 1. And the quality of the algorithm is usually measured in terms of how close the bound gets to the true optimum.

In the running example of maximum cut, given a graph, the task could be to output a

value that's always an upper bound on the size of the maximum cut. A simple algorithm could be to simply return the total number of edges in the graph. Indeed, this is a valid certification algorithm and we could ask if one could do better.

This is fundamentally a different approach to algorithm design. Consider the scenario when we are maximizing some objective function and so we desire an upper bound on the optimal value. Then, designing a certification algorithm can be construed as attacking a problem from *above* as opposed to from *below*, the latter of which is the more standard notion of algorithm design.

The notion of linear programming relaxations already provide such certification algorithms. Given a problem that can be formulated as an integer program (as many are), a natural way to obtain a certification solution is to widen the search space from integral variables to real variables, adding other appropriate constraints as necessary. This is known as relaxing the program. This enables a faster algorithm to attempt to compute the solution, but comes at a loss of only obtaining an approximate solution. But importantly, the objective value obtained by the return solution is a definite bound bound on the optimal solution, no matter the input. This is what a certification algorithm desires. Measuring the quality of the returned output often depends on the type of relaxation considered and problem specific structure.

In many cases, it's possible to obtain an approximation algorithm to a problem by looking at a relaxation of the program, obtaining a non-integral solution and rounding it to a valid solution. For the maximum cut problem, this was done by Goemans and Williamson in their seminal work [72] where they used a semidefinite programming relaxation, which is more powerful than linear programming relaxations.

In this dissertation, we will focus on a specific class of such certification algorithms, namely the Sum-of-Squares (SoS) hierarchy, sometimes referred to as the Lasserre hierarchy. The SoS hierarchy is a series of convex relaxations to a given program. By virtue of being a

relaxation, they can be used for certification. Due to it's tremendous success for various fundamental optimization problems such as maximum cut, constraint satisfaction, etc., the SoS hierarchy has become a powerful optimization technique. This is further amplified by results that say that the SoS hierarchy is the optimal relaxation among a broad class of semidefinite programming relaxations [119], and assuming the famous unique games conjecture, it's the best approximation algorithm for every constraint satisfaction problem [152]. A chief goal of this dissertation is to understand the limits of this powerful technique. We especially focus on the so-called average-case setting, that we will define now.

1.2 Average-case analysis

An important theme in this work is the study of random instances of problems, which is termed average-case analysis. As opposed to traditional worst-case algorithm design, where we wish to design an algorithm that performs well on the worst possible input, there has been an exciting development of research on problems where the input is randomly sampled from a distribution. For instance, in the maximum cut problem, we could assume that the input comes from the Erdős-Rényi family of random graphs, where the number of vertices in the graph is chosen beforehand and each edge is present independently with probability 0.5.

In average-case algorithm design, we wish to design algorithms that perform well on average-case inputs with high probability, as opposed to all inputs. This is important because studying the worst case complexity of a problem may not shed light on the intrinsic hardness of the problem. This happens because the worst-case instance input for an algorithm could be highly artificial and contrived. Put another way, in real world scenarios, the inputs for various optimization or search problems we encounter are unlikely to be such instances. This is seen in practice as well. For example, the simplex method for linear programming [47] is exponentially slow in the worst-case, as was shown by Klee and Minty [106], but

performs extremely well practically. Various works have tried to explain this behavior, e.g. [26, 170, 27, 171]

Tremendous effort has been invested to understand the average-case complexity for a wide variety of problems. And research towards designing average-case algorithms brings about a deeper understanding of the core of the problem, enabling the design of worst-case algorithms as well. This can be seen for example for the famous Densest k-subgraph problem [22]. In this work, we will focus on average-case analysis.

In our pursuit, fundamental mathematical objectS that occur repeatedly are large random matrices. And we often desire to understand their behavior.

1.3 Underlying theme of this work: Random matrices

Random matrices are abundant in computer science, especially in the fields of optimization and statistics. Often, the analysis of an algorithm requires analyzing the behavior of certain random matrices that can be constructed from the input. Even outside computer science, random matrix theory is a fundamental field of it's own right, having been studied since the early 1900s, with applications also extending to many branches of mathematics and physics. For a short survey, see [65].

There has been tremendous effort over the last few decades to develop the theory of random matrices, see the book by Tropp [180]. For example, the matrix-Bernstein inequality studies the behavior of a random weighted sum of matrices; the Wigner semicircle law studies the distribution of the eigenvalues of a random matrix sampled from the Gaussian Orthogonal ensemble. On the other hand, fewer tools are available to understand the behavior of nonlinear random matrices, where each matrix entry is a nonlinear function of the input, say for instance low-degree polynomials.

In our setting, this occurs frequently when trying to analyze the SoS hierarchy for various problems. This is true both when trying to design algorithms via SoS as well as when trying

to study the limitations of SoS algorithms, for example, [11, 88, 165, 131, 95]. Therefore, we begin with this important endeavor of understanding the behavior of nonlinear random matrices. In the first part of this thesis, we are interested specifically in concentration behavior. We emphasize that this is an important research direction in it's own right.

To bound the fluctuations of a random matrix from its mean, measured in terms of spectral or Schatten t-norm of the difference, a simple but powerful technique that has been widely used (including in many of the works cited above) is the so-called trace method. In this method, the (centered) random matrix is raised to a large power and the expected trace of the resulting matrix is bounded. While this method gives satisfactory results, it often requires ingenious observations and highly nontrivial combinatorics.

Another approach is as follows. Consider a random matrix that is a function of several independent input variables. We can study it's behavior by studying how much it deviates when a single uniformly chosen input entry is resampled. By bounding these local fluctuations, we can bound the global fluctuation of the random matrix. This technique gives rise to the Efron-Stein inequalities. Originally, they were developed for scalar random variables (which can be thought of as 1×1 matrix). And in this special case, they turned out to be extremely powerful since they have been shown to recover many standard concentration inequalities. And recently, the work [147] showed a matrix version of the Efron-Stein inequalities. In this work, we build on this to obtain a general framework for proving concentration of large random matrices.

In the second part of this thesis, in the analysis of SoS algorithms, the fundamental difficulty that appears is to analyze the behavior of a large nonlinear random matrix. In particular, we want to argue that this random matrix is positive semidefinite with high probability over the choice of the input. For this, we exhibit an approximate Cholesky decomposition of the matrix and the proof extensively builds on the concentration results we develop above.

In conclusion, the motif in this work is the study of nonlinear random matrices, where we both build a general framework for analyzing concentration and apply them to study algorithms on fundamental problems.

1.4 The Sum of Squares Hierarchy

Given an optimization problem in the form of a program with polynomial inequality constraints, there have been many works proposing generic approaches to relax the program, in order to obtain good solutions efficiently. Some of the more dominant approaches have been the Lovász-Schrijver hierarchy [120] and the Sherali-Adams hierarchy [167]. Informally speaking, these hierarchies of algorithms lift the program to a larger set of variables, tied together via various constraints, relax and solve the larger program, and finally project the solution down to the original variable space. They are parameterized by an integer known as the degree, where larger degrees offer tighter relaxations at the cost of larger running times.

The Sum-of-Squares (SoS) hierarchy is a similar optimization technique that harnesses the power of semidefinite programming. For polynomial optimization problems, the SoS hierarchy, first independently investigated by Shor [169], Nesterov [135], Parillo [144], Lasserre [117] and Grigoriev [73, 74], 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 which are solvable in $n^{O(d)}$ time. This has paved the way for the SoS hierarchy to be almost a blackbox tool for algorithm design. As has been shown in multiple works, it serves as a strong algorithm for various problems, both in the worst case and the average case settings.

Consider our running example of the Maximum Cut problem. The seminal Goemans-Williamson algorithm [71] achieves an approximation factor of ≈ 0.878 for this problem via a semidefinite programming relaxation. As it turns out, this algorithm is just the degree 2 SoS hierarchy. This approximation factor is conjectured to be optimal and there has been

increasing evidence that this is indeed the case. This highlights an example of why the SoS hierarchy is powerful.

Indeed, there has been tremendous success in using the SoS hierarchy to obtain efficient algorithms for combinatorial optimization problems (e.g., [72, 6, 76, 153]) as well as problems stemming from Statistics and Machine Learning (e.g., [11, 15, 89, 150, 110]). In fact, SoS achieves the state-of-the-art approximation guarantees for many fundamental problems such as Sparsest Cut [6], Maximum Cut [72], Tensor PCA [89] and all Max-k-CSPs [152]. As mentioned earlier, for a large class of problems, it's been shown that SoS relaxations are the most efficient among all semidefinite programming relaxations [119].

The term "Sum of Squares" comes from a dual view in proof complexity. Besides being an algorithmic technique, SoS can be equivalently viewed as giving a proof or certificate of a bound on the optimal value of a polynomial optimization problem. This work can be traced back to Hilbert's seventeeth problem which has led to work on a proof complexity result known as the Positivstellensatz, which gives conditions under which polynomial systems can be shown to have no solutions, see e.g. [174, 151, 159]. The algorithmic implications were originally observed by Lasserre [117] and Parillo [144, 145] leading to the interpretation of SoS as an optimization technique as we study in this work. This duality can be completely formalized and has led to the so-called framework of "proofs to algorithms" that has achieved tremendous success, especially recently in robust statistics, see e.g., [110, 96, 82, 8]. The adage is that if we can find an "easy" proof of an identifiability result for a search problem, then it can be automatized to give an algorithm. We will not explore this in detail here, and we refer the reader to the monograph [64].

Next, we move onto SoS lower bounds but before that, we highlight some related techniques that has gained traction in the community recently.

1.4.1 Related Algorithmic Techniques

Apart from search, decision and certification, researchers have also considered other related types of problems. Consider a problem where the input is sampled from one of two known distributions and we would like to identify which distribution it was sampled from. This is known generally as hypothesis testing. For example, one distribution could be the distribution of Erdős-Rényi random graphs while the other could be the distribution of Erdős-Rényi random graphs but with a large cut planted in them. It's clear that this problem is a different flavor of the maximum cut problem on random graphs. Beyond being interesting in their own right, studying these related formulations offer alternate perspectives and interesting insights into the search or certification variants as well. Another type of problem, known as recovery problems, is to recover the planted structure when the input is sampled from the latter distribution.

For all the type of problems considered so far, apart from SoS, there have also been several other framework of algorithms that have been considered and in some cases, extensively studied. Examples include

- Lovász-Schrijver and Sherali-Adams hierarchies - As discussed earlier, these hierarchies lift a program to a larger set of variables and then relax any integrality constraints. The resulting solution is then projected back to the original variables which may then be rounded to an integral solution. These hierarchies are captured by the SoS hierarchy, or in other words, the SoS hierarchy is at least as powerful as these hierarchies [64].

- Low degree polynomials - For hypothesis testing, low degree polynomials can be used to try and distinguish the two distributions. More precisely, if there is a low degree polynomial such that its expected value on the two distributions behave differently and the variance isn't too large, this can be used to distinguish the two distributions. This is related to the SoS hierarchy and we will revisit this point in more detail later.

- Statistical query algorithms - For hypothesis testing, the statistical query model (SQ) is another popular restricted class of algorithms introduced by [99]. In this model, for an underlying distribution, we can access it indirectly by querying expected values of functions, upto some error. Given access to this oracle, we would like to hypothesis test. SQ algorithms capture a broad class of algorithmic techniques in statistics and machine learning including spectral methods, moment and tensor methods (see e.g. [61, 62]). SQ algorithms has also been used to study information-computation tradeoffs and more broadly has been studied in other contexts [60]. There has also been significant work trying to understand the limits of SQ algorithms (e.g. [61, 63, 51]). Recent work [32] has shown that low degree polynomials and statistical query algorithms have equivalent power under mild conditions.

- Approximate message passing and other statistical physics techniques such as belief propagation, see e.g. the review [188].

- Local algorithms, see e.g. [55, 56, 92].

- Circuit models of computation of bounded size, see e.g. [161, 162].

1.5 Lower bounds against The Sum of Squares Hierarchy

Because of the incredible success of the SoS hierarchy for a variety of problems, it's an important research direction to study the limits of the SoS hierarchy, which we endeavour in this dissertation. In particular, we will focus on average-case problems and as we will see, most of the technical difficulty boils down to the analysis of nonlinear random matrices, to handle which we develop various techniques.

There are many reasons for why studying lower bounds against the SoS hierarchy is important. The SoS hierarchy is general enough to capture a broad class of algorithmic reasoning [64]. In particular, SoS captures the Lovász-Schrijver and Sherali-Adams hierarchies,

statistical query algorithms and algorithms based on low degree polynomials. Therefore, SoS lower bounds indicate to the algorithm designer the intrinsic hardness of the problem and suggest 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 techniques such as SoS have served as strong evidence of computational hardness [85, 91]. Thus, understanding the power of the SoS hierarchy on these problems is an important step towards understanding the approximability of these problems. See also the surveys [17, 130] for more on this.

There have been relatively fewer works on SoS lower bounds, as opposed to some other classes of algorithms we have discussed, which can be attributed to the sheer technical difficulty of proving such lower bounds. For example, the works [74, 164, 107] studied SoS lower bounds for random constraint satisfaction problems. A series of works [57, 127, 49, 13, 140] studied SoS lower bounds for maximum clique on random graphs. Some other SoS lower bounds, not including the ones in this thesis, are the works [121, 109, 128, 113, 108].

1.6 A summary of our main results

In the first part of this work, we study concentration behavior of nonlinear random matrices. In the second part, we study lower bounds against the SoS hierarchy for several fundamental problems.

1.6.1 Nonlinear matrix concentration via Matrix Efron-Stein

We start by giving a general theorem on concentration of random matrices whose entries are polynomials of independent random variables. The famous matrix-Bernstein inequality answers this question when we only have linear polynomials. But understanding the setting of non-linear polynomials is just as important yet it poses significant challenges. When they

arise in various applications in the literature, the usual way to handle such random matrices has been the so-called trace method. While this method gives the desired results, sometimes to great effect, applying it usually turns out to be highly nontrivial. In this work, we propose an alternate way to prove matrix concentration via the Matrix Efron-Stein inequalities. We propose a general matrix concentration inequality, the proof of which relies on the powerful method of exchangeable pairs. We show some applications of this inequality and expect it to have significant applications outside what we have explored here.

1.6.2 Sum of Squares lower bounds

We obtain strong sub-exponential time lower bounds against the SoS hierarchy for a variety of fundamental problems in computer science. All our applications start with the so-called pseudocalibration heuristic, reducing the problem to analyzing the behavior of a large random matrix, known as the *moment matrix*. Our conceptual and technical innovations happen at this step. The results we present are as follows.

Sherrington-Kirkpatrick Hamiltonian

An important problem in statistical physics, the Sherrington-Kirkpatrick problem is to optimize the quadratic form of a random matrix sampled from the Gaussian Orthogonal Ensemble, over boolean vectors. It's been known for a long time that the true optimal value concentrates at a particular constant, upto scaling. And recently, an efficient algorithm has also been proposed for this optimization problem. Certification on the other hand was widely believed to be hard beyond the simple spectral algorithm. We provide strong evidence for this by exhibiting lower bounds against SoS for this problem. This work requires us to understand the nullspace of the moment matrix and *nullify it* before applying our matrix concentration tools. Conceptually, this work provides a lot of insight into the behavior of SoS on other fundamental problems such as maximum cut and learning mixtures of Gaussians.

Sparse PCA

Sparse PCA is a variant of principal components analysis (PCA), a fundamental routine in statistics and machine learning. We work with the spiked Wishart model, which is the most natural version of this problem, but which has proved quite hard to analyze in SoS. Prior works have predicted the computational barrier of the recovery of the sparse component, as a tradeoff between the dimension, sparsity and number of samples. We confirm this barrier by proving lower bounds, matching known algorithms, against sub-exponential time SoS. This work involves splitting the random moment matrix into different matrices and using innovative combinatorial charging arguments to study how these matrices interact with each other. Conceptually, this work confirms the computational barrier diagram for this problem, that has been predicted and believed to be true for a long time.

Planted Slightly Denser subgraph

Finding a dense subgraph in a given graph is an important problem that has received much scrutiny over the years, both algorithmically as well as from the algorithmic hardness angle. For random instances of the problem under certain parameter regimes, the difficulty of this problem has been conjectured, usually referred to as the PDS conjecture, and this problem has been used as a canonical hard problem to reduce to various other problems and study their computational barriers. Moreover, these hard instances have also been used as a basis for cryptographic schemes. Therefore, SoS lower bounds against this problem go a long way towards confirming this conjecture. In this work, we exhibit such sub-exponential time lower bounds for certain parameter regimes, where it has been widely believed to require sub-exponential time.

Tensor PCA

Tensor PCA is the average-case version of the problem of optimizing homogeneous polynomials over the sphere, which is a fundamental and important problem in optimization due to it's connections to a variety of fields. In this work, we prove SoS lower bounds matching known algorithms for this problem, settling the computational barrier for SoS for this problem. It also offers insight on the approximability-inapproximability threshold for general homogeneous polynomial optimization and suggests that random instances may not be the hardest for this problem.

1.7 Excluded work

This dissertation contains the main body of my research conducted during my PhD but there has also been other research directions that have been left out, regrettably. This includes the following works.

1.7.1 SoS Lower bounds for Sparse Independent Set

In our work [95], we show SoS lower bounds for the maximum independent set problem on sparse Erdős-Rényi random graphs, matching the Lovász theta function upto low order terms. To do this, we build on the tools developed in this dissertation as well as develop a variety of new techniques. In particular, this work is the first venture in the important research direction of understanding the limitations of SoS on sparse random graphs. We highlight that for this work, our nonlinear matrix concentration tools from Chapter 2 are very useful. We will elaborate on this result in Chapter 8 since it builds on much of the work we will develop in this dissertation.

1.7.2 Causal Inference

Causal inference is the study of discovering and understanding causal relationships in observed data, with diverse applications in medicine, genetics, economics, epidemics, artificial intelligence, etc. In [157], we focus on the problem of learning a class of causal models known as Bayesian Networks (BN), from data. This is a classical and fundamental problem since BNs are compact, modular and offer intuitive causal interpretation, which has made them very useful in various fields. We propose and study a new practical algorithm for this problem. It is efficient, provably differs from the widely used Greedy-Equivalence-Search algorithm, and since the algorithm is a general-purpose score-based learning algorithm, it is widely applicable. Also, under some statistical assumptions that are inspired from and which generalize recent works, our algorithm provably recovers the true Bayesian Network, even for non-parametric models making no assumptions on linearity, additivity, independent noise or faithfulness. It also suggests interesting potential connections to other machine learning fields such as clustering, forward-backward greedy methods, and kernel methods.

In [105], we study a relatively understudied but just as important setting, where we have unobserved (sometimes even unmeasurable!) latent causes or confounders for the observed variables. We focus on the setting of probabilistic mixture models, which naturally comes up in machine learning, economics, finance, biology, etc. Under some natural assumptions on the model, we develop an algorithm that takes the observed data and uncovers the hidden variables and the underlying causal relationships. Prior works related to this problem have usually focused on special settings such as linear models. We instead propose an algorithm to this problem in the highly nonlinear mixture models setting which works atop existing algorithms for mixture model order estimation (which is easier than density estimation).

1.8 Organization of the thesis

In Chapter 2, we develop our nonlinear matrix concentration results and show it's applications towards various nonlinear random matrices that have arisen in the literature. We then introduce the Sum of Squares hierarchy in Chapter 3, introduce the technique of pseudocalibration used for showing SoS lower bounds and show it's connections to low-degree algorithms. In Chapter 4, we formally state the main SoS lower bounds we show in this thesis and put them in context with known prior works. In Chapter 5, we prove the SoS lower bound for the Sherrington-Kirkpatrick problem. And in the next two chapters, Chapter 6 and Chapter 7, we prove the SoS lower bounds for Planted Slightly Denser Subgraph, Tensor PCA and Sparse PCA. We conclude with followup and potential future works in Chapter 8.

CHAPTER 2

NONLINEAR MATRIX CONCENTRATION

In this chapter, we will describe our techniques for nonlinear matrix concentration via Efron-Stein inequalities. The material in this chapter is adapted from [158], which is joint work with Madhur Tulsiani.

2.1 Introduction

In optimization, statistics, and spectral algorithms, we often want to understand the concentration of various random matrices. To do this, we can appeal to the powerful theory of matrix-deviation inequalities [180]. For example, the matrix-Bernstein inequality addresses random matrices of the form

$$\mathbf{M} \;=\; x_1 \cdot \mathbf{C_1} + \cdots + x_n \cdot \mathbf{C_n}$$

where $x_1, \ldots, x_n$ are independent scalar random variables, and $\mathbf{C_1}, \ldots, \mathbf{C_n}$ are fixed matrices. A large selection of such inequalities are available when the random matrix (say) $\mathbf{M}$ is a *linear* function of independent random variables. However, several recent works require us to understand random matrices which are *non-linear* functions, and in particular low-degree polynomial functions, of scalar random variables. This forms the focus of our work.

As a motivating example, consider the random matrix $\mathbf{M} \in \mathbb{R}^{[n]^2 \times [n]^2}$ obtained as

$$\mathbf{M} \;=\; \mathbf{A_1} \otimes \mathbf{A_1} + \cdots + \mathbf{A_m} \otimes \mathbf{A_m}\,,$$

where $\mathbf{A_1}, \ldots, \mathbf{A_m} \in \mathbb{R}^{[n] \times [n]}$ are independent random matrices, with i.i.d. entries uniformly distributed in $\{-1, 1\}$. It is easy to see that the entries of the matrix $\mathbf{M}$ are degree-2 polynomial functions of the independent random variables describing the entries of $\mathbf{A_1}, \ldots, \mathbf{A_m}$.

The concentration of such a matrix was analyzed by Hopkins et al. [88, 81], who use it to design spectral algorithms for a variant of the principal components analysis (PCA). This matrix is a special case of a more general setting that we study in this work.

Matrix-valued polynomial functions. In the example above, the entries of the matrices are low-degree polynomials in independent (Rademacher) random variables. In this work, we consider a general setting where we take an n-tuple $Z = (Z_1, \ldots, Z_n)$ of independent and identically distributed random variables[1] distributed in Ω. We consider random matrices given by a matrix-valued function $\mathbf{F}(Z)$ taking values in $\mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ for arbitrary index sets $\mathcal{I}, \mathcal{J}$, where each entry $\mathbf{F}[I, J](Z)$ is a polynomial in $Z_1, \ldots, Z_n$. We develop a general framework to analyze concentration of such matrices. Our matrix concentration results are simpler to state in the case when $Z_1, \ldots, Z_n$ are independent Rademacher variables uniformly distributed in $\{-1, 1\}$, but apply for the general case as well.

Special cases of such non-linear random matrices have been used in several applications in spectral algorithms and lower bounds. We now briefly discuss a few examples below.

1. **Tensor networks.** Random matrices such as the above were viewed as a special case of "flattened tensor networks" by Moitra and Wein [131], who also considered spectral algorithms obtained via somewhat larger tensor networks. A tensor network is a graph with nodes corresponding to tensors (see the figure below for an example). An edge between two nodes corresponds to shared indices for one of the dimensions and the degree of each node is equal to the order of the corresponding tensor (the number of dimensions). Such networks indicate how tensors of different orders can be multiplied to obtain larger ones. For example, the first network in the figure below illustrates the network corresponding to simple multiplication $\mathbf{A} \cdot \mathbf{B}$ of two matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$, where the red and blue edges indicate the row and column indices

1. Our framework also applies when the variables are not necessarily identically distributed, as long as they are independent.

respectively. Similarly, the second network in the figure below illustrates the network corresponding to the application by Hopkins et al. [87], where $\mathbf{T} \in \mathbb{R}^{n \times n \times m}$ is a random tensor with i.i.d. entries in $\{-1, 1\}$. While the latter network yields an order-4 tensor, they obtain a matrix in $\mathbb{R}^{n^2 \times n^2}$ by "flattening" it, where the row is indicated by the indices in the red edges and the column is indicated by the indices in the blue edges. In the figure, we also indicate the index sets corresponding to each of the edges (though these are often supressed in the diagrams). Moitra and Wein [131] analyzed a larger tensor network, with a graph consisting of 10 nodes, in their algorithm for the continuous multi-reference alignment problem.

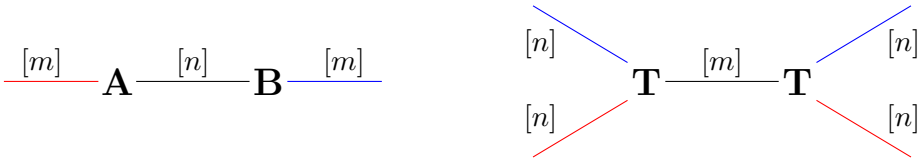

Figure 2.1: Tensor networks for matrix multiplication and the algorithm in [87]

2. **Graph matrices.** Another setting of nonlinear concentration arises from the analysis of the so-called "graph matrices" [126, 1]. Graph matrices play an important role in lower bounds for average-case problems, against algorithms based on the powerful Sum-of-Squares (SoS) SDP hierarchy running in polynomial time and even sub-exponential time [127, 49, 84, 154, 13, 128**?** , 149, 95].

   Let $\mathbf{X}$ be the $\{\pm 1\}$-adjacency matrix of a random graph in $\mathcal{G}_{n,1/2}$ i.e., $\mathbf{X}[i, j]$ is uniform $\{-1, 1\}$ when $i \neq j$ and 0 when $i = j$. Graph matrices are random matrices corresponding to the occurences of a small graph pattern called a "shape". A shape τ is a small, fixed graph with two ordered subsets U_τ, V_τ of vertices. For simplicity, let τ be a shape of a fixed size, where the vertex set $V(\tau)$ is partitioned into two ordered sets $V(\tau) = U_\tau \sqcup V_\tau$. For such a shape τ, the corresponding *graph matrix* $\mathbf{M}_\tau$ has rows and columns indexed by $[n]^{|U_\tau|}$ and $[n]^{|V_\tau|}$ respectively, and we view the row and

column indices I and J as defining a (unique in this case) map $\varphi : U_\tau \sqcup V_\tau \to [n]$. The corresponding entry is given by

$$\mathbf{M}_\tau[I, J] \;=\; \mathbf{M}_\tau[\varphi(U_\tau), \varphi(V_\tau)] \;=\; \begin{cases} \prod_{(u,v)\in E(\tau)} \mathbf{X}[\varphi(u), \varphi[v]] & \text{if } \varphi \text{ is injective} \\ 0 & \text{otherwise} \end{cases}$$

In the case of general graph matrices (defined formally in Section 2.4.2), U_τ, V_τ are arbitrary ordered subsets of the vertex set of τ, and we sum over all feasible injective maps φ. As an example, consider the case shown in Fig. 2.2, where τ is a triangle on three vertices $\{u_1, v_1, v_2\}$ with $U_\tau = (u_1)$ and $V_\tau = (v_1, v_2)$. Then, the corresponding matrix is given by

$$\mathbf{M}_\tau[i_1, (i_2, i_3)] \;=\; \mathbf{X}[i_1, i_2] \cdot \mathbf{X}[i_2, i_3] \cdot \mathbf{X}[i_3, i_1]\,,$$

where $\mathbf{X}$ automatically enforces injectivity.

Graph matrices are closely related to tensor networks (ignoring the injectivity constraint on φ). For instance, the above matrix can be viewed as the flattened tensor network below, where the tensor $\mathbf{I}$ denotes the "diagonal" tensor of order 3 with entries being 1 if all indices are equal and 0 otherwise.

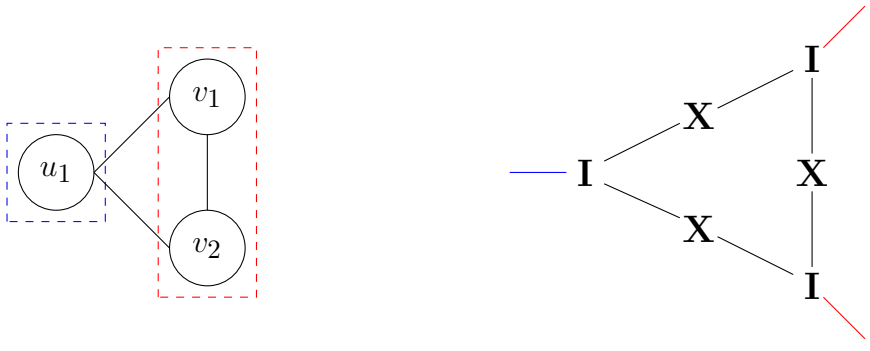

Figure 2.2: The graph τ and corresponding flattened tensor network

Analyzing concentration Recall that our objective is to analyze the concentration of polynomial random matrices. To motivate our approach, consider first the problem of obtaining concentration bounds on a *scalar* polynomial $f(Z)$ with mean zero. To obtain such bounds, because of Markov's inequality, it suffices to compute moment estimates

$$\mathbb{P}\left[|f(Z)| \geq \lambda\right] \;=\; \mathbb{P}\left[(f(Z))^{2t} \geq \lambda^{2t}\right] \;\leq\; \lambda^{-2t} \cdot \mathbb{E}\left[(f(Z))^{2t}\right]$$

While in some cases $\mathbb{E}[(f(Z))^{2t}]$ can be computed by direct expansion, it often involves an intricate analysis of the structure of terms with degrees growing with t, and therefore indirect methods may be more convenient. One such method is based on hypercontractive inequalities. In particular for Rademacher variables, the hypercontractive inequality [136] gives that for a polynomial f of degree d_p, we have

$$\mathbb{E}\left[(f(Z))^{2t}\right] \;\leq\; (2t-1)^{d_p \cdot t} \cdot \left(\mathbb{E}\left[(f(Z))^{2}\right]\right)^{t}.$$

Thus, for (scalar) polynomial functions, the hypercontractive inequality gives moment estimates using $(f(Z))^2$, which is convenient because $(f(Z))^2$ is a polynomial of *fixed* degree and therefore is much easier to understand. In fact, it can often be conveniently analyzed using the Fourier coefficients of f.

The matrix analog of the above argument involves the Schatten-$2t$ norm $\|.\|_{2t}$, which is defined for a matrix $\mathbf{M}$ with non-zero singular values $\sigma_1, \ldots, \sigma_r$ as $\|\mathbf{M}\|_{2t}^{2t} \;:=\; \sum_{j\in[r]} \sigma_j^{2t}$. For a function $\mathbf{F}$ with $\mathbb{E}[\mathbf{F}(Z)] = 0$, we have the following bound using Schatten norms.

$$\mathbb{P}\left[\sigma_1(\mathbf{F}) \geq \lambda\right] \;\leq\; \lambda^{-2t} \cdot \mathbb{E}\,\|\mathbf{F}\|_{2t}^{2t} \;=\; \lambda^{-2t} \cdot \mathbb{E}\,\mathrm{tr}\left[(\mathbf{F}(Z)\mathbf{F}(Z)^{\intercal})^{t}\right]$$

Known norm bounds for tensor networks [131] (which involves Gaussian variables) and graph matrices [1, 95] rely on direct expansion of the trace above. They analyze terms in the

expansion as being formed by $2t$ copies of the network/shape, which leads them to consider graphs formed by $2t$ copies of the network/shape, with possibly overlapping vertex sets. To analyze such graphs, they both rely on intricate combinatorics.

While hypercontractive inequalities are also known for matrix-valued functions of Rademacher variables [18], their form involves Schatten-p norms for $p \in [1, 2]$ and (to the best of our knowledge) are not known to imply matrix concentration. To get around this, we consider another indirect method based on Efron-Stein inequalities. In the scalar case, Efron-Stein inequalities gives us a slight weakening of the above scalar bound. Interestingly, it turns out that this can indeed be generalized to the matrix case.

Efron-Stein inequalities. Efron-Stein inequalities bound the global variance of a function of independent random variables, in terms of local variance estimates obtained by changing one variable at a time. For $i \in [n]$ and tuple $Z = (Z_1, \ldots, Z_n)$, let $Z^{(i)}$ denote the tuple $(Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n)$, where $\widetilde{Z}_i$ is an independent copy of Z_i. For a scalar function $f(Z)$, the Efron-Stein inequality states that

$$\mathrm{Var}\left[f(Z)\right] \;=\; \mathbb{E}\left[(f(Z) - \mathbb{E}\, f)^2\right] \;\leq\; \frac{1}{2} \cdot \sum_{i \in [n]} \mathbb{E}\left[\left(f(Z) - f\left(Z^{(i)}\right)\right)^2\right] \;=\; \mathbb{E}\left[V(Z)\right],$$

where $V(Z) \;:=\; \sum_{i \in [n]} \mathbb{E}\left[\left(f(Z) - f\left(Z^{(i)}\right)\right)^2 | Z\right]$. For Rademacher variables, $\mathbb{E}[V(Z)]$ is equal to the total influence from boolean Fourier analysis and indeed, the above inequality can also be observed via Fourier analysis. In fact, when f is a polynomial of degree d_p, the two sides are within a factor d_p.

A moment version of the Efron-Stein inequality was developed by Boucheron et al. [28], who obtain bounds in terms of $V(Z)$ (in fact, in terms of more refined quantities $V_+(Z)$ and $V_-(Z)$) which serves as a proxy for the variance. Their results imply that for a function f,

$$\mathbb{E}\left[(f(Z) - \mathbb{E}\, f)^{2t}\right] \;\leq\; (C_0 \cdot t)^t \cdot \mathbb{E}\left[(V(Z))^t\right].$$

A beautiful matrix generalization of the above inequality (Theorem 2.1.1 below) was obtained by Paulin, Mackey and Tropp [147], via the method of exchangeable pairs (see also [93] for a different proof). Their inequality is stated for Hermitian matrix valued functions $\mathbf{H}$. But we can also use it for non-Hermitian functions $\mathbf{F}$, where we simply apply it to the Hermitian dilation $\mathbf{H} = \begin{bmatrix} 0 & \mathbf{F} \\ \mathbf{F}^\intercal & 0 \end{bmatrix}$ instead.

Theorem 2.1.1 ([147])**.** *Let $\mathbf{H}(Z)$ be a Hermitian matrix valued function of independent random variables $Z = (Z_1, \ldots, Z_n)$ with $\mathbb{E}\,\|\mathbf{H}\| < \infty$. Then, for each natural number $t \geq 1$,*

$$\mathbb{E}\,\mathrm{tr}\left[(\mathbf{H} - \mathbb{E}\,\mathbf{H})^{2t}\right] \;\leq\; (4t-2)^t \cdot \mathbb{E}\,\mathrm{tr}\left[\mathbf{V}^t\right] ,$$

where $\mathbf{V}(Z)$ is the variance proxy defined as

$$\mathbf{V}(Z) := \frac{1}{2} \cdot \sum_{i=1}^{n} \mathbb{E}\left[\left(\mathbf{H}(Z) - \mathbf{H}\left(Z^{(i)}\right)\right)^2 \mid Z\right] .$$

A simple bound for Rademacher variables. The form of the variance proxy suggests a recursive approach for polynomial functions (say of degree d_p) of Rademacher variables. Consider the scalar case again in particular the Efron-Stein inequality by Boucheron et al. [28], where the variance proxy can be written as

$$\begin{aligned} V(Z) \;=\; \frac{1}{2} \cdot \sum_{i\in[n]} \mathbb{E}\left[\left(f(Z) - f\left(Z^{(i)}\right)\right)^2 \mid Z\right] \;&=\; \frac{1}{2} \cdot \sum_{i\in[n]} \mathbb{E}\left[(Z_i - \widetilde{Z}_i)^2 \cdot \left(\frac{\partial f(Z)}{\partial Z_i}\right)^2 \mid Z\right] \\ &=\; \sum_{i\in[n]} \left(\frac{\partial f(Z)}{\partial Z_i}\right)^2 \;=\; \|\mathbf{f}_1(Z)\|_2^2 , \end{aligned}$$

where $\mathbf{f}_1(Z)$ is a vector-valued function given by $\mathbf{f}_1[i](Z) = \frac{\partial f(Z)}{\partial Z_i}$. Thus, to estimate $\mathbb{E}\,(f(Z))^{2t}$, we just need to estimate $\mathbb{E}\,\|\mathbf{f}_1(Z)\|_2^{2t}$, where $\mathbf{f}_1(Z)$ is now a vector valued function. The key observation is that $\mathbf{f}_1(Z)$ has entries of degree at most $d_p - 1$. This suggests that we

can apply this inequality recursively until we end up with constant polynomials, which we fully understand. We can do a similar computation for matrix-valued functions $\mathbf{F}(Z)$ using Theorem 2.1.1. This yields two matrices $\mathbf{F}_{0,1}$ and $\mathbf{F}_{1,0}$ of partial derivatives, where an extra index i is added either to the row or column indices. Iterating this yields the following result, which we state in terms of the partial derivative operators $\nabla_\alpha(f) = \left(\prod_{i:\alpha_i=1} \frac{\partial}{\partial Z_i}\right)(f)$ for $\alpha \in \{0,1\}^n$ (extended entry-wise to matrices).

Theorem 2.1.2 (Rademacher recursion)**.** *Let $\mathbf{F} : \{-1,1\}^n \to \mathbb{R}^{\mathcal{I}\times\mathcal{J}}$ be a matrix valued polynomial function of degree at most d_p. Then, for each natural number $t \geq 1$,*

$$\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \;\leq\; \sum_{1\leq a+b\leq d_p} (16td_p)^{(a+b)\cdot t} \cdot \left\|\mathbb{E}\,\mathbf{F}_{a,b}\right\|_{2t}^{2t},$$

where $\mathbf{F}_{a,b}$ is a matrix of partial derivatives indexed by the sets $\mathcal{I}\times\{0,1\}^n$ and $\mathcal{J}\times\{0,1\}^n$ with

$$\mathbf{F}_{a,b}[(\cdot,\alpha),(\cdot,\beta)] = \begin{cases} \nabla_{\alpha+\beta}(\mathbf{F}) & \text{if } |\alpha| = a, |\beta| = b, \alpha\cdot\beta = 0 \\ 0 & \text{otherwise} \end{cases}$$

Similar to the hypercontractive bound for the scalar case, the bound above is in terms of a small number ($O(d_p^2)$) of matrices that arise from polynomials of fixed degree (not growing with t), but importantly, they are *deterministic* matrices. Because they are deterministic, analyzing them is considerably easier. When we apply this theorem to the case $\mathbf{F} = \mathbf{M}_\tau$, the graph matrix of a shape τ, we obtain bounds in terms of combinatorial objects known as "vertex separators" of the shape τ. This recovers the bounds by Ahn et al. [1] and perhaps surprisingly (to the authors), this gives an alternative and direct derivation of these combinatorial structures such as vertex separators, compared to the ingenious observations made in Ahn et al. [1]. We cover this and other applications of the Rademacher framework in Section 2.4.

Extending the framework to general product distributions. A key contribution of our work is to show how the above framework can be extended to arbitrary product distributions (with bounded moments). A motivating example of this is norm bounds for the so-called "sparse graph matrices". In sparse graph matrices, the variables Z_i can be thought of as (normalized) edges of a $\mathcal{G}_{n',p}$ graph, that is, $Z_i = -\sqrt{\frac{1-p}{p}}$ with probability p and $Z_i = \sqrt{\frac{p}{1-p}}$ with probability $1-p$. These variables are standard in p-biased Fourier analysis [**?**] and are chosen to satisfy $\mathbb{E}\, Z_i = 0$ and $\mathbb{E}\, Z_i^2 = 1$. Sparse graph matrices naturally arise when analyzing average case problems on $\mathcal{G}_{n,p}$ graphs for $p = o(1)$, as opposed to $\mathcal{G}_{n,1/2}$ graphs.

Until recently, little was known about norm bounds for sparse graph matrices. The difficulty stems partly from the fact that when $p = o(1)$, it is important that sparse graph matrix norm bounds have the right dependence on p and not just on n. Such norm bounds were obtained recently by Jones et al. [95], via the trace power method which involved a delicate combinatorial counting argument. On the other hand, we obtain similar norm bounds using our framework but in a more mechanical fashion. We can also readily apply our framework in the even more general case of sub-Gaussian random variables and our bounds will depend on the sub-Gaussian norm of the distributions.

To extend our framework to general product distributions, we could take inspiration from the Rademacher case and could attempt to simply recursively apply the Efron-Stein inequality. Unfortunately, this idea will fail. The issue can be observed by again considering the scalar case. Assume that $Z_1, \ldots, Z_n$ are i.i.d. with $\mathbb{E}\, Z_i = 0$ and $\mathbb{E}\, Z_i^2 = 1$ for all $i \in [n]$. Also assume for simplicity that $f(Z)$ is a multi-linear polynomial of degree d_p. Analyzing the variance proxy as before, we get

$$V(Z) \;=\; \frac{1}{2} \cdot \sum_{i \in [n]} \mathbb{E}\left[(Z_i - \widetilde{Z}_i)^2 \cdot \left(\frac{\partial f(Z)}{\partial Z_i} \right)^2 \mid Z \right] \;=\; \frac{1}{2} \sum_{i \in [n]} \mathbb{E}\left[(Z_i - \widetilde{Z}_i)^2 | Z \right] \cdot \left(\frac{\partial f(Z)}{\partial Z_i} \right)^2 .$$

In the Rademacher case, we had $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2|Z] = 2$. This left us with the polynomials corresponding to partial derivatives but which importantly had a strictly lower degree. However, for a general product distribution, we instead have $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2|Z] = 1 + Z_i^2$. This gives back a term $\left(Z_i \cdot \frac{\partial f}{\partial Z_i}\right)^2$ where the polynomial inside the square could have degree possibly still equal to d_p. This means that in the next step of the recursion, we may again have to consider a derivative with respect to Z_i and may again end up with the same polynomial f. Therefore, the recursion is stalled! A similar issue occurs for matrices, which is elaborated in Section 2.5. To get around this, we generalize the work of [147].

Generalizing [147] via explicit inner kernels. To resolve the above issue, we modify the proof of [147] and our proof techniques may be of independent interest.

We first recall how the matrix Efron-Stein inequality, Theorem 2.1.1, was proved in [147]. Their basic strategy is to utilize the theory of *exchangeable pairs* [172, 173, 38, 39], in particular *kernel Stein pairs*. A kernel Stein pair is an exchangeable pair of random matrices that has a "kernel", a bivariate function that "reproduces" the matrices in the pair. More concretely, consider an exchangeable pair of random variables (Z, Z') (which means (Z', Z) has the same distribution). For this exchangeable pair, a bivariate matrix-valued function $\mathbf{K}(z, z')$ is said to be a kernel for a matrix-valued function $\mathbf{F}$ if it satisfies

- Anti-symmetry: $\mathbf{K}(z', z) = -\mathbf{K}(z, z')$ for all inputs (z, z').
- Reproducing property: $\mathbb{E}[\mathbf{K}(Z, Z') \mid Z] = \mathbf{F}(Z)$.

If such a kernel $\mathbf{K}$ exists, then the pair of random variables $(\mathbf{F}(Z), \mathbf{F}(Z'))$ is said to be a kernel Stein pair.

Building on ideas from [173, 38], Paulin, Mackey and Tropp [147] first show the existence of a kernel, by exhibiting it as a limit of coupled Markov Chains. By studying the evolution of this kernel coupling, they prove analytic properties of the kernel. Then, using this kernel,

they employ the powerful method of exchangeable pairs to evaluate moments of the random matrix, which in turn will imply concentration.

For a Hermitian random matrix $\mathbf{X}$, they introduce two matrices - the *conditional variance* $\mathbf{V_X}$ which measures the squared fluctuations of $\mathbf{X}$ when resampling a coordinate of Z; and the *kernel conditional variance* $\mathbf{V^K}$ which measures the squared fluctation of the kernel when resampling a coordinate of Z. With these matrices in hand, they bound the Schatten $2t$-norm of $\mathbf{X}$ by the Schatten t-norm of $s\mathbf{V_X} + s^{-1}\mathbf{V^K}$ for any parameter $s > 0$. Finally, they choose s appropriately to make these two quantities approximately equal, in which case it simplifies to the variance proxy $\mathbf{V}$, proving Theorem 2.1.1.

In our setting, no such choice of s is feasible because for any choice of s, either the conditional variance term $s\mathbf{V_X}$ will dominate $\mathbf{X}^2$ or the kernel conditional variance term $s^{-1}\mathbf{V^K}$ will dominate $\mathbf{X}^2$. This will make the main inequality Theorem 2.1.1 trivial.

To get around this, we will exploit the structure of the matrix we have, i.e. $\mathbf{F} = \mathbf{DGD}$ where $\mathbf{D}$ is a diagonal matrix that encodes all variables that have already been differentiated on and $\mathbf{G}$ is a polynomial matrix of the remaining variables. Since $\mathbf{D}$ is a simple diagonal matrix with low degrees, most of the deviations exhibited by $\mathbf{F}$ are in fact likely to be exhibited by $\mathbf{G}$. To capture this intuition, we consider a kernel for only the inner matrix $\mathbf{G}$ instead of $\mathbf{F}$ as a whole. We call this an *inner kernel*.

This helps us avoid the root cause of the issue, i.e. differentiating on variables we have already encountered (which correspond to entries in $\mathbf{D}$). Therefore, the recursion will not stall!

However, in general, this is not realizable since $\mathbf{D}$ and the kernel of $\mathbf{G}$ can interact in unexpected ways. To study this interaction, we construct explicit polynomial kernels (Theorem 2.7.3) (compared to [147] who show the existence of the kernel but for all functions).

We study how this explicit inner kernel interacts with $\mathbf{D}$ (see Lemma 2.7.6) and use it to obtain a generalization of the inequalities by [147] (generalized because setting $\mathbf{D} = \mathbf{I}$ will

give back their result) stated in Lemma 2.7.10.

A subtle issue is that the conditional variance of $\mathbf{X}$ may still have additional deviations due to the diagonal matrices $\mathbf{D}$ (which still involve random variables). We control the additional deviations using Jensen's operator trace inequality (for non-commuting averages) [78] (stated in Lemma 2.2.4). Putting these ideas together lets us obtain a version of the Efron-Stein inequality where the variance proxy only corresponds to the conditional variance of the inner kernel. In the setting of polynomial functions, this inequality generalizes the work of [147].

With the modified Efron-Stein inequality from above, we cannot guarantee that the matrices $\mathbf{F}$ at intermediate steps are of lower degree, but on the other hand, the degree of the inner matrix $\mathbf{G}$ reduces at each step. Therefore, we can recursively apply this inequality to obtain our final bounds. The final bounds are then stated in terms of norm bounds for the simplified matrices of the form $\mathbf{DGD}$ where $\mathbf{G}$ are deterministic matrices and $\mathbf{D}$ are diagonal matrices which are still functions of Z. While random, these matrices can be easily analyzed via simple scalar concentration tools.

The main theorem is stated in Section 2.6, in particular Theorem 2.6.6, with the proof following in Section 2.7. While our proof builds on the work by [147], the argument here is self-contained.

Applications. Our framework is suitable for many nonlinear concentration results obtained in the literature [11, 69, 88, 126, 1, 87, 165, 81, 86, 131, 95]. We show a few of these applications in Section 2.4 and Section 2.8. We expect similar future applications to benefit from our framework because the task is mechanically reduced to analyzing considerably simpler matrices.

In Section 2.4.2, we derive norm bounds on dense graph matrices. In earlier works, dense graph matrices have been used extensively in analysis of semidefinite programming hierarchies, especially the Sum-of-Squares (SoS) hierarchy [127, 49, 84, 154, 13, 128**?** , 149].

For more applications and a detailed treatment of graph matrices, see [1].

In Section 2.8, we derive norm bounds for sparse graph matrices. Sparse graph matrices have been relatively less understood until recently, when [95] obtained norm bounds for such matrices via the trace power method. They use these bounds to prove SoS lower bounds for the maximum independent set problem on sparse graphs.

Potential extensions In this work, we assumed that the input forms a product distribution. In other words, the variables $Z_1, \ldots, Z_n$ are independent. A natural extension is the case when they are not independent. This has important applications for many problems such as when the input is a uniform d-regular graph, or when the input is sampled from a distribution with a global constraint, etc. In such cases, the input variables are not independent but it may be possible to use similar ideas to analyze concentration.

More concretely, to study concentration in the non-independent setting, one can use the recent work of Huang and Tropp [93] on matrix concentration from Poincaré inequalities, together with our framework. For this, we just need to exhibit a Markov process that converges to our desired distribution.

Organization of the chapter We start with preliminaries in Section 2.2. In Section 2.3, we state and prove the Rademacher recursion. We illustrate some applications of this framework in Section 2.4. In Section 2.5, we explain why similar ideas may not be enough in the general case. We then propose our general framework in Section 2.6 and prove it in Section 2.7. We end with an application of the general framework to sparse graph matrices in Section 2.8.

2.2 Preliminaries

Notation We use boldface letters such as $\mathbf{I}, \mathbf{M}, \mathbf{X} \ldots$, to denote matrices. Entries of a matrix $\mathbf{X} \in \mathbb{R}^{\mathcal{I} \times \mathcal{J}}$ will be denoted by $\mathbf{X}[I, J]$ for $I \in \mathcal{I}, J \in \mathcal{J}$. Let $\mathbb{H}^n$ denote the set of

$n \times n$ real symmetric matrices. The trace of a matrix $\mathbf{X} \in \mathbb{H}^n$ equals $\sum_{i \in [n]} \mathbf{X}[i,i]$ and is denoted by $\operatorname{tr} \mathbf{X}$.

Multi-index notation

For any pair of vectors $\alpha, \beta \in \mathbb{N}^n$ and scalar $c \in \mathbb{N}$, we define $\alpha + \beta, \alpha \cdot \beta, c\alpha$ entrywise. We also define the orderings $\alpha \leq \beta$ and $\alpha \trianglelefteq \beta$ where we say $\alpha \leq \beta$ if for each i, $\alpha_i \leq \beta_i$, and $\alpha \trianglelefteq \beta$ if for each i, α_i is either 0 or β_i. We denote by $|\alpha|_0$ the number of nonzero entries of α and by $|\alpha|_1$, the sum of entries of α. For a boolean vector $\gamma \in \{0,1\}^n$, we define $1 - \gamma$ the vector with all its bits flipped.

Derivatives

For variables $Z_1, \ldots, Z_n$ and $\alpha \in \mathbb{N}^n$, define the monomial $Z^\alpha := \prod_{i=1}^n Z_i^{\alpha_i}$. This forms a standard basis for polynomials.

For $\alpha \in \mathbb{N}^n$, we define the linear operator ∇_α that acts on polynomials by defining its action on the elements Z^β as follows and then extend linearly to all polynomials.

$$\nabla_\alpha(Z^\beta) = \begin{cases} Z^{\beta - \alpha} & \text{if } \alpha \trianglelefteq \beta \\ 0 & \text{o.w.} \end{cases}$$

Informally, for a polynomial f written as a linear combination of the standard basis polynomials Z^β, $\nabla_\alpha(f)$ isolates the terms that precisely contain the powers $Z_i^{\alpha_i}$ for all i such that $\alpha_i \neq 0$ and then truncates these powers. In other words, it's the coefficient of Z^α in f. In particular, observe that $\nabla_\alpha(f)$ does not depend on Z_i for any i such that $\alpha_i \neq 0$.

Supose f is multilinear, as we can assume in the Rademacher case when we are working with $Z_i \in \{-1, 1\}$. For $\alpha \in \{0,1\}^n$ with nonzero indices $i_1, \ldots, i_k \in [n]$, we have $\nabla_\alpha(f) = \frac{\partial}{\partial Z_{i_1}} \cdots \frac{\partial}{\partial Z_{i_k}} f$. So this linear operator generalizes the partial derivative operator. But note that in general, ∇ is not simply the standard partial derivative operator.

Matrix Analysis

Linear operators that act on polynomials can also be naturally defined to act on matrices by acting on each entry.

We define $\mathbf{I}_m$ to be the $m \times m$ identity matrix. We drop the subscript when it's clear. For matrices $\mathbf{F}, \mathbf{G}$, define $\mathbf{F} \oplus \mathbf{G}$ to be the matrix $\begin{bmatrix} 0 & \mathbf{F} \\ \mathbf{G} & 0 \end{bmatrix}$. For a matrix $\mathbf{F}$, define its Hermitian dilation $\overline{\mathbf{F}}$ as $\mathbf{F} \oplus \mathbf{F}^T$. Denote by $\preceq$ the Loewner order, that is, $\mathbf{A} \preceq \mathbf{B}$ for $\mathbf{A}, \mathbf{B} \in \mathbb{H}^n$ if and only if $\mathbf{B} - \mathbf{A}$ is positive semi-definite.

Definition 2.2.1. *For a matrix $\mathbf{F}$ and an integer $t \geq 0$, define the Schatten $2t$-norm as*

$$\|\mathbf{F}\|_{2t}^{2t} = \mathrm{tr}[(\mathbf{F}\mathbf{F}^T)^t]$$

Fact 2.2.2. *For real symmetric matrices $\mathbf{X}_1, \ldots, \mathbf{X}_n$, we have*

$$(\mathbf{X}_1 + \ldots + \mathbf{X}_n)^2 \preceq n(\mathbf{X}_1^2 + \ldots + \mathbf{X}_n^2)$$

Fact 2.2.3. *For positive semidefinite matrices $\mathbf{X}, \mathbf{X}_1, \ldots, \mathbf{X}_n$ such that $\mathbf{X} \preceq \mathbf{X}_1 + \ldots + \mathbf{X}_n$ and for any integer $t \geq 1$,*

$$\mathrm{tr}[\mathbf{X}^t] \leq n^{t-1}(\mathrm{tr}[\mathbf{X}_1^t] + \ldots + \mathrm{tr}[\mathbf{X}_n^t])$$

Proof. By Hölder's inequality, $n^{t-1}(\mathrm{tr}[\mathbf{X}_1^t] + \ldots + \mathrm{tr}[\mathbf{X}_n^t]) \geq (\|\mathbf{X}_1\|_t + \ldots + \|\mathbf{X}_n\|_t)^t$. By triangle inequality of Schatten norms, this is at least $\|\mathbf{X}_1 + \ldots + \mathbf{X}_n\|_t^t$. Finally, because $\mathbf{X}_1 + \ldots + \mathbf{X}_n \succeq \mathbf{X} \succeq 0$, we can use the monotonicity of trace functions (see [148, Proposition 1]) where we use the increasing function $f(x) = x^t$ on $x \in [0, \infty)$. This proves the result. ∎

Lemma 2.2.4 (Jensen's operator trace inequality)**.** *[78, Corollary 2.5] Let f be a convex, continuous function defined on an interval I and suppose that $0 \in I$ and $f(0) \leq 0$. Then,*

for all integers $m, n \geq 1$, for every tuple $\mathbf{B}_1, \ldots, \mathbf{B}_n$ of real symmetric $m \times m$ matrices with spectra contained in I and every tuple $\mathbf{A}_1, \ldots, \mathbf{A}_n$ of $m \times m$ matrices with $\sum_{i=1}^{n} \mathbf{A}_i \mathbf{A}_i^T \preceq \mathbf{I}$, we have

$$\mathrm{tr}[f(\sum_{i=1}^{n} \mathbf{A}_i^T \mathbf{B}_i \mathbf{A}_i)] \leq \mathrm{tr}[\sum_{i=1}^{n} \mathbf{A}_i^T f(\mathbf{B}_i) \mathbf{A}_i]$$

2.3 The basic framework for Rademacher random variables

Let $Z = (Z_1, \ldots, Z_n)$ be sampled uniformly from $\{-1, 1\}^n$. We will consider matrix-valued functions $\mathbf{F} : \{-1, 1\}^n \to \mathbb{R}^{\mathcal{I} \times \mathcal{J}}$, with rows and columns indexed by arbitrary sets $\mathcal{I}, \mathcal{J}$ respectively such that for all $I \in \mathcal{I}, J \in \mathcal{J}$,

$$\mathbf{F}[I, J] \;=\; f_{I,J}(Z)$$

where $f_{I,J}$ are polynomials of $Z_1, \ldots, Z_n$. Since $Z_i \in \{-1, 1\}$, we can assume without loss of generality that $f_{I,J}$ are multilinear. Let d_p be the maximum degree of any $f_{I,J}$ in $\mathbf{F}$. In this section, we will give a general framework using which we can obtain bounds on $\mathbb{E} \|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t}$ for any integer $t \geq 1$.

Theorem 2.1.2 (Rademacher recursion)**.** *Let $\mathbf{F} : \{-1, 1\}^n \to \mathbb{R}^{\mathcal{I} \times \mathcal{J}}$ be a matrix valued polynomial function of degree at most d_p. Then, for each natural number $t \geq 1$,*

$$\mathbb{E} \|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \;\leq \sum_{1 \leq a+b \leq d_p} (16 t d_p)^{(a+b) \cdot t} \cdot \left\|\mathbb{E}\,\mathbf{F}_{a,b}\right\|_{2t}^{2t},$$

where $\mathbf{F}_{a,b}$ is a matrix of partial derivatives indexed by the sets $\mathcal{I} \times \{0, 1\}^n$ and $\mathcal{J} \times \{0, 1\}^n$ with

$$\mathbf{F}_{a,b}[(\cdot, \alpha), (\cdot, \beta)] = \begin{cases} \nabla_{\alpha+\beta}(\mathbf{F}) & \text{if } |\alpha| = a, |\beta| = b, \alpha \cdot \beta = 0 \\ 0 & \text{otherwise} \end{cases}$$

Remark 2.3.1. *Note that while the matrices $\mathbf{F}_{a,b}$ are stated above as having rows and*

colmns indexed by $\mathcal{I} \times \{0,1\}^n$ and $\mathcal{J} \times \{0,1\}^n$ for convenience, we only need to consider the submatrices with $|\mathcal{I}| \cdot \binom{n}{a}$ rows and $|\mathcal{J}| \cdot \binom{n}{b}$ columns, since all other entries will be zero (when $|\alpha| \neq a$ or $|\beta| \neq b$).

Remark 2.3.2. *To obtain high probability norm bounds from moment estimates, we can set $t = \text{polylog}(n)$ and invoke Markov's inequality. Since we do not attempt to optimize the dependence on the logarithmic factors, we do not attempt to optimize the exponent of t in the main theorem.*

To prove this, we will prove Lemma 2.3.3 and then recursively apply it.

For each $i \leq n$, define the random vector

$$Z^{(i)} := (Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n)$$

where $\widetilde{Z}_i$ is an independent copy of Z_i, that is, is independently resampled from $\{-1, 1\}$.

Let $\mathbf{X} := \mathbf{F} - \mathbb{E}\,\mathbf{F}$. When the input is Z, we denote the matrices as $\mathbf{F}, \mathbf{X}$, etc and when the input is $Z^{(i)}$, denote the corresponding matrices as $\mathbf{F}^{(i)}, \mathbf{X}^{(i)}$, etc. That is, for $I \in \mathcal{I}, J \in \mathcal{J}$, we have $\mathbf{F}^{(i)}[I, J] = f_{I,J}(Z^{(i)})$. Define $\mathbf{X}_{a,b} = \mathbf{F}_{a,b} - \mathbb{E}\,\mathbf{F}_{a,b}$.

Lemma 2.3.3. *For integers $a, b \geq 0$, we have*

$$\mathbb{E}\left\|\mathbf{X}_{a,b}\right\|_{2t}^{2t} \leq (16td_p)^t(\mathbb{E}\left\|\mathbf{X}_{a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\mathbf{X}_{a+1,b}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{a,b+1}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{a+1,b}\right\|_{2t}^{2t})$$

Using this lemma, we can complete the proof of the main theorem.

Proof of Theorem 2.1.2. Observing that $\mathbf{X}$ is a principal submatrix of $\mathbf{X}_{0,0}$ with all other entries being 0, we can apply Lemma 2.3.3 repeatedly until $\mathbf{X}_{a,b} = 0$, which will be the case if $a + b > d_p$. ∎

In the rest of this section, we will prove Lemma 2.3.3. We start with a basic fact. Let $\mathbf{e}_i \in \{0,1\}^n$ be the vector with a unique nonzero entry $(\mathbf{e}_i)_i = 1$.

Proposition 2.3.4. *For a multilinear polynomial $f(Z) = f(Z_1, \ldots, Z_n)$, we have*

$$f(Z) - f(Z^{(i)}) \;=\; (Z_i - \widetilde{Z}_i) \cdot \nabla_{\mathbf{e}_i} f(Z)$$

Proof of Lemma 2.3.3. Consider the Hermitian dilation $\overline{\mathbf{F}}_{a,b} = \mathbf{F}_{a,b} \oplus \mathbf{F}_{a,b}^T$. Define $\overline{\mathbf{X}}_{a,b} = \overline{\mathbf{F}}_{a,b} - \mathbb{E}\,\overline{\mathbf{F}}_{a,b} = \mathbf{X}_{a,b} \oplus \mathbf{X}_{a,b}^T$. By Theorem 2.1.1 applied to $\overline{\mathbf{X}}_{a,b}$,

$$\mathbb{E}\,\mathrm{tr}\left[\overline{\mathbf{X}}_{a,b}^{2t}\right] \leq (2(2t-1))^t\,\mathbb{E}\,\mathrm{tr}\left[\mathbf{V}_{a,b}^t\right]$$

where $\mathbf{V}_{a,b}$ is the variance proxy

$$\mathbf{V}_{a,b} = \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(\overline{\mathbf{X}}_{a,b} - \overline{\mathbf{X}}_{a,b}^{(i)})^2 | Z]$$

Firstly, by a simple computation,

$$\mathbb{E}\,\mathrm{tr}\left[\overline{\mathbf{X}}_{a,b}^{2t}\right] = \mathbb{E}\,\mathrm{tr}\left[(\mathbf{X}_{a,b}\mathbf{X}_{a,b}^{\intercal})^t\right] + \mathbb{E}\,\mathrm{tr}\left[(\mathbf{X}_{a,b}^{\intercal}\mathbf{X}_{a,b})^t\right] = 2\,\mathbb{E}\left\|\mathbf{X}_{a,b}\right\|_{2t}^{2t}$$

and

$$\begin{aligned}
\mathbf{V}_{a,b} &= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[(\overline{\mathbf{X}}_{a,b} - \overline{\mathbf{X}}_{a,b}^{(i)})^2 | Z] \\
&= \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[\begin{bmatrix} (\mathbf{X}_{a,b} - \mathbf{X}_{a,b}^{(i)})(\mathbf{X}_{a,b} - \mathbf{X}_{a,b}^{(i)})^{\intercal} & 0 \\ 0 & (\mathbf{X}_{a,b} - \mathbf{X}_{a,b}^{(i)})^{\intercal}(\mathbf{X}_{a,b} - \mathbf{X}_{a,b}^{(i)}) \end{bmatrix} | Z\right] \\
&= \frac{1}{2}\begin{bmatrix} \sum_{i=1}^{n} \mathbb{E}[(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^{\intercal} | Z] & 0 \\ 0 & \sum_{i=1}^{n} \mathbb{E}[(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^{\intercal}(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)}) | Z] \end{bmatrix}
\end{aligned}$$

We will use the following claim that we will prove later.

Claim 2.3.5. *We have the following relations.*

$$\sum_{i=1}^{n} \mathbb{E}[(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^{\intercal} | Z] = 2(b+1)\mathbf{F}_{a,b+1}\mathbf{F}_{a,b+1}^{\intercal}$$

$$\sum_{i=1}^{n} \mathbb{E}[(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})^{\intercal}(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)}) | Z] = 2(a+1)\mathbf{F}_{a+1,b}^{\intercal}\mathbf{F}_{a+1,b}$$

This gives $\mathbb{E}\operatorname{tr}\left[\mathbf{V}_{a,b}^{t}\right] = (b+1)^t \mathbb{E}\left\|\mathbf{F}_{a,b+1}\right\|_{2t}^{2t} + (a+1)^t \mathbb{E}\left\|\mathbf{F}_{a+1,b}\right\|_{2t}^{2t}$. Therefore, we get

$$\begin{aligned}
2\,\mathbb{E}\,\|\mathbf{X}_{a,b}\|_{2t}^{2t} &= \mathbb{E}\operatorname{tr}\left[\overline{\mathbf{X}}_{a,b}^{2t}\right] \\
&\leq (2(2t-1))^t\,\mathbb{E}\operatorname{tr}\left[\mathbf{V}_{a,b}^{t}\right] \\
&\leq (2(2t-1))^t((b+1)^t\,\mathbb{E}\,\|\mathbf{F}_{a,b+1}\|_{2t}^{2t} + (a+1)^t\,\mathbb{E}\,\|\mathbf{F}_{a+1,b}\|_{2t}^{2t}) \\
&\leq (2(2t-1))^t((b+1)^t\,\mathbb{E}\,\|\mathbf{X}_{a,b+1} + \mathbb{E}\,\mathbf{F}_{a,b+1}\|_{2t}^{2t} + (a+1)^t\,\mathbb{E}\,\|\mathbf{X}_{a+1,b} + \mathbb{E}\,\mathbf{F}_{a+1,b}\|_{2t}^{2t}) \\
&\leq (16t)^t((b+1)^t(\mathbb{E}\,\|\mathbf{X}_{a,b+1}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a,b+1}\|_{2t}^{2t}) + (a+1)^t(\mathbb{E}\,\|\mathbf{X}_{a+1,b}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a+1,b}\|_{2t}^{2t}) \\
&\leq (16td_p)^t(\mathbb{E}\,\|\mathbf{X}_{a,b+1}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a,b+1}\|_{2t}^{2t} + \mathbb{E}\,\|\mathbf{X}_{a+1,b}\|_{2t}^{2t} + \|\mathbb{E}\,\mathbf{F}_{a+1,b}\|_{2t}^{2t})
\end{aligned}$$

■

It remains to prove the claim.

Proof of Claim 2.3.5. We will prove the first equality. The second one is analogous. For $I \in \mathcal{I}, J \in \mathcal{J}, \alpha, \beta \in \{0,1\}^n$, we have

$$(\mathbf{F}_{a,b} - \mathbf{F}_{a,b}^{(i)})[(I,\alpha),(J,\beta)] = \begin{cases} \nabla_{\alpha+\beta}(f_{I,J}(Z) - f_{I,J}(Z^{(i)})) & \text{if } |\alpha|_0 = a, |\beta|_0 = b, \alpha \cdot \beta = 0 \\ 0 & \text{o.w.} \end{cases}$$

By Proposition 2.3.4, the first expression simplifies to $(Z_i - \widetilde{Z}_i)\nabla_{\mathbf{e}_i}\nabla_{\alpha+\beta}f_{I,J}(Z)$. Define the matrix $\mathbf{F}_{a,b,i}$ to be the matrix with the same set of rows and columns as $\mathbf{F}_{a,b}$ and whose

only nonzero entries are given by

$$\mathbf{F}_{a,b,i}[(I,\alpha),(J,\beta+\mathbf{e}_i)] = \nabla_{\mathbf{e}_i}\nabla_{\alpha+\beta} f_{I,J}(Z) \text{ if } |\alpha|_0 = a, |\beta|_0 = b, \beta\cdot\mathbf{e}_i = 0, \alpha\cdot(\beta+\mathbf{e}_i) = 0$$

Then, it's easy to see that $\sum_{i=1}^{n}\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}} = (b+1)\mathbf{F}_{a,b+1}\mathbf{F}_{a,b+1}^{T}$ and $(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^{\mathsf{T}} = (Z-\widetilde{Z}_i)^2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}}$. The latter equality implies

$$\mathbb{E}[(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^{\mathsf{T}}|Z] = \mathbb{E}[(Z_i-\widetilde{Z}_i)^2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}}|Z] = 2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}}$$

Therefore,

$$\sum_{i=1}^{n}\mathbb{E}[(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})(\mathbf{F}_{a,b}-\mathbf{F}_{a,b}^{(i)})^{\mathsf{T}}|Z] = 2\sum_{i=1}^{n}\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}} = 2(b+1)\mathbf{F}_{a,b+1}\mathbf{F}_{a,b+1}^{\mathsf{T}}$$

■

2.4 Applications

To illustrate our framework, we apply it to obtain concentration bounds for nonlinear random matrices that have been considered in the literature before. The first one is a simple tensor network that arose in the analysis of spectral algorithms for a variant of principal components analysis (PCA) [88, 81]. The second application is to obtain norm bounds on dense graph matrices [126, 1]. In the second application, the norm bounds are governed by a combinatorial structure called *the minimum vertex separator of a shape*. We will see how this notion arises naturally under our framework, while prior works that derived such bounds used the trace power method and required nontrivial combinatorial insights.

2.4.1 A simple tensor network

We consider the following result from [88, 81].

Lemma 2.4.1 ([81], Theorem 6.7.1)**.** *Let $c \in \{1, 2\}$ and let $d \geq 1$ be an integer. Let $\mathbf{A}_1, \ldots, \mathbf{A}_{n^c}$ be i.i.d. random matrices uniformly sampled from $\{-1, 1\}^{n^d \times n^d}$. Then, with probability $1 - O(n^{-100})$,*

$$\left\| \sum_{k \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k - \mathbb{E} \sum_{k \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k \right\| \leq C\sqrt{d} n^{(2d+c)/2} (\log n)^{1/2}$$

for an absolute constant $C > 0$.

Using our framework, we will prove a slightly relaxed version of the inequality where $\sqrt{d}(\log n)^{1/2}$ is replaced by $\log n$. We remark that we have not attempted to optimize these extra factors in front of the dominating term $n^{(2d+c)/2}$, so it's plausible that a more careful analysis can obtain a slightly better bound.

Proof of the relaxed bound. Let the i, j-th entry of $\mathbf{A}_k$ be $a_{k,i,j}$. Let $\mathbf{F} = \sum_{i \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k - \mathbb{E} \sum_{i \leq n^c} \mathbf{A}_k \otimes \mathbf{A}_k$ be a random matrix on the variables $a_{k,i,j}$ for $k \leq n^c, i, j \leq n^d$. So $\mathbb{E}\,\mathbf{F} = 0$ and we are looking for bounds on $\|\mathbf{F}\|$. The entries are given by

$$\mathbf{F}[(i_1, i_2), (j_1, j_2)] = \begin{cases} \sum_{k \leq n^c} a_{k,i_1,j_1} a_{k,i_2,j_2} & \text{if } (i_1, j_1) \neq (i_2, j_2) \\ 0 & \text{if } (i_1, j_1) = (i_2, j_2) \end{cases}$$

The nonzero entries are homogeneous polynomials of degree 2. Using Theorem 2.1.2,

$$\mathbb{E}\,\|\mathbf{F}\|_{2t}^{2t} \leq (32t)^{2t} (\left\|\mathbb{E}\,\mathbf{F}_{2,0}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{1,1}\right\|_{2t}^{2t} + \left\|\mathbb{E}\,\mathbf{F}_{0,2}\right\|_{2t}^{2t})$$

We will consider each of these terms. In the following arguments, we restrict attention to indices i_1, i_2, j_1, j_2 such that $(i_1, j_1) \neq (i_2, j_2)$.

1. $\mathbb{E}\,\mathbf{F}_{2,0}$ has nonzero entries in row $((i_1, i_2), \{(k, i_1, j_1), (k, i_2, j_2)\})$ and column (j_1, j_2) and all these entries are 1. The Schatten norm does not change when we permute the

rows and columns. So, we can group the rows on k, i_1, i_2 and within each group, we can sort j_1, j_2 in both rows and columns. We get a matrix having n^{2d+c} identity matrices, each of dimensions $n^{2d} \times n^{2d}$, stacked on top of each other. Using the definition, the Schatten-$2t$ norm of this matrix is easily computed to be $\left\|\mathbb{E}\,\mathbf{F}_{2,0}\right\|_{2t}^{2t} = n^{c+4d} n^{t(2d+c)}$.

2. $\mathbb{E}\,\mathbf{F}_{1,1}$ has nonzero entries in either row $((i_1, i_2), \{(k, i_1, j_1)\})$ and column $((j_1, j_2), \{(k, i_2, j_2)\})$; or row $((i_1, i_2), \{(k, i_2, j_2)\})$ and column $((j_1, j_2), \{(k, i_1, j_1)\})$ and all these entries are 1. So we can write $\mathbb{E}\,\mathbf{F}_{1,1} = \mathbf{A} + \mathbf{B}$ corresponding to the 2 sets of entries. Arguing just as in the previous case, we can obtain $\|\mathbf{A}\|_{2t}^{2t} = n^{c+4d} n^{t(2d+c)}$ where we group the rows on k, i_2, j_1 and $\|\mathbf{B}\|_{2t}^{2t} = n^{c+4d} n^{t(2d+c)}$ where we group the rows on k, i_1, j_2. Therefore, $\left\|\mathbb{E}\,\mathbf{F}_{1,1}\right\|_{2t}^{2t} \leq 2^{2t}(\|\mathbf{A}\|_{2t}^{2t} + \|\mathbf{B}\|_{2t}^{2t}) = 2^{2t+1} n^{c+4d} n^{t(2d+c)}$.

3. The case $\mathbb{E}\,\mathbf{F}_{0,2}$ is identical to $\mathbb{E}\,\mathbf{F}_{2,0}$.

Putting them together, $\mathbb{E}\,\|\mathbf{F}\|_{2t}^{2t} \leq (C't)^{2t} n^{c+4d} n^{t(2d+c)}$ for an absolute constant $C' > 0$. Now, we apply Markov's inequality to get

$$Pr[\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\| \geq \theta] \;\leq\; Pr[\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \geq \theta^{2t}] \;\leq\; \theta^{-2t}\,\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \;\leq\; \theta^{-2t}(C't)^{2t} n^{c+4d} n^{t(2d+c)}$$

We now set $\theta = \varepsilon^{-1/(2t)}(C't) n^{(c+4d)/t} n^{(2d+c)/2}$ to make this expression at most ε. Plug in $\varepsilon = n^{-100}$ and set $t = \log n$ to obtain that $\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\| \leq C n^{(2d+c)/2} \log n$ holds with probability $1 - n^{-100}$, where $C > 0$ is an absolute constant. ∎

2.4.2 Graph matrices

In this section, we first define graph matrices and then show how to obtain norm bounds for *dense graph matrices*, i.e. the case when $G \sim \mathcal{G}_{n,1/2}$, using our framework. Handling *sparse graph matrices*, i.e. the case when $G \sim \mathcal{G}_{n,p}$ for $p = o(1)$, may not work well with our basic framework as we will explain in Section 2.5. Instead, our general framework in Section 2.6 will handle this case well and we obtain sparse graph matrix norm bounds in Section 2.8.

Definitions

Define by $\mathcal{G}_{n,p}$ the Erdős-Rényi random graph on the vertex set $[n]$ with n vertices, where each edge is present independently with probability p. Let the graph be encoded by variables $G_{i,j} \in \Omega = \{-\sqrt{\frac{1-p}{p}}, \sqrt{\frac{p}{1-p}}\}$ where $-\sqrt{\frac{1-p}{p}}$ indicates the presence of the edge $\{i,j\}$ and $\sqrt{\frac{p}{1-p}}$ indicates absence, for all $1 \leq i,j \leq n$.

So, each $G_{i,j}$ for $i < j$ is sampled from Ω where $G_{i,j}$ takes the value $-\sqrt{\frac{1-p}{p}}$ with probability p and takes the value $\sqrt{\frac{p}{1-p}}$ otherwise. Here, Ω has been normalized so that $\mathbb{E}_{x\sim\Omega}[x] = 0, \mathbb{E}_{x\sim\Omega}[x^2] = 1$. as is standard in p-biased Fourier analysis.

When $p = 1/2$, we are in the setting of *dense graph matrices*. Then, $\mathcal{G}_{n,1/2}$ can be thought of as a sampling of the $G_{i,j}, i < j$ independently and uniformly from $\Omega = \{-1, 1\}$.

For a set of edges $E \subseteq \binom{[n]}{2}$, define $G_E := \prod_{e\in E} G_e$. When $p = 1/2$, the G_E correspond to the Fourier basis for functions of the graph.

Define $\mathcal{I}$ to be the set of sub-tuples of $[n]$, including the empty tuple. Graph matrices will have rows and columns indexed by $\mathcal{I}$. Each graph matrix has a succinct representation as a graph with some extra information, that is called a *shape*.

Definition 2.4.2 (Shape)**.** *A shape is a tuple $\tau = (V(\tau), E(\tau), U_\tau, V_\tau)$ where $(V(\tau), E(\tau))$ is a graph and U_τ, V_τ are ordered subsets of the vertices.*

Definition 2.4.3 (Realization)**.** *Given a shape τ, a realization of τ is an injective map $\varphi : V(\tau) \to [n]$.*

Definition 2.4.4 (Graph matrices)**.** *Let τ be a shape. The graph matrix $\mathbf{M}_\tau : \{\pm 1\}^{\binom{n}{2}} \to \mathbb{R}^{\mathcal{I}\times\mathcal{I}}$ is defined to be the matrix-valued function with I, J-th entry defined as follows.*

$$\mathbf{M}_\tau[I,J] := \sum_{\substack{\text{Realization } \varphi \\ \varphi(U_\tau)=I, \varphi(V_\tau)=J}} G_{\varphi(E(\tau))} = \sum_{\substack{\text{Realization } \varphi \\ \varphi(U_\tau)=I, \varphi(V_\tau)=J}} \prod_{(u,v)\in E(\tau)} G_{\varphi(u),\varphi(v)}$$

In other words, we sum over all realizations of τ that map U_τ, V_τ to I, J respectively and

for each such realization, we have a term corresponding to the Fourier character that the realization gives.

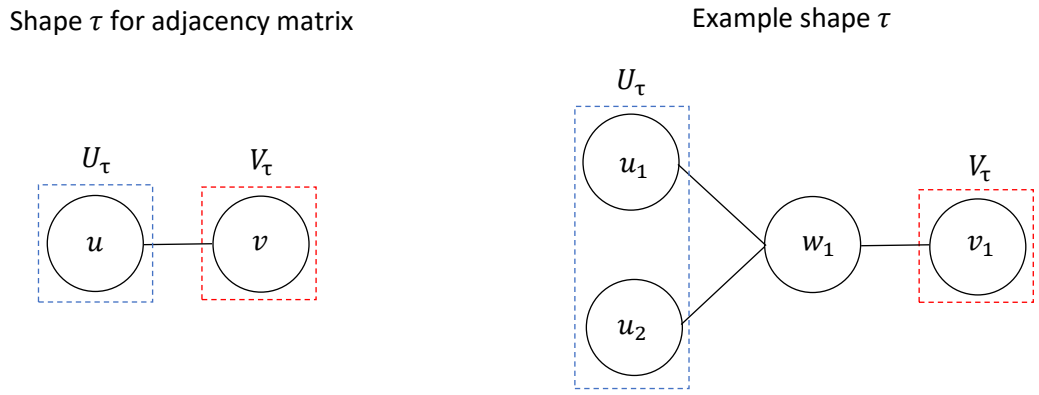

Figure 2.3: Left: Shape corresponding to adjacency matrix, Right: Example of a more complicated shape

The following examples illustrate some simple graph matrices.

Example 2.4.5 (Adjacency matrix)**.** *Let τ be the shape on the left in Fig. 2.3, with two vertices $V(\tau) = \{u, v\}$ and a single edge $E(\tau) = \{\{u, v\}\}$. U_τ, V_τ are $(u), (v)$ respectively where we use tuples to indicate ordering. Then $\mathbf{M}_\tau$ has nonzero entries $\mathbf{M}_\tau[(i), (j)](G) = G_{i,j}$ for all $i \neq j$. If $G \in \{\pm 1\}^{\binom{n}{2}}$ is thought of as a graph, then $\mathbf{M}_\tau$ has as principal submatrix the ± 1 adjacency matrix of G with zeros on the diagonal, and the other entries are 0.*

Example 2.4.6. *In Fig. 2.3, consider the shape τ on the right. We have $U_\tau = (u_1, u_2), V_\tau = (v_1), V(\tau) = \{u_1, u_2, v_1, w_1\}$ and $E(\tau) = \{\{u_1, w_1\}, \{u_2, w_1\}, \{w_1, v_1\}\}$. $\mathbf{M}_\tau$ is a matrix with rows and columns indexed by sub-tuples of $[n]$. Its nonzero entries are in rows I and columns J with $|I| = |U_\tau| = 2$ and $|J| = |V_\tau| = 1$ respectively. Specifically, for all distinct a_1, a_2, b_1, the entry corresponding to row (a_1, a_2) and column (b_1) is $\sum_{c_1 \in [n] \setminus \{a_1, a_2, b_1\}} G_{a_1, c_1} G_{a_2, c_1}, G_{c_1, b_1}$. Here, each term is obtained via the realization φ that maps u_1, u_2, w_1, v_1 to a_1, a_2, c_1, b_1 re-*

spectively. Succinctly,

$$\mathbf{M}_\tau = \begin{array}{c} \\ \textit{row } (a_1, a_2) \rightarrow \end{array} \begin{array}{c} \textit{column } (b_1) \\ \downarrow \\ \begin{pmatrix} \vdots \\ \cdots\cdots \sum_{c_1 \in [n] \setminus \{a_1, a_2, b_1\}} G_{a_1,c_1} G_{a_2,c_1}, G_{c_1,b_1} \cdots\cdots\cdots \\ \vdots \end{pmatrix} \end{array}$$

Intuitively, graph matrices are symmetrizations of the Fourier basis, where the symmetry is incorporated by summing over all realizations of "free" vertices $V(\tau) \setminus U_\tau \setminus V_\tau$ of the shape τ. For more examples of graph matrices and why they can be a useful tool to work with, see [1].

Norm bounds for dense graph matrices

In this section, we study the concentration of the so-called "dense graph matrices" which is a term that refers to graph matrices M_τ in the setting $p = 1/2$. Since the edges of a random graph sampled from $\mathcal{G}_{n,1/2}$ can be viewed as independent Rademacher random variables, we can apply our framework in this setting.

In particular, we will obtain bounds on $\mathbb{E} \|\mathbf{M}_\tau - \mathbb{E}\, \mathbf{M}_\tau\|_{2t}^{2t}$. The $G_{i,j} \in \{-1, 1\}$ correspond to the Z_is in Section 2.3 and for a fixed shape τ, $\mathbf{M}_\tau$ will be the matrix $\mathbf{F}$ we are interested in analyzing. For $I, J \in \mathcal{I}$, $\mathbf{M}_\tau[I, J]$ is a nonzero polynomial only when there exists at least one realization of τ that maps U_τ, V_τ to I, J respectively. In particular, we must have $|I| = |U_\tau|$ and $|J| = |V_\tau|$. In this case, $\mathbf{M}_\tau[I, J]$ is a homogenous polynomial of degree $|E(\tau)|$.

By Theorem 2.1.2, we have

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \;\leq\; \sum_{\substack{a+b\geq 1\\ a,b\geq 0}} (16t|E(\tau)|)^{(a+b)t}\, \left\|\mathbb{E}\,\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t}$$

where for integers $a, b \geq 0$, $\mathbf{M}_{\tau,a,b}$ is defined to be the matrix with rows and columns each indexed by $\mathcal{I} \times \{0,1\}^{\binom{n}{2}}$ such that for all $I, J \in \mathcal{I}$, we have

$$\mathbf{M}_{\tau,a,b}[(I,\alpha),(J,\beta)] \;=\; \begin{cases} \nabla_{\alpha+\beta}\mathbf{M}_\tau[I,J] & \text{if } |\alpha|_0 = a, |\beta|_0 = b, \alpha\cdot\beta = 0 \\ 0 & \text{o.w.} \end{cases}$$

For any multilinear homogenous polynomial f of degree d, since $\mathbb{E}[G_{i,j}] = 0$ for all i, j, we have $\nabla_\alpha f = 0$ whenever $|\alpha|_0 < d$. Therefore, $\mathbb{E}\,\mathbf{M}_{\tau,a,b} = 0$ for all $a + b < |E(G)|$. Moreover, $\mathbb{E}\,\mathbf{M}_{\tau,a,b} = 0$ whenever $a + b \neq |E(G)|$ otherwise $\mathbb{E}\,\mathbf{M}_{\tau,a,b} = \mathbf{M}_{\tau,a,b}$. So, we can further simplify the above expression to

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \;\leq\; \sum_{\substack{a+b=|E(\tau)|\\ a,b\geq 0}} (16t|E(\tau)|)^{|E(\tau)|t}\, \left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t}$$

It remains to analyze $\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t}$ for $a + b = |E(G)|$. We will see that analyzing these matrices is much simpler since they are deterministic matrices and simple computations using the Frobenius norm bound will work well. To state our final bounds, we need to define the notion of vertex separators of shapes.

Remark 2.4.7. *As we will see, when analyzing the Frobenius norms for these deterministic matrices, the notion of the minimum vertex separator arises naturally. In prior trace method calculations (e.g. [126], [1]), this required ingenious combinatorial observations.*

Definition 2.4.8 (Vertex separator)**.** *For a shape τ, define a vertex separator to be a subset of vertices $S \subseteq V(\tau)$ such that there is no path from U_τ to V_τ in $\tau \setminus S$, which is the shape*

obtained by deleting all the vertices of S (including all edges they're incident on).

For a shape τ, denote by S_τ a vertex separator of the smallest size. Also, let I_τ be the set of isolated vertices (vertices with degree 0) in $V(\tau) \setminus U_\tau \setminus V_\tau$, so the presence of these vertices essentially scale the matrix by a scalar factor.

Theorem 2.4.9. *For a shape τ and any integer $t \geq 1$,*

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \leq \left(C^{t|E(\tau)|} n^{|V(\tau)|} t^{t|E(\tau)|} |E(\tau)|^{2t|E(\tau)|}\right) n^{t(|V(\tau)|-|S_\tau|+|I_\tau|)}$$

for an absolute constant $C > 0$.

Upto lower order terms, the same result has been shown before in [126, 1]. To interpret this bound, assume that τ has a constant number of vertices. By setting $t \approx \text{polylog}(n)$, we get

$$\|\mathbf{M}_\tau\| = \widetilde{\mathrm{O}}\left(\sqrt{n}^{|V(\tau)|-|S_\tau|+|I_\tau|}\right)$$

with high probability, where $\widetilde{\mathrm{O}}$ hides logarithmic factors. This is obtained by applying Markov's inequality on the bound on $\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t}$. If τ has at least one edge, then $\mathbb{E}\,\mathbf{M}_\tau = 0$ and Theorem 2.4.9 yields such bounds. If τ has no edges, then it's quite simple to obtain such a bound and we include it in Lemma 2.4.10 for the sake of completeness. Corollary 2.4.11 makes precise the high probability bound above. Therefore, this power of n is essentially what controls the norm bound and this is utilized heavily in applications (e.g. [13**?** , 149]).

Proof of Theorem 2.4.9. We first argue that we can assume $I_\tau = \emptyset$. This is because of the following reason. Each distinct vertex in τ of degree 0 essentially scales the matrix by a factor of at most n. And in the right hand side of the inequality, each vertex in I_τ contributes a factor of n^{2t} accordingly, from $n^{t|V(\tau)|}$ and from $n^{t|I_\tau|}$, and the other changes only weaken the inequality.

Now, fix $a, b \geq 0$ such that $a + b = |E(\tau)|$ and consider $\mathbf{M}_{\tau,a,b}$. For $I, J \in \mathcal{I}, \alpha, \beta \in \{0,1\}^{\binom{n}{2}}$ such that $|\alpha|_0 = a, |\beta|_0 = b, \alpha \cdot \beta = 0$, by definition,

$$\begin{aligned}\mathbf{M}_{\tau,a,b}[(I,\alpha),(J,\beta)] &= \nabla_{\alpha+\beta}\left(\sum_{\varphi:\varphi(U_\tau)=I,\varphi(V_\tau)=J}\prod_{u,v\in E(\tau)} G_{\varphi(u),\varphi(v)}\right)\\ &= |\{\varphi \mid \varphi(U_\tau) = I, \varphi(V_\tau) = J, \varphi(E(\tau)) = \mathrm{Supp}(\alpha+\beta)\}|\end{aligned}$$

where Supp(.) denotes the support. We will now obtain norm bounds on these deterministic matrices by reinterpreting them as graph matrices for different shapes.

Let $P = (E_1, E_2)$ denote the partition of $E(\tau) = E_1 \sqcup E_2$ into two ordered sets E_1, E_2, where $\sqcup$ denotes disjoint union. Then, we can write $\mathbf{M}_{\tau,a,b} = \sum_{P\in\mathcal{P}} \mathbf{M}_{\tau,a,b,P}$ where

$$\mathbf{M}_{\tau,a,b,P}[(I,\alpha),(J,\beta)] = |\{\varphi \mid \varphi(U_\tau) = I, \varphi(V_\tau) = J, \varphi(E_1) = \mathrm{Supp}(\alpha), \varphi(E_2) = \mathrm{Supp}(\beta)\}|$$

Let the set of ordered partitions P be $\mathcal{P}$. Then, $|\mathcal{P}| \leq (4|E(\tau)|)^{|E(\tau)|}$ and so, by Fact 2.2.3,

$$\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} \leq (4|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \left\|\mathbf{M}_{\tau,a,b,P}\right\|_{2t}^{2t}$$

Each $\mathbf{M}_{\tau,a,b,P}$ can be interpreted as a graph matrix for a different shape τ_P, with the same vertex set and no edges. Let $V(\tau_P) = V(\tau), E(\tau_P) = \emptyset$ and set $U_{\tau_P} = U_\tau \cup V(E_1), V(\tau_P) = V_\tau \cup V(E_2)$ using a canonical ordering. Then, $\mathbf{M}_{\tau,a,b}$ is equal to $\mathbf{M}_{\tau_P}$ upto renaming of the rows and columns. For an illustration, see Fig. 2.4.

This graph matrix has a block diagonal structure indexed by the realizations of the set of common vertices $S = U_{\tau_P} \cap V_{\tau_P}$. Indeed, for $K \in [n]^S$, let $\mathbf{M}_{\tau_P,K}$ be the block of $\mathbf{M}_{\tau_P}$

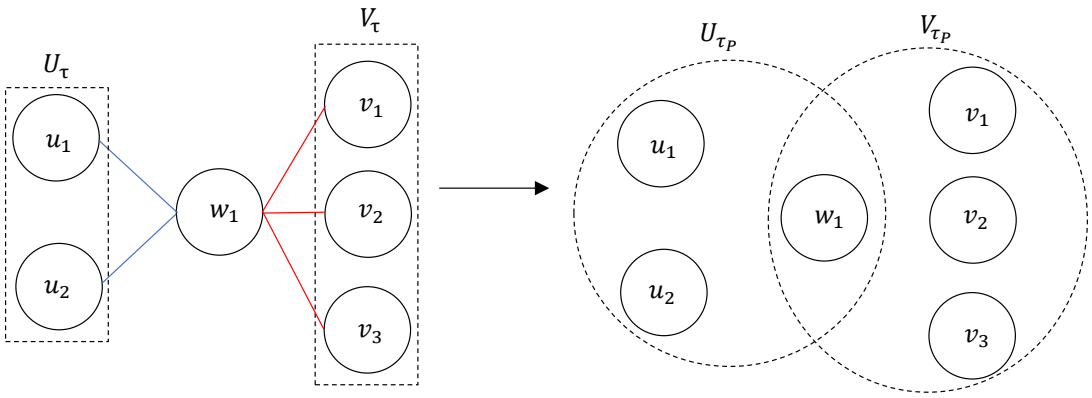

Figure 2.4: An example illustrating how τ_P is defined. In this example, P constraints the blue and red edges to go to α and β respectively. U_{τ_P}, V_{τ_P} have an ordering on the vertices (not shown here).

with $\varphi(S) = K$. Then, $\mathbf{M}_{\tau_P,K}\mathbf{M}_{\tau_P,K'}^\intercal = \mathbf{M}_{\tau_P,K}^\intercal\mathbf{M}_{\tau_P,K'} = 0$ for $K \neq K'$ and so,

$$\begin{aligned}
\mathbb{E}\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} &\leq (4|E(\tau)|)^{t|E(\tau)|}\sum_{P\in\mathcal{P}}\left\|\mathbf{M}_{\tau_P}\right\|_{2t}^{2t} \\
&= (4|E(\tau)|)^{t|E(\tau)|}\sum_{P\in\mathcal{P}}\sum_{T\in[n]^S}\left\|\mathbf{M}_{\tau_P,T}\right\|_{2t}^{2t} \\
&\leq (4|E(\tau)|)^{t|E(\tau)|}\sum_{P\in\mathcal{P}}\sum_{T\in[n]^S}\left(\left\|\mathbf{M}_{\tau_P,T}\right\|_2^2\right)^t
\end{aligned}$$

where we bounded the Schatten norm by the appropriate power of the Frobenius norm.

For any fixed $K \in [n]^S$, the entries of $\mathbf{M}_{\tau_P,K}$ take values in $\{0,1\}$ and the number of nonzero entries is at most $n^{|V(\tau)|-|S|}$ because the realizations of vertices in S are fixed and the other vertices have at most n choices each. Therefore, $\left\|\mathbf{M}_{\tau_P,K}\right\|_2^2 \leq n^{|V(\tau)|-|S|}$.

Finally, we bound $|S|$ to estimate how large this term can be over all possibilities of P. We argue that S blocks all paths from U_τ to V_τ. To see this, consider any path from U_τ to V_τ, it must contain an edge $(u,v) \in E(\tau)$ such that $u \in U_{\tau_P}, v \in V_{\tau_P}$. We must either have $(u,v) \in E_1$, in which case $u,v \in U_{\tau_P}$ and $v \in S$, or $(u,v) \in E_2$, in which case $u,v \in V_{\tau_P}$ and $u \in S$. In either case, S must contain either u or v. This argument implies S must be a vertex separator of τ, giving $|S| \geq |S_\tau|$. For a proof by picture, see Fig. 2.5.

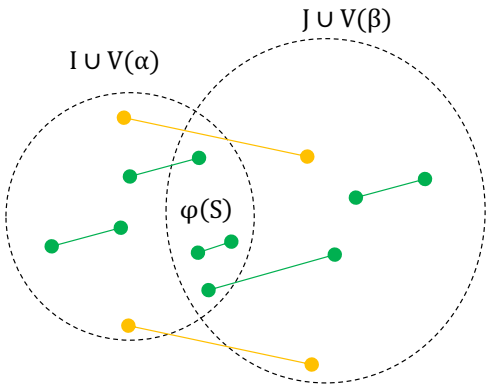

Figure 2.5: Proof by picture that $|S| \geq |S_\tau|$. Green edges can occur in τ, orange edges cannot, so S blocks all paths from U_τ to V_τ.

We also have the trivial upper bound $|S| \leq |V(\tau)|$. Ultimately, this gives

$$\begin{aligned}\left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} &\leq (4|E(\tau)|)^{t|E(\tau)|} \sum_{P\in\mathcal{P}} \sum_{T\in[n]^S} n^{t(|V(\tau)|-|S_\tau|)} \\ &\leq (4|E(\tau)|)^{t|E(\tau)|}(4|E(\tau)|)^{|E(\tau)|}{}_n{}^{|V(\tau)|}{}_n{}^{t(|V(\tau)|-|S_\tau|)}\end{aligned}$$

Along with our prior discussion, we get

$$\begin{aligned}\mathbb{E}\left\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t} &\leq \sum_{a+b=|E(\tau)|} (16t|E(\tau)|)^{|E(\tau)|t} \left\|\mathbf{M}_{\tau,a,b}\right\|_{2t}^{2t} \\ &\leq \sum_{a+b=|E(\tau)|} (16t|E(\tau)|)^{|E(\tau)|t}(4|E(\tau)|)^{t|E(\tau)|}(4|E(\tau)|)^{|E(\tau)|}{}_n{}^{|V(\tau)|}{}_n{}^{t(|V(\tau)|-|S_\tau|)} \\ &\leq \left(C^{t|E(\tau)|}{}_n{}^{|V(\tau)|}t^{t|E(\tau)|}|E(\tau)|^{2t|E(\tau)|}\right) n^{t(|V(\tau)|-|S_\tau|)}\end{aligned}$$

for an absolute constant $C > 0$. ∎

In the proof above, our analysis of the shape τ_P which has no edges, applies in general to any shape τ with no edges. For the sake of completeness, we state it explicity in the following lemma.

Lemma 2.4.10. *For a shape τ with no edges and any integer $t \geq 1$,*

$$\mathbb{E} \|\mathbf{M}_\tau\|_{2t}^{2t} \leq n^{|U_\tau \cap V_\tau|} n^{t(V(\tau)-|U_\tau \cap V_\tau|+|I_\tau|)}$$

Note that this has the same form as Theorem 2.4.9 because for a shape τ with no edges, the minimum vertex separator S_τ is just $U_\tau \cap V_\tau$.

The following corollary obtains high probability norm bounds for norms of graph matrices via Markov's inequality.

Corollary 2.4.11. *For a shape τ, for any constant $\varepsilon > 0$, with probability $1 - \varepsilon$,*

$$\|\mathbf{M}_\tau\| \leq (C|E(\tau)| \log(n^{|V(\tau)|}/\varepsilon))^{|E(\tau)|} \cdot \sqrt{n}^{|V(\tau)|-|S_\tau|+|I_\tau|}$$

for an absolute constant $C > 0$.

Proof. If $E(\tau) = \emptyset$, we invoke Lemma 2.4.10. Otherwise, $\mathbb{E}\, \mathbf{M}_\tau = 0$ and we invoke Theorem 2.4.9. By an application of Markov's inequality,

$$\begin{aligned} Pr[\|\mathbf{M}_\tau\| \geq \theta] &\leq Pr[\|\mathbf{M}_\tau\|_{2t}^{2t} \geq \theta^{2t}] \\ &\leq \theta^{-2t}\, \mathbb{E}\, \|\mathbf{M}_\tau\|_{2t}^{2t} \\ &\leq \theta^{-2t} \left((C')^{t|E(\tau)|} n^{|V(\tau)|} t^{t|E(\tau)|} |E(\tau)|^{2t|E(\tau)|} \right) n^{t(|V(\tau)|-|S_\tau|+|I_\tau|)} \end{aligned}$$

for an absolute constant $C' > 0$. We now set

$$\theta = \left(\varepsilon^{-1/(2t)} (C'')^{|E(\tau)|} n^{|V(\tau)|/(2t)} t^{|E(\tau)|/2} |E(\tau)|^{|E(\tau)|} \right) \sqrt{n}^{|V(\tau)|-|S_\tau|+|I_\tau|}$$

for an absolute constant $C'' > 0$, to make this expression at most ε. Set $t = \frac{1}{2} \log(n^{|V(\tau)|}/\varepsilon)$ to complete the proof. ∎

2.5 Why a naïve application of [147] may fail for general product distributions

In this section, we elaborate on the difficulties that arise when working with random variables that are not necessarily Rademacher. In this case, note that we cannot assume that the polynomial entries are multilinear as well.

To recall the setting, we are given a random matrix $\mathbf{F}$ whose entries are low degree polynomials in random variables $Z_1, \ldots, Z_n$ which are independently sampled from arbitrary distributions. And we wish to obtain concentration bounds on how much $\mathbf{F}$ can deviate from its mean, by way of controlling $\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t}$.

Building on the ideas from Section 2.3, we could attempt to use matrix Efron-Stein, Theorem 2.1.1 and hope to obtain a similar recursion framework. We now discuss what happens if we do this. Assume $\mathbb{E}[Z_i] = 0, \mathbb{E}[Z_i^2] = 1$. We can proceed similar to the proof of Theorem 2.1.2. So, we consider $\mathbf{X}$ as a principal submatrix of $\mathbf{X}_{0,0}$ and follow through Lemma 2.3.3. The main change will happen in Claim 2.3.5. In particular, the equation $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2|Z] = 2$ is no longer true. Instead, we will have $\mathbb{E}[(Z_i - \widetilde{Z}_i)^2|Z] = 1 + Z_i^2$. So, we get the expression

$$\sum_{i=1}^{n}(1 + Z_i^2)\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\intercal} = \sum_{i=1}^{n}\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\intercal} + \sum_{i=1}^{n} Z_i^2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\intercal}$$

The first term can been handled just as in the basic framework. Unfortunately, the second term will be a source of difficulty. To get around this difficulty, we could attempt to apply the matrix Efron-Stein inequality again on an appropriately constructed matrix. To do this, we can interpret the second term as having been obtained after differentiating with respect to the variable Z_i and then *putting the variable back*. In contrast, we didn't need to put it back when working with Rademacher random variables. But after we do this, when we recurse on these extra matrices, the new second term will contain the left hand side as a

sub-term, thereby giving a trivial inequality and stalling the recursion.

To see this more clearly, consider the simplest case $a = b = 0$. Then, the first term $\sum_{i=1}^{n} \mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}}$ will be equal to $\mathbf{F}_{0,1}\mathbf{F}_{0,1}^{\mathsf{T}}$ as we saw earlier. To evaluate the second term $\sum_{i=1}^{n} Z_i^2\mathbf{F}_{a,b,i}\mathbf{F}_{a,b,i}^{\mathsf{T}}$ in a similar manner, we define the matrix $\mathbf{H}$ to be the same as $\mathbf{F}_{0,1}$ except that each entry is now multiplied by Z_i where i is the differentiated variable in the column. That is, $\mathbf{H}[I,(J,\mathbf{e}_i)] = Z_i\mathbf{F}_{0,1}[I,(J,\mathbf{e}_i)]$. Observe that in the definition of $\mathbf{H}$, Z_i has been put back after differentiating with respect to it. Then, the second term will be $\mathbf{H}\mathbf{H}^{\mathsf{T}}$ and we can hope to use Efron-Stein again on this matrix $\mathbf{H}$ recursively.

We could do that and proceed similarly to the proof of Lemma 2.3.3 with appropriate modifications as above. But since $\beta_i = 1$ already, differentiating with respect to Z_i and putting it back, will return the same matrix $\mathbf{H}$! So, we end up with an inequality of the form

$$\mathbb{E}\,\|\mathbf{H}\|_{2t}^{2t} \leq \mathrm{O}(t)^t(\mathbb{E}\,\|\mathbf{H}\|_{2t}^{2t} + \text{ other nonnegative terms})$$

Indeed, this is a tautology and will not be useful to us.

For a quick and dirty bound, suppose we had a parameter L such that $1 + Z_i^2 \leq L$ for our distributions, then we will be able to obtain a similar framework while incurring a loss of $\sqrt{L}$ at each step of the recursion. But unfortunately, this bound will be lossy. For example, if we do this computation for the centered normalized adjacency matrix of $G \sim \mathcal{G}_{n,p}$, we will obtain a norm bound of $\widetilde{\mathrm{O}}(\frac{\sqrt{n(1-p)}}{\sqrt{p}})$ where $\widetilde{\mathrm{O}}$ hides logarithmic factors.. This bound is tight for constant or even inverse polylogarithmic p. But for $p = n^{-\theta}$ for some constant $0 < \theta < 1$, this is not tight because in this regime, the true norm bound is known to be $\widetilde{\mathrm{O}}(\sqrt{n})$ (see the early works of [67, 182] and for tighter bounds, see [19] and references therein).

If we dig into the details of what happened, this example illustrates that the matrix Efron-Stein inequality Theorem 2.1.1 becomes a tautology for certain kinds of matrices, that yield $\mathbf{V} = \mathrm{O}(1)\mathbf{X}\mathbf{X}^{\mathsf{T}} +$ other positive semidefinite matrices.

But in our framework in general, the aforementioned bad matrices occur when we differ-

entiate with respect to variables that have already been differentiated on. In other words, the current definition of the variance proxy $\mathbf{V}$ doesn't take into account whether we have already differentiated with respect to some variable Z_i. So, for the general recursion, we dive into the proof due to [147] and modify it using structural properties of the intermediate matrices we obtain in our framework.

2.6 The general recursion framework

We now assume $Z_1, \ldots, Z_n$ are i.i.d. random variables sampled from a distribution Ω with finite moments. We assume that they are identically distributed for simplicity but our technique easily extends even when they are not identically distributed, as long as they are independent. For each $i \leq n$, define $\widetilde{Z}_i$ to be an independent copy of Z_i and define the vector $Z^{(i)} := (Z_1, \ldots, Z_{i-1}, \widetilde{Z}_i, Z_{i+1}, \ldots, Z_n)$. Define Z' to be the random vector defined by sampling i from $[n]$ uniformly at random and then setting $Z' = Z^{(i)}$.

Let $\mathbf{F} \in \mathbb{R}[Z]^{\mathcal{I} \times \mathcal{J}}$ be a matrix with rows and columns indexed by arbitrary sets $\mathcal{I}, \mathcal{J}$ respectively such that for all $I \in \mathcal{I}, J \in \mathcal{J}$, $\mathbf{F}[I, J]$ are polynomials of $Z_1, \ldots, Z_n$. Let d_p the maximum degree of $\mathbf{F}[I, J]$ over all entries I, J and let d be the maximum degree of Z_i over all entries $\mathbf{F}[I, J]$ and $i \leq n$.

Similar to the Rademacher case, let $\mathbf{X} := \mathbf{F} - \mathbb{E}\,\mathbf{F}$. When the input is Z, we denote the matrices as $\mathbf{F}, \mathbf{X}$, etc and when the input is $Z^{(i)}$, denote the corresponding matrices as $\mathbf{F}^{(i)}, \mathbf{X}^{(i)}$, etc. In this section, we will give a general framework using which we can obtain bounds on $\mathbb{E}\,\|\mathbf{F} - \mathbb{E}\,\mathbf{F}\|_{2t}^{2t}$ for any integer $t \geq 1$.

We set up a few preliminaries in order to state the main theorem.

Definition 2.6.1 (Space $\mathcal{S}$)**.** *Let $\mathcal{S}$ be the space of mean-zero polynomials in $Z_1, \ldots, Z_n$ of degree at most d_p.*

For $\alpha \neq 0$, we also define the centered monomials

$$\chi_\alpha(Z) = \prod_{\alpha_i > 0} (Z_i^{\alpha_i} - \mathbb{E}[Z_i^{\alpha_i}])$$

By definition, $\chi_\alpha \in \mathcal{S}$ for all $\alpha \neq 0, |\alpha|_1 \leq d_p$. The following proposition is straightforward.

Proposition 2.6.2. *The set* $\{\chi_\alpha(Z) | 1 \leq |\alpha|_1 \leq d_p\}$ *forms a basis for* $\mathcal{S}$.

For the general framework, we work over this basis because as we will see in Section 2.7, the "inner kernel matrix" is convenient to state in this basis. The ∇ operator also works nicely with our polynomials χ_β. Indeed, observe that $\nabla_\alpha(\chi_\beta) = \begin{cases} \chi_{\beta-\alpha} & \text{if } \alpha \trianglelefteq \beta \\ 0 & \text{o.w.} \end{cases}$.

For a polynomial $f(Z)$ in $\mathcal{S}$, denote by $\widehat{f}(\alpha)$ the coefficient of $\chi_\alpha(Z)$ in the expansion of f, that is,

$$f(Z) = \sum_{0 \neq \alpha \in \mathbb{N}^n} \widehat{f}(\alpha) \chi_\alpha(Z)$$

We can naturally extend this notation to matrices that have mean 0. So, we can write $\mathbf{X} = \sum_{\alpha \neq 0} \widehat{\mathbf{X}}(\alpha) \chi_\alpha(Z)$ where $\widehat{\mathbf{X}}(\alpha)$ are deterministic matrices. In order to apply our recursion framework, we group this sum into terms based on $|\alpha|_0$. For $k \geq 1$, define $\mathbf{X}_k = \sum_{|\alpha|_0 = k} \widehat{\mathbf{X}}(\alpha) \chi_\alpha(Z)$. Then,

$$\mathbf{X} = \sum_{k \geq 1} \mathbf{X}_k$$

Note that when $k > d_p$, $\mathbf{X}_k = 0$.

Definition 2.6.3 (Indexing set $\mathcal{K}$)**.** *We define* $\mathcal{K} \subseteq \mathbb{N}^n \times \{0,1\}^n$ *to be the set of pairs* (α, γ) *such that* $|\alpha|_1 \leq d_p, \alpha \in \mathbb{N}^n$ *and* $\gamma \leq \alpha, \gamma \in \{0,1\}^n$.

Define the diagonal matrix $\mathbf{D}_1 \in \mathbb{R}[Z]^{\mathcal{I} \times \mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{I} \times \mathcal{K}}$ with nonzero entries

$$\mathbf{D}_1[(I, \alpha, \gamma), (I, \alpha, \gamma)] = \sqrt{\mathbb{E}[Z^{2\alpha \cdot (1-\gamma)}]} Z^{\alpha \cdot \gamma}$$

Similarly, define the diagonal matrix $\mathbf{D}_2 \in \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}}$ with nonzero entries

$$\mathbf{D}_2[(J,\alpha,\gamma),(J,\alpha,\gamma)] = \sqrt{\mathbb{E}[Z^{2\alpha\cdot(1-\gamma)}]}Z^{\alpha\cdot\gamma}$$

Definition 2.6.4 (Matrices $\mathbf{G}_{k,a,b}, \mathbf{F}_{k,a,b}$)**.** *For integers k,a,b such that $k \geq 1, a, b \geq 0$, define the matrix $\mathbf{G}_{k,a,b}$ to have rows and columns indexed by $\mathcal{I}\times\mathcal{K}$ and $\mathcal{J}\times\mathcal{K}$ respectively such that for all $(I,\alpha_1,\gamma_1) \in \mathcal{I}\times\mathcal{K}$, $(J,\alpha_2,\gamma_2) \in \mathcal{J}\times\mathcal{K}$,*

$$\mathbf{G}_{k,a,b}[(I,\alpha_1,\gamma_1),(J,\alpha_2,\gamma_2)] = \begin{cases} \nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J] & \text{if } |\alpha_1|_0 = a, |\alpha_2|_0 = b, \alpha_1\cdot\alpha_2 = 0 \\ 0 & o.w. \end{cases}$$

Also, define $\mathbf{F}_{k,a,b} := \mathbf{D}_1\mathbf{G}_{k,a,b}\mathbf{D}_2$.

Note that when $k > d_p$, $\mathbf{F}_{k,a,b} = 0$.

Proposition 2.6.5. *For integers k,a,b such that $k \geq 1, a, b \geq 0$, suppose $a+b<k$. Then each nonzero entry f of $\mathbf{G}_{k,a,b}$ has the property that $\widehat{f}(\alpha)$ is nonzero only when $|\alpha|_0 = k-a-b$*

Proof. The nonzero entries of $\mathbf{X}_k$ only has terms containing exactly k variables and $\nabla_{\alpha_1+\alpha_2}$ either zeroes out the term, or it truncates exactly $|\alpha_1+\alpha_2|_0 = |\alpha_1|_0+|\alpha_2|_0 = a+b$ variables. ∎

This also immediately implies that $\mathbb{E}[\mathbf{G}_{k,a,b}] = 0$ whenever $a+b<k$. Finally, when $k = a+b$, we have that $\mathbf{G}_{k,a,b}$ is a deterministic matrix independent of the Z_i. These give rise to the matrices $\mathbf{F}_{a+b,a,b}$ that appears in our main theorem.

We are now ready to state the main theorem.

Theorem 2.6.6 (General recursion)**.** *Let the tuple of random variables Z and the function $\mathbf{F}$ be as above. Then, for all integers $t \geq 1$,*

$$\mathbb{E}\,\|\mathbf{F}-\mathbb{E}\,\mathbf{F}\|_{2t}^{2t} \leq \sum_{a,b\geq 0, a+b\geq 1} (Ct^2 dd_p^4)^{(a+b)t}\,\mathbb{E}\,\left\|\mathbf{F}_{a+b,a,b}\right\|_{2t}^{2t}$$

for an absolute constant $C > 0$.

Note that $\mathbf{F}_{a+b,a,b} = \mathbf{D}_1 \mathbf{G}_{a+b,a,b} \mathbf{D}_2$ where $\mathbf{D}_1, \mathbf{D}_2$ are diagonal matrices and $\mathbf{G}_{a+b,a,b}$ is a deterministic matrix that's independent of Z. To analyze the expected Schatten norm of such matrices, we can resort to far simpler techniques. For instance, we can obtain a simple bound using an appropriate power of the Frobenius norm, and apply standard scalar concentration tools. We will see an example of this in Section 2.8.

Remark 2.6.7. *We have made no attempts to optimize the factors in front of the expectation in Theorem 2.6.6, which we suspect can be improved.*

We prove the main theorem by repeatedly applying the following technical lemma, the proof of which we defer to the next section.

Lemma 2.6.8. *For all integers $t \geq 1$, integers $k \geq 1, a, b \geq 0$ such that $a + b < k$,*

$$\mathbb{E} \left\| \overline{\mathbf{F}}_{k,a,b} \right\|_{2t}^{2t} \leq (Ct^2 d d_p^2)^t (\mathbb{E} \left\| \overline{\mathbf{F}}_{k,a,b+1} \right\|_{2t}^{2t} + \mathbb{E} \left\| \overline{\mathbf{F}}_{k,a+1,b} \right\|_{2t}^{2t})$$

for an absolute constant $C > 0$.

Using this lemma, we can complete the proof of the main theorem.

Proof of Theorem 2.6.6. Using Fact 2.2.3, we have $\mathbb{E} \left\| \mathbf{X} \right\|_{2t}^{2t} \leq d_p^{2t} \sum_{k=1}^{d_p} \mathbb{E} \left\| \mathbf{X}_k \right\|_{2t}^{2t}$. Note that for any $k \geq 1$, the matrix $\mathbf{X}_k$ is a principal submatrix of $\mathbf{F}_{k,0,0}$ with all other entries being 0, so $\mathbb{E} \left\| \mathbf{X}_k \right\|_{2t}^{2t} = \mathbb{E} \left\| \mathbf{F}_{k,0,0} \right\|_{2t}^{2t} = \frac{1}{2} \mathbb{E} \left\| \overline{\mathbf{F}}_{k,0,0} \right\|_{2t}^{2t}$. Therefore,

$$\mathbb{E} \left\| \mathbf{X} \right\|_{2t}^{2t} \leq \frac{1}{2} d_p^{2t} \sum_{k=1}^{d_p} \mathbb{E} \left\| \overline{\mathbf{F}}_{k,0,0} \right\|_{2t}^{2t}$$

We now apply Lemma 2.6.8 repeatedly to all our terms until $k = a + b$, ultimately giving

$$\mathbb{E} \left\| \mathbf{X} \right\|_{2t}^{2t} \leq \frac{1}{2} d_p^{2t} (Ct^2 d d_p^2)^{(a+b)t} \sum_{a,b \geq 0, a+b \geq 1} \mathbb{E} \left\| \overline{\mathbf{F}}_{a+b,a,b} \right\|_{2t}^{2t}$$

Observing that $\mathbb{E}\left\|\overline{\mathbf{F}}_{a+b,a,b}\right\|_{2t}^{2t} = 2\,\mathbb{E}\left\|\mathbf{F}_{a+b,a,b}\right\|_{2t}^{2t}$ completes the proof. ∎

2.7 A generalization of [147] and proof of Lemma 2.6.8

In this section, we will prove Lemma 2.6.8 using the high level strategy described in Section 2.1. This requires generalizing the results in [147], and the proof techniques may be of independent interest.

2.7.1 Generalizing [147] via explicit inner kernels

In our setting, observe that (Z, Z') has the same distribution as (Z', Z). This is what is known as an *exchangeable pair* of variables, that will be extremely useful for our analysis. In particular, Z, Z' have the same distribution and $\mathbb{E}\, f(Z, Z') = \mathbb{E}\, f(Z', Z)$ for every integrable function f.

Definition 2.7.1 (Laplacian operator $\mathcal{L}$)**.** *Define the operator $\mathcal{L}$ on the space $\mathcal{S}$ as*

$$\mathcal{L}(f)(Z) = \mathbb{E}[f(Z) - f(Z')|Z]$$

for all polynomials $f \in \mathcal{S}$.

Note that this operator is well-defined since for any $f \in \mathcal{S}$, $\mathbb{E}[L(f)] = \mathbb{E}[\mathbb{E}[f(Z) - f(Z')|Z]] = \mathbb{E}[f(Z) - f(Z')] = 0$ and hence, $L(f) \in \mathcal{S}$.

Lemma 2.7.2. *For all $\alpha \in \mathbb{N}^n$, χ_α is an eigenvector of $\mathcal{L}$ with eigenvalue $\frac{|\alpha|_0}{n}$.*

Proof. Recall that Z' is obtained by choosing $i \in [n]$ uniformly at random and then setting $Z' = Z^{(i)}$. Therefore,

$$\begin{aligned}\mathcal{L}(\chi_\alpha)(Z) &= \mathbb{E}[\chi_\alpha(Z) - \chi_\alpha(Z')|Z] \\ &= \frac{1}{n}\sum_{i \le n} \mathbb{E}[\chi_\alpha(Z) - \chi_\alpha(Z^{(i)})|Z]\end{aligned}$$

When $\alpha_i = 0$, $\chi_\alpha(Z) - \chi_\alpha(Z^{(i)}) = 0$. Otherwise, $\mathbb{E}[\chi_\alpha(Z) - \chi_\alpha(Z^{(i)})|Z] = \chi_\alpha(Z)$. Therefore, the above expression simplifies to $\frac{|\alpha|_0}{n}\chi_\alpha(Z)$. ∎

Theorem 2.7.3 (Explicit Kernel)**.** *For any mean-centered polynomial $f \in \mathcal{S}$, there exists a polynomial K_f on $2n$ variables $z_1, \ldots, z_n, z'_1, \ldots, z'_n$, denoted collectively as (z, z'), with the following properties*

1. $K_f(z', z) = -K_f(z, z')$
2. $\mathbb{E}[K_f(Z, Z')|Z] = f(Z)$ *where (Z, Z') is the exchangeable pair we consider above.*

Proof. Using Proposition 2.6.2 and Lemma 2.7.2, under the basis of polynomials χ_α, the operator $\mathcal{L}$ is a diagonal matrix with nonzero diagonal entries and therefore, $\mathcal{L}^{-1}$ exists and is explicitly given by

$$\mathcal{L}^{-1}(f)(Z) = \sum_\alpha \frac{n}{|\alpha|_0}\widehat{f}(\alpha)\chi_\alpha(Z)$$

We then take $K_f(z, z') = \mathcal{L}^{-1}(f)(z) - \mathcal{L}^{-1}(f)(z')$. The first condition is obvious and for the second condition, we have

$$\mathbb{E}[K_f(Z, Z')|Z] = \mathbb{E}[\mathcal{L}^{-1}(f)(Z) - \mathcal{L}^{-1}(f)(Z')|Z] = \mathcal{L}(\mathcal{L}^{-1}(f)) = f$$

∎

As seen in the proof of Theorem 2.7.3, $\mathcal{L}$ has a well-defined inverse $\mathcal{L}^{-1}$. We now define the matrix $\mathbf{K}_{k,a,b}$ that we call the *inner kernel*.

Definition 2.7.4 (The inner kernel matrix $\mathbf{K}_{k,a,b}$)**.** *For integers $k \geq 1, a, b \geq 0$ such that $a + b < k$, define the matrix $\mathbf{K}_{k,a,b} \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}}$ taking $2n$ variables $(z, z') = (z_1, \ldots, z_n, z'_1, \ldots, z'_n)$ as input as follows*

$$\mathbf{K}_{k,a,b}(z, z') = \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(z) - \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(z')$$

In the rest of this section except where explicitly stated, fix integers $k \geq 1, a, b \geq 0$ such that $a + b < k$. Then, the inner kernel $\mathbf{K}_{k,a,b}$ is well-defined.

Lemma 2.7.5. $\mathbf{K}_{k,a,b}(Z, Z') = \frac{n}{k-a-b}(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))$

Proof.

$$\begin{aligned}
\mathbf{K}_{k,a,b}(Z, Z') &= \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(Z) - \mathcal{L}^{-1}(\mathbf{G}_{k,a,b})(Z') \\
&= \sum_{|\alpha|_0 = k-a-b} \widehat{\mathbf{G}_{k,a,b}}(\alpha)(\mathcal{L}^{-1}(\chi_\alpha)(Z) - \mathcal{L}^{-1}(\chi_\alpha)(Z')) \\
&= \frac{n}{k-a-b} \sum_{|\alpha|_0 = k-a-b} \widehat{\mathbf{G}_{k,a,b}}(\alpha)(\chi_\alpha(Z) - \chi_\alpha(Z')) \\
&= \frac{n}{k-a-b}(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))
\end{aligned}$$

■

The following lemma postulates important properties of the the inner kernel, including how it interacts with $\mathbf{D}_1$ and $\mathbf{D}_2$.

Lemma 2.7.6. $\mathbf{K}_{k,a,b}$ *satisfies the following properties*

1. $\mathbf{K}_{k,a,b}(z', z) = -\mathbf{K}_{k,a,b}(z, z')$
2. $\mathbb{E}[\mathbf{K}_{k,a,b}(Z, Z')|Z] = \mathbf{G}_{k,a,b}(Z)$
3. $(\mathbf{D}_1(Z) - \mathbf{D}_1(Z'))\mathbf{K}_{k,a,b}(Z, Z') = \mathbf{K}_{k,a,b}(Z, Z')(\mathbf{D}_2(Z) - \mathbf{D}_2(Z')) = 0.$

Proof. The first equality is obvious from the definition. For the second equality, note that $\mathbb{E}[\mathbf{G}_{k,a,b}] = 0$ and $\mathbf{K}_{k,a,b}$ is defined by replacing each entry f of $\mathbf{G}_{k,a,b}$ by the kernel polynomial K_f as exhibited in Theorem 2.7.3. Now, we prove the third equality.

Consider the matrix $(\mathbf{D}_1(Z) - \mathbf{D}_1(Z'))\mathbf{K}_{k,a,b}(Z, Z')$ whose $[(I, \alpha_1, \gamma_1), (J, \alpha_2, \gamma_2)]$ entry is given by

$$\frac{n}{k-a-b}\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)}]}(Z^{\alpha_1\cdot\gamma_1} - (Z')^{\alpha_1\cdot\gamma_1})(\nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z) - \nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z'))$$

where we have used Lemma 2.7.5. We will argue that this term is identically 0. We must have $Z' = Z^{(i)}$ for some $i \leq n$. If $(\alpha_1 \cdot \gamma_1)_i = 0$, then $Z^{\alpha_1\cdot\gamma_1} = (Z')^{\alpha_1\cdot\gamma_1}$ and the above term is 0. Otherwise, $(\alpha_1 + \alpha_2)_i \neq 0$ and so $\nabla_{\alpha_1+\alpha_2}$ on any polynomial f will only contain the terms independent of Z_i, in which case $\nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z) = \nabla_{\alpha_1+\alpha_2}\mathbf{X}_k[I,J](Z')$. In this case was well, the above term is 0. The proof of the other equality is analogous. ∎

The reason we call $\mathbf{K}_{k,a,b}$ the inner kernel is because, as seen above, it serves as a kernel for the inner matrix $\mathbf{G}$ in the decomposition $\mathbf{F} = \mathbf{DGD}$.

Since we will need to work with Hermitian dilations, we define

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}$$

We will use the following basic fact extensively in our manipulations.

Fact 2.7.7. *For any matrix* $\mathbf{A} \in \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{J}\times\mathcal{K}}$, $\mathbf{D}\overline{\mathbf{A}}\mathbf{D} = \overline{\mathbf{D}_1\mathbf{A}\mathbf{D}_2}$.

Proof. We have

$$\begin{aligned}\mathbf{D}\overline{\mathbf{A}}\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix}\begin{bmatrix} 0 & \mathbf{A} \\ \mathbf{A}^\intercal & 0 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} &= \begin{bmatrix} 0 & \mathbf{D}_1\mathbf{A} \\ \mathbf{D}_2\mathbf{A}^\intercal & 0 \end{bmatrix}\begin{bmatrix} \mathbf{D}_1 & 0 \\ 0 & \mathbf{D}_2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & \mathbf{D}_1\mathbf{A}\mathbf{D}_2 \\ \mathbf{D}_2\mathbf{A}^\intercal\mathbf{D}_1 & 0 \end{bmatrix} \\ &= \overline{\mathbf{D}_1\mathbf{A}\mathbf{D}_2}\end{aligned}$$

■

We start with a generalized version of a result from [147].

Lemma 2.7.8. *Let* $\mathbf{K} = \overline{\mathbf{K}}_{k,a,b}$. *For any symmetric matrix valued function* $\mathbf{R}$ *on the variables* Z *of the same dimensions as* $\mathbf{K}$, *such that* $\mathbb{E}\left\|\mathbf{K}(Z,Z')\mathbf{R}(Z)\right\| < \infty$, *we have*

$$\mathbb{E}[\overline{\mathbf{F}}_{k,a,b}(Z)\mathbf{R}(Z)] = \frac{1}{2}\,\mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z,Z')\mathbf{D}(Z)(\mathbf{R}(Z) - \mathbf{R}(Z'))]$$

Proof. By Lemma 2.7.6, we have

$$\begin{aligned}
\mathbb{E}[\overline{\mathbf{F}}_{k,a,b}(Z)\mathbf{R}(Z)] &= \mathbb{E}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\mathbf{R}(Z)] \\
&= \mathbb{E}[\mathbf{D}(Z)\,\mathbb{E}[\mathbf{K}(Z,Z')|Z]\mathbf{D}(Z)\mathbf{R}(Z)] \\
&= \mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z,Z')\mathbf{D}(Z)\mathbf{R}(Z)]
\end{aligned}$$

where the first equality follow from condition 2 of Lemma 2.7.6 and the second follows from the pull-through property of expectations. Continuing,

$$\begin{aligned}
\mathbb{E}[\overline{\mathbf{F}}_{k,a,b}(Z)\mathbf{R}(Z)] &= \mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z,Z')\mathbf{D}(Z)\mathbf{R}(Z)] \\
&= \mathbb{E}[\mathbf{D}(Z')\mathbf{K}(Z',Z)\mathbf{D}(Z')\mathbf{R}(Z')] \\
&= -\,\mathbb{E}[\mathbf{D}(Z')\mathbf{K}(Z,Z')\mathbf{D}(Z')\mathbf{R}(Z')] \\
&= -\,\mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z,Z')\mathbf{D}(Z')\mathbf{R}(Z')] \\
&= -\,\mathbb{E}[\mathbf{D}(Z)\mathbf{K}(Z,Z')\mathbf{D}(Z)\mathbf{R}(Z')]
\end{aligned}$$

Here, the second equality follows from the fact that (Z,Z') has the same distribution as (Z',Z), so we can exchange them. The third, fourth and fifth equalities follow from conditions $1,3,3$ of Lemma 2.7.6 respectively. Adding the two displays, we get the result. ■

Definition 2.7.9 (Matrices $\mathbf{U}_{k,a,b}, \mathbf{V}_{k,a,b}$)**.** *We define the following matrices*

$$\mathbf{U}_{k,a,b} = \mathbb{E}[(\overline{\mathbf{F}}_{k,a,b}(Z) - \overline{\mathbf{F}}_{k,a,b}(Z'))^2 | Z]$$

$$\mathbf{V}_{k,a,b} = \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z, Z')\mathbf{D}(Z))^2 | Z]$$

The definition of $\mathbf{U}_{k,a,b}$ is essentially unchanged from [147], where it is called the *conditional variance*. The definition of $\mathbf{V}_{k,a,b}$ is slightly different in our setting. This lets us exploit the specific product structure exhibited by $\overline{\mathbf{F}}_{k,a,b}$ and the special properties of the inner kernel from Lemma 2.7.6.

We will now prove a lemma which is similar to a lemma shown in [147].

Lemma 2.7.10. *For any $s > 0$ and for any integer $t \geq 1$,*

$$\mathbb{E} \left\| \overline{\mathbf{F}}_{k,a,b} \right\|_{2t}^{2t} \leq \left(\frac{2t-1}{4} \right)^t \mathbb{E} \left\| s\mathbf{U}_{k,a,b} + s^{-1}\mathbf{V}_{k,a,b} \right\|_t^t$$

To prove this, we will need the following inequality.

Lemma 2.7.11 (Polynomial mean value trace inequality, [147])**.** *For all matrices $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{H}^d$, all integers $q \geq 1$ and all $s > 0$,*

$$\operatorname{tr}[\mathbf{C}(\mathbf{A}^q - \mathbf{B}^q)]| \leq \frac{q}{4} \operatorname{tr}[(s(\mathbf{A} - \mathbf{B})^2 + s^{-1}\mathbf{C}^2)(\mathbf{A}^{q-1} + \mathbf{B}^{q-1})]$$

Proof of Lemma 2.7.10. We start by invoking Lemma 2.7.8 by setting $\mathbf{R}(Z) = \overline{\mathbf{F}}_{k,a,b}^{2t-1}(Z)$.

$$\begin{aligned}
\mathbb{E} \left\| \overline{\mathbf{F}}_{k,a,b} \right\|_{2t}^{2t} &= \mathbb{E} \operatorname{tr}[\overline{\mathbf{F}}_{k,a,b} \cdot \overline{\mathbf{F}}_{k,a,b}^{2t-1}] \\
&= \frac{1}{2} \mathbb{E}[\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z, Z')\mathbf{D}(Z)(\overline{\mathbf{F}}_{k,a,b}^{2t-1}(Z) - \overline{\mathbf{F}}_{k,a,b}^{2t-1}(Z'))]
\end{aligned}$$

Applying Lemma 2.7.11,

$$\begin{aligned}
&\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \\
&\leq (\frac{2t-1}{8})\,\mathbb{E}\,\mathrm{tr}[(s(\overline{\mathbf{F}}_{k,a,b}(Z)-\overline{\mathbf{F}}_{k,a,b}(Z'))^2+s^{-1}(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2)(\overline{\mathbf{F}}_{k,a,b}^{2t-2}(Z)+\overline{\mathbf{F}}_{k,a,b}^{2t-2}(Z'))] \\
&= (\frac{2t-1}{4})\,\mathbb{E}\,\mathrm{tr}[(s(\overline{\mathbf{F}}_{k,a,b}(Z)-\overline{\mathbf{F}}_{k,a,b}(Z'))^2+s^{-1}(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2)\overline{\mathbf{F}}_{k,a,b}^{2t-2}(Z)]
\end{aligned}$$

where the last line used the fact that (Z, Z') has the same distribution as (Z', Z) and applied condition 3 of Lemma 2.7.6. Using the definitions of $\mathbf{U}_{k,a,b}$ and $\mathbf{V}_{k,a,b}$, we get

$$\begin{aligned}
\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} &\leq \frac{2t-1}{4}\,\mathbb{E}\,\mathrm{tr}[(s\mathbf{U}_{k,a,b}+s^{-1}\mathbf{V}_{k,a,b})\overline{\mathbf{F}}_{k,a,b}^{2t-2}] \\
&\leq \frac{2t-1}{4}\left(\mathbb{E}\left\|s\mathbf{U}_{k,a,b}+s^{-1}\mathbf{V}_{k,a,b}\right\|_t^t\right)^{1/t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t})^{(t-1)/t}
\end{aligned}$$

where we used Hölder's inequality for the trace and Hölder's inequality for the expectation. Rearranging gives the result. ∎

2.7.2 Proof of Lemma 2.6.8

Lemma 2.7.10 suggests that in order to bound $\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$, it suffices to bound $\mathbb{E}\left\|\mathbf{U}_{k,a,b}\right\|_t^t$ and $\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t$. Indeed, this will be our strategy. To bound $\mathbb{E}\left\|\mathbf{U}_{k,a,b}\right\|_t^t$, we will bound it via the matrices that we define below.

Definition 2.7.12 (Matrices $\mathbf{\Delta}_1^{k,a,b}, \mathbf{\Delta}_2^{k,a,b}, \mathbf{\Delta}_3^{k,a,b}$)**.** *Define the matrices*

$$\mathbf{\Delta}_1^{k,a,b} = \mathbb{E}[((\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z))^2|Z]$$

$$\mathbf{\Delta}_2^{k,a,b} = \mathbb{E}[(\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z)-\overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z))^2|Z]$$

$$\mathbf{\Delta}_3^{k,a,b} = \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z')))^2|Z]$$

Lemma 2.7.13. $\mathbf{U}_{k,a,b} \preceq 3(\mathbf{\Delta}_1^{k,a,b}+\mathbf{\Delta}_2^{k,a,b}+\mathbf{\Delta}_3^{k,a,b})$.

To prove this lemma, we will use the following lemma.

Lemma 2.7.14. *We have the relations*

$$(\mathbf{D}(Z) - \mathbf{D}(Z'))(\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z')) = 0$$

$$(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))(\mathbf{D}(Z) - \mathbf{D}(Z')) = 0$$

Proof sketch. The proof is similar to the proof of third equality in Lemma 2.7.6. When Z' is set to $Z^{(i)}$ for some $i \leq n$, when a diagonal entry of $\mathbf{D}(Z) - \mathbf{D}(Z')$ is nonzero, then the corresponding row of $\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z')$ will be 0. The second equality is analogous. ∎

Proof of Lemma 2.7.13. We have

$$\begin{aligned}
&(\overline{\mathbf{F}}_{k,a,b}(Z) - \overline{\mathbf{F}}_{k,a,b}(Z'))^2 \\
&= (\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z) - \mathbf{D}(Z')\overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z'))^2 \\
&= \left(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z) - \mathbf{D}(Z')) + \mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z') + (\mathbf{D}(Z) - \mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z')\mathbf{D}(Z')\right)^2 \\
&= \left(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z) - \mathbf{D}(Z')) + \mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z) + (\mathbf{D}(Z) - \mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\right)^2
\end{aligned}$$

where the last equality follows from Lemma 2.7.14. Taking expectations conditioned on Z and applying Fact 2.2.2, we immediately get $\mathbf{U}_{k,a,b} \preceq 3(\mathbf{\Delta}_1^{k,a,b} + \mathbf{\Delta}_2^{k,a,b} + \mathbf{\Delta}_3^{k,a,b})$. ∎

In subsequent sections, we will prove the following technical bounds on the matrices we have considered so far.

Lemma 2.7.15. *For all integers $t \geq 1$, $\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t \leq \frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$.*

Lemma 2.7.16. *For all integers $t \geq 1$, $\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t \leq (2d_p)^t n^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$.*

Lemma 2.7.17. *For all integers $t \geq 1$, $\mathbb{E}\left\|\mathbf{\Delta}_1^{k,a,b}\right\|_t^t \leq \frac{(8dd_p)^t}{n^t}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$.*

Lemma 2.7.18. *For all integers $t \geq 1$, $\mathbb{E}\left\|\boldsymbol{\Delta}_3^{k,a,b}\right\|_t^t \leq \frac{(4d_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$.*

Assuming the above lemmas, we can complete the proof of Lemma 2.6.8, which we restate for convenience.

Lemma 2.6.8. *For all integers $t \geq 1$, integers $k \geq 1, a, b \geq 0$ such that $a + b < k$,*

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \leq (Ct^2dd_p^2)^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

for an absolute constant $C > 0$.

Proof of Lemma 2.6.8. Using Lemma 2.7.10, Lemma 2.7.13, we get that for any $s > 0$,

$$\begin{aligned}
\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} &\leq (\frac{2t-1}{4})^t\,\mathbb{E}\left\|s\mathbf{U}_{k,a,b} + s^{-1}\mathbf{V}_{k,a,b}\right\|_t^t \\
&\leq t^t(s^t\,\mathbb{E}\left\|\mathbf{U}_{k,a,b}\right\|_t^t + s^{-t}\,\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t) \\
&\leq (9st)^t(\mathbb{E}\left\|\boldsymbol{\Delta}_1^{k,a,b}\right\|_t^t + \mathbb{E}\left\|\boldsymbol{\Delta}_2^{k,a,b}\right\|_t^t + \mathbb{E}\left\|\boldsymbol{\Delta}_3^{k,a,b}\right\|_t^t) + t^ts^{-t}\,\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t
\end{aligned}$$

Let $\rho = s/n$. Since the inequality is true for any choice of $s > 0$, it is true for any choice of $\rho > 0$. Now, using Lemma 2.7.17, Lemma 2.7.18,

$$\begin{aligned}
(9st)^t(\mathbb{E}\left\|\boldsymbol{\Delta}_1^{k,a,b}\right\|_t^t + \mathbb{E}\left\|\boldsymbol{\Delta}_3^{k,a,b}\right\|_t^t) &\leq (9st)^t\left(\frac{(8dd_p)^t}{n^t} + \frac{(4d_p)^t}{n^t}\right)\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \\
&= \rho^t(C_1tdd_p)^t\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}
\end{aligned}$$

for an absolute constant $C_1 > 0$. Using Lemma 2.7.15, Lemma 2.7.16,

$$\begin{aligned}
(9st)^t\,\mathbb{E}\left\|\boldsymbol{\Delta}_2^{k,a,b}\right\|_t^t + t^ts^{-t}\,\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t &\leq \left((9st)^t\frac{(2d_p)^t}{n^t} + t^ts^{-t}(2d_p)^tn^t\right)(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t}) \\
&\leq (\rho^tC_2^t + \rho^{-t}C_3^t)(td_p)^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})
\end{aligned}$$

for absolute constants $C_2, C_3 > 0$. Therefore,

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \leq \rho^t (C_1 t d d_p)^t \,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} + (\rho^t C_2^t + \rho^{-t} C_3^t)(t d_p)^t (\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

We choose $\rho > 0$ so that $\rho^t (C_1 t d d_p)^t = \frac{1}{2}$ to get

$$\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \leq \frac{1}{2}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} + \frac{1}{2}(C t^2 d d_p^2)^t (\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$$

for an absolute constant $C > 0$. Rearranging yields the result. ∎

2.7.3 Bounding $\boldsymbol{\Delta}_2^{k,a,b}$ and $\mathbf{V}_{k,a,b}$

The next lemma relates $\mathbf{V}_{k,a,b}$ to $\boldsymbol{\Delta}_2^{k,a,b}$ upto a factor of n^2 which will be enough for us. We can then focus on bounding $\boldsymbol{\Delta}_2^{k,a,b}$.

Lemma 2.7.19. $\mathbf{V}_{k,a,b} \preceq n^2 \boldsymbol{\Delta}_2^{k,a,b}$

Proof. Using Lemma 2.7.5,

$$\begin{aligned}
\mathbf{V}_{k,a,b} &= \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{K}}_{k,a,b}(Z,Z')\mathbf{D}(Z))^2|Z] \\
&= \mathbb{E}[(\mathbf{D}(Z)\left(\frac{n}{k-a-b}(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))\right)\mathbf{D}(Z))^2|Z] \\
&\preceq n^2\,\mathbb{E}[(\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z))^2|Z] \\
&= n^2 \boldsymbol{\Delta}_2^{k,a,b}
\end{aligned}$$

∎

For $1 \leq i \leq n$ and $1 \leq l \leq d$, let $\mathbf{e}_{i,l} \in \mathbb{N}^n$ denote the vector α with $\alpha_i = l$ and $\alpha_j = 0$ for $j \neq i$. We note the following simple proposition.

Proposition 2.7.20. *For any polynomial f such that the degree of Z_i is at most d,*

$$f(Z) - f(Z^{(i)}) = \sum_{1 \le l \le d} (Z_i^l - \widetilde{Z}_i^{\,l}) \nabla_{\mathbf{e}_{i,l}}(f)$$

We now restate and prove Lemma 2.7.15.

Lemma 2.7.15. *For all integers $t \ge 1$, $\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t \le \frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$.*

Proof. Consider

$$\begin{aligned}
\mathbf{\Delta}_2^{k,a,b} &= \mathbb{E}[(\mathbf{D}(Z)(\overline{\mathbf{G}}_{k,a,b}(Z) - \overline{\mathbf{G}}_{k,a,b}(Z'))\mathbf{D}(Z))^2 | Z] \\
&= \mathbb{E}\left[\begin{bmatrix} \mathbf{M}\mathbf{M}^\intercal & 0 \\ 0 & \mathbf{M}^\intercal\mathbf{M} \end{bmatrix} | Z \right] \\
&= \begin{bmatrix} \mathbb{E}[\mathbf{M}\mathbf{M}^\intercal | Z] & 0 \\ 0 & \mathbb{E}[\mathbf{M}^\intercal\mathbf{M} | Z] \end{bmatrix}
\end{aligned}$$

where $\mathbf{M} = \mathbf{D}_1(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))\mathbf{D}_2(Z)$. Using Proposition 2.7.20,

$$\begin{aligned}
\mathbb{E}[\mathbf{M}\mathbf{M}^T | Z] &= \mathbb{E}[\mathbf{D}_1(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))\mathbf{D}_2(Z) \cdot \mathbf{D}_2(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z'))^\intercal \mathbf{D}_1(Z) | Z] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathbf{D}_1(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z^{(i)}))\mathbf{D}_2(Z) \cdot \mathbf{D}_2(Z)(\mathbf{G}_{k,a,b}(Z) - \mathbf{G}_{k,a,b}(Z^{(i)}))^\intercal \mathbf{D}_1(Z) | Z] \\
&= \frac{1}{n}\sum_{i=1}^n \sum_{l=1}^d \mathbb{E}[(Z_i^l - \widetilde{Z}_i^{\,l})^2 | Z] \cdot \mathbf{D}_1(Z)(\nabla_{\mathbf{e}_{i,l}}\mathbf{G}_{k,a,b})(Z)\mathbf{D}_2(Z) \cdot \mathbf{D}_2(Z)(\nabla_{\mathbf{e}_{i,l}}\mathbf{G}_{k,a,b})(Z)^\intercal \mathbf{D}_1(Z)
\end{aligned}$$

Define $\mathbf{N}_{i,l}(Z) := \mathbf{D}_1(Z)(\nabla_{\mathbf{e}_{i,l}}\mathbf{G}_{k,a,b})(Z)\mathbf{D}_2(Z)$. Then,

$$\begin{aligned}
\mathbb{E}[\mathbf{M}\mathbf{M}^T | Z] &= \frac{1}{n}\sum_{i=1}^n \sum_{l=1}^d \mathbb{E}[(Z_i^l - \widetilde{Z}_i^{\,l})^2 | Z] \cdot \mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal \\
&\preceq \frac{2}{n}\sum_{i=1}^n \sum_{l=1}^d (Z_i^{2l} + \mathbb{E}[Z_i^{2l}]) \cdot \mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal
\end{aligned}$$

Similarly,

$$\mathbb{E}[\mathbf{M}^\intercal\mathbf{M}|Z] \preceq \frac{2}{n}\sum_{i=1}^{n}\sum_{l=1}^{d}(Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)^\intercal\mathbf{N}_{i,l}(Z)$$

Claim 2.7.21. *We have the relations*

$$\sum_{i=1}^{n}\sum_{l=1}^{d}(Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal = (b+1)\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^\intercal$$

$$\sum_{i=1}^{n}\sum_{l=1}^{d}(Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\cdot\mathbf{N}_{i,l}(Z)^\intercal\mathbf{N}_{i,l}(Z) = (a+1)\mathbf{F}_{k,a+1,b}^\intercal\mathbf{F}_{k,a+1,b}$$

Using this claim, we have

$$\mathbb{E}[\mathbf{M}\mathbf{M}^T|Z] \preceq \frac{2(b+1)}{n}\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^\intercal \preceq \frac{2d_p}{n}\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^\intercal$$

$$\mathbb{E}[\mathbf{M}^\intercal\mathbf{M}|Z] \preceq \frac{2(a+1)}{n}\mathbf{F}_{k,a+1,b}^\intercal\mathbf{F}_{k,a+1,b} \preceq \frac{2d_p}{n}\mathbf{F}_{k,a+1,b}^\intercal\mathbf{F}_{k,a+1,b}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left\|\mathbf{\Delta}_2^{k,a,b}\right\|_t^t &= \mathbb{E}\left\|\mathbb{E}[\mathbf{M}\mathbf{M}^\intercal|Z]\right\|_t^t + \mathbb{E}\left\|\mathbb{E}[\mathbf{M}^\intercal\mathbf{M}|Z]\right\|_t^t \\ &\le \frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\mathbf{F}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\mathbf{F}_{k,a+1,b}\right\|_{2t}^{2t}) \\ &\le \frac{(2d_p)^t}{n^t}(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})\end{aligned}$$

■

It remains to prove the claim.

Proof of Claim 2.7.21. We will prove the first relation, the second is analogous. For a fixed $i \le n, l \le d$, consider any nonzero entry $[(I_1,\alpha_1,\gamma_1),(I_2,\alpha_2,\gamma_2)]$ of $\sum_{i=1}^n\sum_{l=1}^d(Z_i^{2l} + \mathbb{E}[Z_i^{2l}])\mathbf{N}_{i,l}(Z)\mathbf{N}_{i,l}(Z)^\intercal$, where $I_1, I_2 \in \mathcal{I}, (\alpha_1,\gamma_1),(\alpha_2,\gamma_2) \in \mathcal{K}$. We must have $|\alpha_1|_0 =$

$|\alpha_2|_0 = a$, in which case the entry is equal to

$$\sum_{\substack{(J,\alpha_3,\gamma_3)\in\mathcal{J}\times\mathcal{K}\\ |\alpha_3|=b\\ \alpha_1\alpha_3=\alpha_2\alpha_3=0}} (Z_i^{2l}+\mathbb{E}[Z_i^{2l}])\cdot(\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_1\cdot\gamma_1+\alpha_3\cdot\gamma_3}\nabla_{\mathbf{e}_{i,l}}\nabla_{\alpha_1+\alpha_3}\mathbf{X}_k[I_1,J])$$
$$\cdot(\sqrt{\mathbb{E}[Z^{2\alpha_2\cdot(1-\gamma_2)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_2\cdot\gamma_2+\alpha_3\cdot\gamma_3}\nabla_{\mathbf{e}_{i,l}}\nabla_{\alpha_2+\alpha_3}\mathbf{X}_k[I_2,J])$$

Note that the term inside the summation is nonzero only when $\mathbf{e}_{i,l}\cdot(\alpha_1+\alpha_3) = \mathbf{e}_{i,l}\cdot(\alpha_2+\alpha_3) = 0$. Hence, this sum can be written as

$$\sum_{\substack{(J,\alpha_3,\gamma_3)\in\mathcal{J}\times\mathcal{K}\\ |\alpha_3|=b+1\\ \mathbf{e}_{i,l}\trianglelefteq\alpha_3,\alpha_1\alpha_3=\alpha_2\alpha_3=0}} (\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_1\cdot\gamma_1+\alpha_3\cdot\gamma_3}\nabla_{\alpha_1+\alpha_3}\mathbf{X}_k[I_1,J])$$
$$\cdot(\sqrt{\mathbb{E}[Z^{2\alpha_2\cdot(1-\gamma_2)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_2\cdot\gamma_2+\alpha_3\cdot\gamma_3}\nabla_{\alpha_2+\alpha_3}\mathbf{X}_k[I_2,J])$$

When we add this entry over all $i \le n, l \le d$, this simplifies to

$$(b+1)\cdot\sum_{\substack{(J,\alpha_3,\gamma_3)\in\mathcal{J}\times\mathcal{K}\\ |\alpha_3|=b+1\\ \alpha_1\alpha_3=\alpha_2\alpha_3=0}} (\sqrt{\mathbb{E}[Z^{2\alpha_1\cdot(1-\gamma_1)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_1\cdot\gamma_1+\alpha_3\cdot\gamma_3}\nabla_{\alpha_1+\alpha_3}\mathbf{X}_k[I_1,J])$$
$$\cdot(\sqrt{\mathbb{E}[Z^{2\alpha_2\cdot(1-\gamma_2)+2\alpha_3\cdot(1-\gamma_3)}]}Z^{\alpha_2\cdot\gamma_2+\alpha_3\cdot\gamma_3}\nabla_{\alpha_2+\alpha_3}\mathbf{X}_k[I_2,J])$$

The factor of $(b+1)$ came because the index i could have been chosen from among all the active indices in α_3. But this is precisely the $[(I_1,\alpha_1,\gamma_1),(I_2,\alpha_2,\gamma_2)]$ entry of $(b+1)\mathbf{F}_{k,a,b+1}\mathbf{F}_{k,a,b+1}^{\intercal}$, proving the claim. ∎

We restate and prove Lemma 2.7.16.

Lemma 2.7.16. *For all integers $t \ge 1$, $\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t \le (2d_p)^t n^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t}+\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})$.*

Proof. Using Lemma 2.7.19 and Lemma 2.7.15, we get

$$\begin{aligned}\mathbb{E}\left\|\mathbf{V}_{k,a,b}\right\|_t^t &\leq n^{2t}\,\mathbb{E}\left\|\boldsymbol{\Delta}_2^{k,a,b}\right\|_t^t \\ &\leq (2d_p)^t n^t(\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b+1}\right\|_{2t}^{2t} + \mathbb{E}\left\|\overline{\mathbf{F}}_{k,a+1,b}\right\|_{2t}^{2t})\end{aligned}$$

■

2.7.4 Bounding $\boldsymbol{\Delta}_1^{k,a,b}$ and $\boldsymbol{\Delta}_3^{k,a,b}$

Define $\sqcup$ to be the disjoint union of sets. For $1 \leq i \leq n$ and $1 \leq l \leq d$, define the diagonal matrices $\boldsymbol{\Pi}_{i,l}, \boldsymbol{\Pi}'_{i,l}, \boldsymbol{\Pi}_i, \boldsymbol{\Pi}'_i \in \mathbb{R}^{(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})} \times \mathbb{R}^{(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})}$ (the same dimensions as $\mathbf{D}$) as

$$\boldsymbol{\Pi}_{i,l}[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases}1 & \text{if } (\alpha\cdot\gamma)_i \neq 0 \text{ and } \alpha_i = l \\ 0 & \text{o.w.}\end{cases} \qquad \boldsymbol{\Pi}_i[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases}1 & \text{if } (\alpha\cdot\gamma)_i \neq 0 \\ 0 & \text{o.w.}\end{cases}$$

$$\boldsymbol{\Pi}'_{i,l}[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases}1 & \text{if } \alpha_i \neq 0 \text{ and } \alpha_i = l \\ 0 & \text{o.w.}\end{cases} \qquad \boldsymbol{\Pi}'_i[(I,\alpha,\beta),(I,\alpha,\beta)] = \begin{cases}1 & \text{if } \alpha_i \neq 0 \\ 0 & \text{o.w.}\end{cases}$$

for all $I \in \mathcal{I} \sqcup \mathcal{J}$. Note that for all $i \leq n$, $\boldsymbol{\Pi}_i = \sum_{l=1}^d \boldsymbol{\Pi}_{i,l}$.

Also, for all $1 \leq i \leq n$, we define the permutation matrices $\boldsymbol{\Sigma}_i \in \mathbb{R}^{(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})} \times \mathbb{R}^{(\mathcal{I}\times\mathcal{K})\sqcup(\mathcal{J}\times\mathcal{K})}$ as follows. Consider the permutation σ_1 on $\mathcal{I}\times\mathcal{K}$ that transposes (I,α,γ) and $(I,\alpha,\gamma+\mathbf{e}_i)$ for all $(I,\alpha,\gamma) \in \mathcal{I}\times\mathcal{K}$ such that $\alpha_i \neq 0$. Here, $\mathbf{e}_i \in \{0,1\}^n$ has exactly one nonzero entry, which is in the ith position, and $\gamma+\mathbf{e}_i$ is the usual addition over $\mathbb{F}_2$. σ_1 leaves other positions fixed. Let $\boldsymbol{\Sigma}_i^{(1)}$ be the permutation matrix for σ. Similarly, let $\boldsymbol{\Sigma}_i^{(2)}$ be the permutation matrix of the permutation σ_2 on $\mathcal{J}\times\mathcal{K}$ that transposes (J,α,γ) and $(J,\alpha,\gamma+\mathbf{e}_i)$ for all $(J,\alpha,\gamma) \in \mathcal{J}\times\mathcal{K}$ such that $\alpha_i \neq 0$, and leaves all other positions fixed. Then, we define $\boldsymbol{\Sigma}_i = \begin{bmatrix}\boldsymbol{\Sigma}_i^{(1)} & 0 \\ 0 & \boldsymbol{\Sigma}_i^{(2)}\end{bmatrix}$. The following fact is easy to verify.

Fact 2.7.22. $\boldsymbol{\Pi}'_{i,l}\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i\boldsymbol{\Pi}'_{i,l}$ *and* $\boldsymbol{\Pi}'_i\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i\boldsymbol{\Pi}'_i$.

We are now ready to prove Lemma 2.7.17 which we restate for convenience.

Lemma 2.7.17. *For all integers $t \geq 1$, $\mathbb{E}\left\|\boldsymbol{\Delta}_1^{k,a,b}\right\|_t^t \leq \frac{(8dd_p)^t}{n^t}\,\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$.*

Proof. Firstly,

$$\begin{aligned}
\boldsymbol{\Delta}_1^{k,a,b} &= \mathbb{E}[((\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z))^2|Z] \\
&= \mathbb{E}[(\mathbf{D}(Z)-\mathbf{D}(Z'))\overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\cdot\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z'))|Z] \\
&= \mathbb{E}[(\mathbf{D}(Z)-\mathbf{D}(Z'))\mathbf{M}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z'))|Z]
\end{aligned}$$

where we define $\mathbf{M}(Z) = \overline{\mathbf{G}}_{k,a,b}(Z)\mathbf{D}(Z)\cdot\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)$. Recall that $Z' = Z^{(i)}$ for some i randomly chosen from $[n]$ uniformly. Observing that $\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}) = \boldsymbol{\Pi}_i(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))$ for all i, we get

$$\begin{aligned}
\boldsymbol{\Delta}_1^{k,a,b} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))\mathbf{M}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))]|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))\mathbf{M}(Z)(\mathbf{D}(Z)-\mathbf{D}(Z^{(i)}))\boldsymbol{\Pi}_i]|Z] \\
&\preceq 2\left(\mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Pi}_i]|Z] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z]\right) \\
&\preceq 2\left(\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Pi}_i] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z]\right) \\
&\preceq 2(\boldsymbol{\Delta}_{10}+\boldsymbol{\Delta}_{11})
\end{aligned}$$

where we define

$$\boldsymbol{\Delta}_{10} = \mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\overline{\mathbf{F}}_{k,a,b}^2\boldsymbol{\Pi}_i], \qquad \boldsymbol{\Delta}_{11} = \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\boldsymbol{\Pi}_i]|Z]$$

Invoking Lemma 2.2.4 over the interval $[0,\infty)$ with the convex continuous function $f(x) = x^t$,

$\mathbf{B}_i = \overline{\mathbf{F}}^2_{k,a,b}, \mathbf{A}_i = \frac{1}{\sqrt{d_p}}\mathbf{\Pi}_i$ where we observe that $\sum_{i=1}^n \mathbf{A}_i\mathbf{A}_i^T = \frac{1}{d_p}\sum_{i=1}^n \mathbf{\Pi}_i^2 \preceq \mathbf{I}$, we get

$$\begin{aligned}
\mathbb{E}\,\|\mathbf{\Delta}_{10}\|_t^t = \mathbb{E}\,\mathrm{tr}[\mathbf{\Delta}_{10}^t] = \mathbb{E}\,\mathrm{tr}[\Big(\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{\Pi}_i\overline{\mathbf{F}}^2_{k,a,b}\mathbf{\Pi}_i]\Big)^t] &= \frac{1}{n^t}\,\mathbb{E}\,\mathrm{tr}[\Big(\sum_{i=1}^n \mathbf{\Pi}_i\overline{\mathbf{F}}^2_{k,a,b}\mathbf{\Pi}_i\Big)^t] \\
&\le \frac{d_p^{t-1}}{n^t}\,\mathbb{E}\,\mathrm{tr}[\Big(\sum_{i=1}^n \mathbf{\Pi}_i\overline{\mathbf{F}}^{2t}_{k,a,b}\mathbf{\Pi}_i\Big)] \\
&\le \frac{d_p^{t-1}}{n^t}\,\mathbb{E}\,\mathrm{tr}[\Big(\sum_{i=1}^n \mathbf{\Pi}_i^2\Big)\overline{\mathbf{F}}^{2t}_{k,a,b}] \\
&\le \frac{d_p^{t}}{n^t}\,\mathbb{E}\,\mathrm{tr}[\overline{\mathbf{F}}^{2t}_{k,a,b}] \\
&= \frac{d_p^{t}}{n^t}\,\mathbb{E}\,\big\|\overline{\mathbf{F}}_{k,a,b}\big\|_{2t}^{2t}
\end{aligned}$$

Now, consider

$$\begin{aligned}
\mathbf{\Delta}_{11} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{\Pi}_i\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\mathbf{\Pi}_i]|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\sum_{l=1}^d \mathbf{\Pi}_{i,l})\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})(\sum_{l=1}^d \mathbf{\Pi}_{i,l})]|Z] \\
&\preceq d\cdot\mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\sum_{l=1}^d \mathbf{\Pi}_{i,l}\mathbf{D}(Z^{(i)})\mathbf{M}(Z)\mathbf{D}(Z^{(i)})\mathbf{\Pi}_{i,l}]|Z] \\
&= d\cdot\mathop{\mathbb{E}}_{i\in[n]}[\sum_{l=1}^d \frac{\mathbb{E}[Z_i^{2l}]}{Z_i^{2l}}\mathbf{\Pi}_{i,l}\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\mathbf{\Pi}_{i,l}] \\
&= \frac{d}{n}\sum_{i=1}^n\sum_{l=1}^d \frac{\mathbb{E}[Z_i^{2l}]}{Z_i^{2l}}\mathbf{\Pi}_{i,l}\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\mathbf{\Pi}_{i,l} \\
&= \frac{d}{n}\sum_{i=1}^n\sum_{l=1}^d \mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\mathbf{D}(Z)\mathbf{M}(Z)\mathbf{D}(Z)\mathbf{\Sigma}_i^\intercal\mathbf{\Pi}_{i,l} \\
&= \frac{d}{n}\sum_{i=1}^n\sum_{l=1}^d \mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}^2_{k,a,b}\mathbf{\Sigma}_i^\intercal\mathbf{\Pi}_{i,l}
\end{aligned}$$

We now invoke Lemma 2.2.4 on dd_p terms with $\mathbf{B}_{i,l} = \overline{\mathbf{F}}^2_{k,a,b}$ and $\mathbf{A}_{i,l} = \frac{1}{\sqrt{d_p}}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i$ where

we observe that

$$\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{A}_{i,l}\mathbf{A}_{i,l}^{T}=\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l}^{\intercal}=\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}^{2}\preceq\mathbf{I}$$

to get

$$\begin{aligned}\mathbb{E}\left\|\mathbf{\Delta}_{11}\right\|_t^t=\mathbb{E}\operatorname{tr}[\mathbf{\Delta}_{11}^t]&\leq\frac{d^t}{n^t}\mathbb{E}\operatorname{tr}[(\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^{2}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l})^t]\\&\leq\frac{(dd_p)^t}{n^t}\mathbb{E}\operatorname{tr}[\left(\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^{2t}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l}\right)]\\&=\frac{(dd_p)^t}{n^t}\mathbb{E}\operatorname{tr}[\left(\frac{1}{d_p}\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Pi}_{i,l}\mathbf{\Pi}_{i,l}\mathbf{\Sigma}_i\overline{\mathbf{F}}_{k,a,b}^{2t}\right)]\end{aligned}$$

To simplify this, we use Fact 2.7.22 to get

$$\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Sigma}_i^{\intercal}(\mathbf{\Pi}_{i,l})^2\mathbf{\Sigma}_i\preceq\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Sigma}_i^{\intercal}(\mathbf{\Pi}'_{i,l})^2\mathbf{\Sigma}_i=\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}'_{i,l}\mathbf{\Sigma}_i^{\intercal}\mathbf{\Sigma}_i\mathbf{\Pi}'_{i,l}=\sum_{i=1}^{n}\sum_{l=1}^{d}\mathbf{\Pi}'_{i,l}\mathbf{\Pi}'_{i,l}\preceq d_p\mathbf{I}$$

Therefore,

$$\mathbb{E}\left\|\mathbf{\Delta}_{11}\right\|_t^t\leq\frac{(dd_p)^t}{n^t}\mathbb{E}\operatorname{tr}[\overline{\mathbf{F}}_{k,a,b}^{2t}]=\frac{(dd_p)^t}{n^t}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$$

Putting them together, using Fact 2.2.3,

$$\begin{aligned}\mathbb{E}\left\|\mathbf{\Delta}_1^{k,a,b}\right\|_t^t&\leq 4^t(\mathbb{E}\left\|\mathbf{\Delta}_{10}\right\|_t^t+\mathbb{E}\left\|\mathbf{\Delta}_{11}\right\|_t^t)\\&\leq\frac{(8dd_p)^t}{n^t}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}\end{aligned}$$

■

We now restate and prove Lemma 2.7.18.

Lemma 2.7.18. *For all integers $t\geq 1$, $\mathbb{E}\left\|\mathbf{\Delta}_3^{k,a,b}\right\|_t^t\leq\frac{(4d_p)^t}{n^t}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$.*

Proof. Recall that $Z' = Z^{(i)}$ for i sampled uniformly from $[n]$. Then,

$$\begin{aligned}
\boldsymbol{\Delta}_3^{k,a,b} &= \mathbb{E}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z) - \mathbf{D}(Z')))^2|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)(\mathbf{D}(Z) - \mathbf{D}(Z^{(i)})))^2]|Z] \\
&= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[(\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}(Z)\boldsymbol{\Pi}_i(\mathbf{D}(Z) - \mathbf{D}(Z^{(i)})))^2]|Z]
\end{aligned}$$

where we use the fact that $\mathbf{D}(Z) - \mathbf{D}(Z^{(i)}) = \boldsymbol{\Pi}_i(\mathbf{D}(Z) - \mathbf{D}(Z^{(i)}))$ for all i. Define $\mathbf{M}(Z) = \mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}$ to get

$$\begin{aligned}
\boldsymbol{\Delta}_3^{k,a,b} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i(\mathbf{D}(Z) - \mathbf{D}(Z^{(i)}))^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z] \\
&\preceq 2(\mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z]) \\
&= 2(\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal] + \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z]) \\
&= 2(\boldsymbol{\Delta}_{30} + \boldsymbol{\Delta}_{31})
\end{aligned}$$

where we define

$$\boldsymbol{\Delta}_{30} = \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal], \qquad \boldsymbol{\Delta}_{31} = \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z]$$

We have

$$\begin{aligned}
\boldsymbol{\Delta}_{30} = \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z)^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal] &= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\mathbf{D}(Z)\boldsymbol{\Pi}_i\boldsymbol{\Pi}_i\mathbf{D}(Z)\mathbf{M}(Z)^\intercal] \\
&= \mathbf{M}(Z)\mathbf{D}(Z)(\frac{1}{n}\sum_{i=1}^n \boldsymbol{\Pi}_i^2)\mathbf{D}(Z)\mathbf{M}(Z)^\intercal \\
&\preceq \frac{d_p}{n}\mathbf{M}(Z)\mathbf{D}(Z)\mathbf{D}(Z)\mathbf{M}(Z)^\intercal \\
&= \frac{d_p}{n}\overline{\mathbf{F}}_{k,a,b}^2
\end{aligned}$$

For the other term, using Fact 2.7.22,

$$\begin{aligned}
\boldsymbol{\Delta}_{31} &= \mathbb{E}[\mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\mathbf{D}(Z^{(i)})^2\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal]|Z] \\
&= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i\boldsymbol{\Sigma}_i\mathbf{D}(Z)^2\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i\mathbf{M}(Z)^\intercal] \\
&\preceq \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Pi}_i'\boldsymbol{\Sigma}_i\mathbf{D}(Z)^2\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i'\mathbf{M}(Z)^\intercal] \\
&= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{M}(Z)\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i'\mathbf{D}(Z)^2\boldsymbol{\Pi}_i'\boldsymbol{\Sigma}_i\mathbf{M}(Z)^\intercal] \\
&= \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\boldsymbol{\Sigma}_i\boldsymbol{\Pi}_i'\mathbf{D}(Z)^2\boldsymbol{\Pi}_i'\boldsymbol{\Sigma}_i\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)]
\end{aligned}$$

Observe that $\overline{\mathbf{G}}_{k,a,b}\boldsymbol{\Sigma}_i = \overline{\mathbf{G}}_{k,a,b}$ because the entries of $\overline{\mathbf{G}}$ only depend on α and not on γ, so permuting the γs will not have any effect on the matrix. Therefore,

$$\begin{aligned}
\boldsymbol{\Delta}_{31} &\preceq \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\boldsymbol{\Pi}_i'\mathbf{D}(Z)^2\boldsymbol{\Pi}_i'\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)] \\
&\preceq \mathop{\mathbb{E}}_{i\in[n]}[\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)\boldsymbol{\Pi}_i'\boldsymbol{\Pi}_i'\mathbf{D}(Z)\overline{\mathbf{G}}_{k,a,b}\mathbf{D}(Z)] \\
&= \mathop{\mathbb{E}}_{i\in[n]}\overline{\mathbf{F}}_{k,a,b}\boldsymbol{\Pi}_i'\boldsymbol{\Pi}_i'\overline{\mathbf{F}}_{k,a,b} \\
&= \frac{1}{n}\sum_{i=1}^{n}\overline{\mathbf{F}}_{k,a,b}\boldsymbol{\Pi}_i'\boldsymbol{\Pi}_i'\overline{\mathbf{F}}_{k,a,b} \\
&\preceq \frac{d_p}{n}\overline{\mathbf{F}}_{k,a,b}^2
\end{aligned}$$

where we used the fact that $\sum_{i=1}^n \boldsymbol{\Pi}_i'\boldsymbol{\Pi}_i' \preceq d_p\mathbf{I}$. Putting them together,

$$\mathbb{E}\left\|\boldsymbol{\Delta}_3^{k,a,b}\right\|_t^t \le 2^t(\mathbb{E}\left\|\boldsymbol{\Delta}_{30}\right\|_t^t + \mathbb{E}\left\|\boldsymbol{\Delta}_{31}\right\|_t^t) \le 2^t\cdot 2\frac{d_p^t}{n^t}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t} \le \frac{(4d_p)^t}{n^t}\mathbb{E}\left\|\overline{\mathbf{F}}_{k,a,b}\right\|_{2t}^{2t}$$

■

2.8 Application: Sparse graph matrices

We now consider sparse graph matrices, i.e., the setting $G \sim \mathcal{G}_{n,p}$ for $p \leq \frac{1}{2}$. The main difference from dense graph matrices is the contribution of the edge factors. Naïvely bounding the contribution of each edge by it's absolute value, as explained in Section 2.5, each edge in the shape contributes a factor of $\sqrt{\frac{1-p}{p}}$. But in many cases, these bounds are not tight. In fact, they are not tight even in the basic case of the adjacency matrix. In this section, we obtain tighter bounds using our general recursion. As we will see, the improved bound will contain the edge factors only for edges within the vertex separator.

Let $\mathbf{M}_\tau$ be the graph matrix corresponding to shape τ where we use p-biased Fourier characters $G_{i,j}$. In this section, we obtain bounds on $\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t}$ and use it to obtain high probability bounds on $\|\mathbf{M}_\tau\|$. Since many of the details are similar to Section 2.4.2 and the proof of Theorem 2.4.9, we will pass lightly over some details. We recommend the reader to read that section first.

The $G_{i,j}$ correspond to the Z_is in Section 2.6 and $\mathbf{F}$ corresponds to $\mathbf{M}_\tau$. Let $\mathcal{I}$ denote the set of sub-tuples of $[n]$. Each nonzero entry of $\mathbf{M}_\tau$ is a homogenous polynomial of degree $|E(\tau)|$. If $E(\tau) = \emptyset$, then, $\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau = 0$ so we can focus on the case when τ has at least one edge. Moreover, since degree-0 vertices in $V(\tau) \setminus U_\tau \setminus V_\tau$ simply scale the matrix by a factor of at most n, we can handle them separately and for our main analysis, we assume there are no such vertices in τ.

We will use Theorem 2.6.6 but the matrices and the statement can be drastically simplified in our application. Instate the notation of Section 2.6. Since we are dealing with multilinear polynomials, in the definition of $\mathcal{K}$, we can restrict our attention to $\alpha \in \{0,1\}^{\binom{n}{2}}$ because for any other $\alpha \in \mathbb{N}^n$, the corresponding row or column of $\mathbf{G}_{a+b,a,b}$ and hence $\mathbf{F}_{a+b,a,b}$, will be 0. So, we can accordingly redefine $\mathcal{K}$ to only contain these (α, γ), hence $\mathcal{K} \subseteq \{0,1\}^n \times \{0,1\}^n$.

Next, the diagonal matrices $\mathbf{D}_1, \mathbf{D}_2$ will both be equal to the diagonal matrix $\mathbf{D} \in$

$\mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}} \times \mathbb{R}[Z]^{\mathcal{I}\times\mathcal{K}}$ with nonzero entries

$$\mathbf{D}[(I,\alpha,\gamma),(I,\alpha,\gamma)] = \sqrt{\mathbb{E}[\prod_{i,j} G_{ij}^{2\alpha_{ij}(1-\gamma)_{ij}}]} \prod_{i,j} G_i^{\alpha_{ij}\gamma_{ij}} = \prod_{i,j} G_i^{\alpha_{ij}\gamma_{ij}}$$

where we used the fact that for any i, j, $\mathbb{E}[G_{ij}^2] = 1$.

For integers $a, b \geq 0$ such that $a + b = |E(\tau)|$, define the matrix $\mathbf{M}_{\tau,a,b}$ to be the matrix $\mathbf{G}_{a+b,a,b}$. We use this notation in order to be streamlined with Section 2.4.2. That is, $\mathbf{M}_{\tau,a,b}$ has rows and columns indexed by $\mathcal{I} \times \mathcal{K}$ such that for all $(I,\alpha_1,\gamma_1),(J,\alpha_2,\gamma_2) \in \mathcal{I}\times\mathcal{K}$,

$$\mathbf{M}_{\tau,a,b}[(I,\alpha_1,\gamma_1),(J,\alpha_2,\gamma_2)] = \begin{cases} \nabla_{\alpha_1+\alpha_2}\mathbf{M}_\tau[I,J] & \text{if } |\alpha_1|_0 = a, |\alpha_2|_0 = b, \alpha_1\cdot\alpha_2 = 0 \\ 0 & \text{o.w.} \end{cases}$$

This is almost identical to the $\mathbf{M}_{\tau,a,b}$ matrix defined in Section 2.4.2, with the difference being that the row and column indices now have γ in them. Therefore, for $I, J \in \mathcal{I}, (\alpha_1,\gamma_1),(\alpha_2,\gamma_2) \in \mathcal{K}$ such that $|\alpha_1|_0 = a, |\alpha_2|_0 = b, \alpha_1\cdot\alpha_2 = 0$, the entry in row (I,α_1,γ_1) and column (J,α_2,γ_2) is the number of realizations φ of τ such that

- U_τ, V_τ map to I, J respectively under φ, and
- Under φ, the edges of τ map to the edges in α_1 and α_2 viewed as a set.

By Theorem 2.6.6, for integers $t \geq 1$,

$$\begin{aligned} \mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} &\leq \sum_{a,b\geq 0, a+b\geq 1} (Ct^2 dd_p^4)^{(a+b)t}\,\mathbb{E}\left\|\mathbf{F}_{a+b,a,b}\right\|_{2t}^{2t} \\ &= \sum_{a,b\geq 0, a+b=|E(\tau)|} (Ct^2|E(\tau)|^4)^{t|E(\tau)|}\,\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t} \end{aligned}$$

for an absolute constant $C > 0$.

Now, we would like to analyze $\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t}$. Just as in the proof of Theorem 2.4.9, let P specify which edges of $E(\tau)$ go to α_1, α_2 respectively and in what order. Moreover, we

now store extra information in P that indicates which entries of γ_1, γ_2 (relative to α_1, α_2) are set to 1. Let the set of such information P be denoted $\mathcal{P}$, then $|\mathcal{P}| \leq (4|E(\tau)|)^{t|E(\tau)|} 2^{|E(\tau)|}$. Thus,

$$\mathbb{E} \left\| \mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D} \right\|_{2t}^{2t} \leq (8|E(\tau)|)^{t|E(\tau)|} \sum_{P \in \mathcal{P}} \mathbb{E} \left\| \mathbf{D}\mathbf{M}_{\tau,a,b,P}\mathbf{D} \right\|_{2t}^{2t}$$

where we define $\mathbf{M}_{\tau,a,b,P}$ similar to $\mathbf{M}_{\tau,a,b}$ with the extra condition that $\varphi, \alpha_1, \alpha_2, \gamma_1, \gamma_2$ must respect P.

At this point, in contrast to the proof of Theorem 2.4.9, note that the matrices $\mathbf{M}_{\tau,a,b,P}$ here have rows and columns indexed by $\mathcal{I} \times \mathcal{K}$. We will again define the shape τ_P that is equal to the nonzero block of the matrix $\mathbf{D}\mathbf{M}_{\tau,a,b,P}\mathbf{D}$, upto renaming of the rows and columns. $V(\tau_P), U_{\tau_P}, V_{\tau_P}$ are defined the same way as in Section 2.4.2 but to incorporate the action of $\mathbf{D}$ on these entries, we simply keep the edges that are active in γ_1 or γ_2, as prescribed by P. For an illustration, see Fig. 2.6.

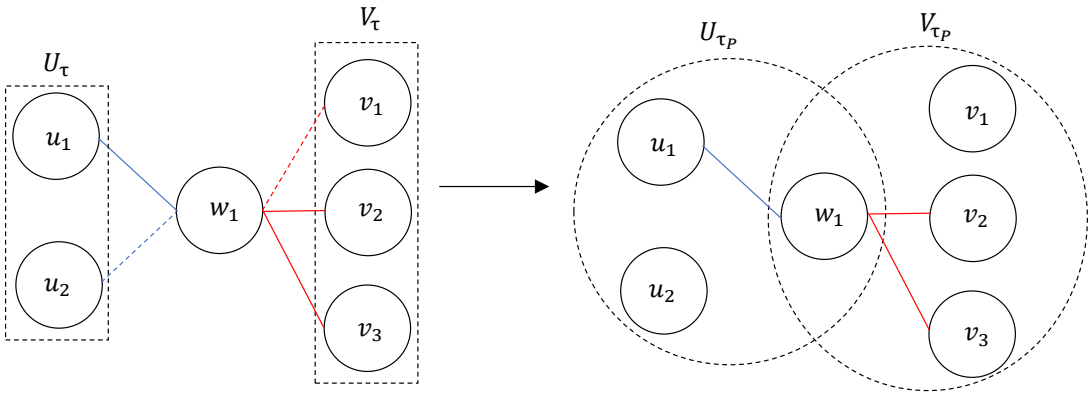

Figure 2.6: An example illustrating how τ_P is defined. In this example, P constraints the blue and red edges to go to α_1 and α_2 respectively. Moreover, P indicates that some edges are active in γ_1, γ_2 (indicated by a solid edge) and some are not active (indicated by a dashed edge) in γ_1, γ_2. We keep the solid edges in τ_P. U_{τ_P}, V_{τ_P} also have an ordering on the vertices (not shown here).

Then, by similar renaming of the rows and columns of $\mathbf{D}\mathbf{M}_{\tau,a,b,P}\mathbf{D}$ and dropping the γs,

we obtain $\mathbf{M}_{\tau_P}$. We therefore obtain the bound

$$\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t} \leq (8|E(\tau)|)^{t|E(\tau)|}\sum_{P\in\mathcal{P}}\mathbb{E}\left\|\mathbf{M}_{\tau_P}\right\|_{2t}^{2t}$$

We would like to analyze norm bounds on the matrices $\mathbf{M}_{\tau_P}$. Observe that τ_P are shapes with the properties

- there are no vertices in $V(\tau_P)\setminus U_{\tau_P}\setminus V_{\tau_P}$
- each edge is either entirely contained in U_{τ_P} or entirely contained in V_{τ_P}

Call such shapes *simple*.

In the following lemma, whose proof is deferred to the next section, we prove norm bounds on simple shapes. Recall that in Lemma 2.4.10, we analyzed the norm bounds of simple shapes with no edges (because in this case, the graph distribution doesn't matter). The analysis for simple shapes is very similar but this time, we use scalar concentration tools to bound the Frobenius norm.

For a set S of vertices, denote by $E(S)$ the set of edges with both endpoints in S.

Lemma 2.8.1. *For all even integers $t \geq 2$, if τ is a simple shape,*

$$\mathbb{E}\left\|\mathbf{M}_\tau\right\|_{2t}^{2t} \leq \left(n^{|V(\tau)|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}\right)\max_{U_\tau\cap V_\tau\subseteq S\subseteq V(\tau)}\left(\frac{1-p}{p}\right)^{t|E(S)|}n^{t(|V(\tau)|-|S|)}$$

for an absolute constant $C>0$.

For simple shapes, the main difference from norm bounds on corresponding dense graph matrices is that each edge within S contributes a factor of $\sqrt{\frac{1-p}{p}}$. Edge contributions are unavoidable when handling sparse graph matrices, but we have identified that we need not consider all edges in the shape but only a subset of it.

Using this lemma, we can obtain norm bounds on general graph matrices. We recall the definition of a vertex separator.

Definition 2.4.8 (Vertex separator)**.** *For a shape τ, define a vertex separator to be a subset of vertices $S \subseteq V(\tau)$ such that there is no path from U_τ to V_τ in $\tau \setminus S$, which is the shape obtained by deleting all the vertices of S (including all edges they're incident on).*

Let I_τ be the set of isolated vertices (vertices of degree 0) in $V(\tau) \setminus U_\tau \setminus V_\tau$, so they essentially scale the matrix by a scalar factor. We now state the main theorem of this section.

Theorem 2.8.2. *For all even integers $t \geq 2$, for any shape τ,*

$$\mathbb{E}\,\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\|_{2t}^{2t} \leq \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(Ct^3|E(\tau)|^5)^{t|E(\tau)|}\right) \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|+|I_\tau|)}$$

where the maximum is over all vertex separators S.

To interpret this bound, if we assume that there are a constant number of vertices in τ, then by choosing $t \approx \text{polylog}(n)$, we get

$$\|\mathbf{M}_\tau\| = \widetilde{\mathrm{O}}\left(\max_{\text{vertex separator } S} \left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|} \sqrt{n}^{|V(\tau)|-|S|+|I_\tau|}\right)$$

with high probability, where $\widetilde{\mathrm{O}}$ hides logarithmic factors. This result follows from Theorem 2.8.2 if τ has at least one edge, but also applies if τ has no edges, in which case we can directly use the far simpler Lemma 2.4.10. A precise form of the above characterization is given in Corollary 2.8.3.

Theorem 2.8.2 gives us the right dependence on p, n for norm bounds in the case of sparse graph matrices. The same bound, upto lower order terms, was also obtained in [95] via the trace power method, where they use these bounds to prove semidefinite-programming lower bounds for the maximum independent set problem on sparse graphs.

Proof of Theorem 2.8.2. If $E(\tau) = \emptyset$, then $\mathbf{M}_\tau = \mathbb{E}\,\mathbf{M}_\tau$ and we are done. So, assume $E(\tau) \neq \emptyset$. Since vertices in I_τ only scale the matrix by a factor of at most n, we can

handle them separately and our bound has the appropriate power of n coming from these. Therefore, we can assume $I_\tau = \emptyset$. Continuing our prior discussions, for an absolute constant $C_1 > 0$,

$$\begin{aligned}\mathbb{E}\left\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t} &\leq \sum_{a,b\geq 0, a+b=|E(\tau)|} (C_1 t^2 |E(\tau)|^4)^{t|E(\tau)|}\,\mathbb{E}\left\|\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}\right\|_{2t}^{2t} \\ &\leq \sum_{a,b\geq 0, a+b=|E(\tau)|} (C_1 t^2 |E(\tau)|^4)^{t|E(\tau)|}(8|E(\tau)|)^{t|E(\tau)|} \sum_{\psi\in\Gamma_{a,b}} \mathbb{E}\left\|\mathbf{M}_\psi\right\|_{2t}^{2t}\end{aligned}$$

where $\Gamma_{a,b}$ are the set of simple shapes we obtain for $\mathbf{D}\mathbf{M}_{\tau,a,b}\mathbf{D}$, as per our discussion above. Using Lemma 2.8.1, for an absolute constant $C_2 > 0$, we have

$$\begin{aligned}&\mathbb{E}\left\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t} \\ &\leq \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(C_2 t^3|E(\tau)|^5)^{t|E(\tau)|}\right) \sum_{a,b\geq 0, a+b=|E(\tau)|} \sum_{\psi\in\Gamma_{a,b}} \max_{U_\psi\cap V_\psi \subseteq S\subseteq V(\psi)} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\psi)|-|S|)}\end{aligned}$$

For any a, b, consider any simple shape $\psi \in \Gamma_{a,b}$ that can be obtained. As observed in the proof of Theorem 2.4.9 (see in particular Fig. 2.5), $U_\psi \cap V_\psi$ must be a vertex separator of τ. Therefore, any $S \supseteq U_\psi \cap V_\psi$ must be a vertex separator of τ. It's easy to see that as S ranges over all sets such that $U_\psi \cap V_\psi \subseteq S \subseteq V(\psi)$, it ranges over all vertex separators of τ.

Also, the number of different ψ is at most $4^{|E(\tau)|}$ since each edge can go either to U_ψ or V_ψ and for each such choice, it can either be active in γ or not. Therefore,

$$\begin{aligned}&\mathbb{E}\left\|\mathbf{M}_\tau - \mathbb{E}\,\mathbf{M}_\tau\right\|_{2t}^{2t} \\ &\leq \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(C_2 t^3|E(\tau)|^5)^{t|E(\tau)|}\right) 4^{|E(\tau)|} \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)} \\ &\leq \left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(Ct^3|E(\tau)|^5)^{t|E(\tau)|}\right) \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)}\end{aligned}$$

for an absolute constant $C > 0$. ∎

The following corollary obtains high probability norm bounds for norms of graph ma-

trices via Markov's inequality. We assume the graph has at least one edge, otherwise it is deterministic and its norm bound was already analyzed in Lemma 2.4.10, Corollary 2.4.11, where we observe that the distinction between sparse and dense graph matrices does not matter if the random matrix is deterministic.

Corollary 2.8.3. *For a shape τ with at least one edge, for any constant $\varepsilon > 0$, with probability $1 - \varepsilon$,*

$$\|\mathbf{M}_\tau\| \leq \left(|V(\tau)|^{|V(\tau)|/2}(C|E(\tau)|^5 \log^3(n^{|V(\tau)|}/\varepsilon))^{|E(\tau)|/2}\right) \cdot \max_{\text{vertex separator } S} \left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|} \sqrt{n}^{|V(\tau)-|S|+|I_\tau|}$$

for an absolute constant $C > 0$.

Proof. Since $|E(\tau)| \geq 1$, $\mathbb{E}\,\mathbf{M}_\tau = 0$. By an application of Markov's inequality,

$$\begin{aligned} Pr[\|\mathbf{M}_\tau\| \geq \theta] &\leq Pr[\|\mathbf{M}_\tau\|_{2t}^{2t} \geq \theta^{2t}] \\ &\leq \theta^{-2t}\,\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} \\ &\leq \theta^{-2t}\left(n^{|V(\tau)|}|V(\tau)|^{t|V(\tau)|}(C't^3|E(\tau)|^5)^{t|E(\tau)|}\right) \max_{\text{vertex separator } S} \left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|+|I_\tau|)} \end{aligned}$$

for an absolute constant $C' > 0$. We now set

$$\begin{aligned} \theta =& \left(\varepsilon^{-1/(2t)}(C'')^{|E(\tau)|} n^{|V(\tau)|/(2t)}|V(\tau)|^{|V(\tau)|/2} t^{3|E(\tau)|/2}|E(\tau)|^{5|E(\tau)|/2}\right) \\ & \cdot \max_{\text{vertex separator } S} \left(\sqrt{\frac{1-p}{p}}\right)^{|E(S)|} \sqrt{n}^{|V(\tau)-|S|+|I_\tau|} \end{aligned}$$

for an absolute constant $C'' > 0$, to make this expression at most ε. Set $t = \frac{1}{2}\log(n^{|V(\tau)|}/\varepsilon)$ to complete the proof. ∎

2.8.1 Norm bounds on simple graph matrices

In this section, we will prove Lemma 2.8.1. First, we recall the following scalar concentration result from [166].

Schudy-Sviridenko moment bound

The definitions and main bound in this section are from [166].

Definition 2.8.4. *A random variable Z is central moment bounded with real parameter $L > 0$ if for any integer $i \geq 1$,*

$$\mathbb{E}[|Z - \mathbb{E}[Z]|^i] \leq i \cdot L \cdot \mathbb{E}[|Z - \mathbb{E}[Z]|^{i-1}]$$

Proposition 2.8.5. *The p-biased Bernoulli random variable Z is central moment bounded with real parameter $L = \sqrt{\frac{1-p}{p}}$.*

Proof. We have $\mathbb{E}[Z] = 0$ and for $p \leq \frac{1}{2}$, $|Z| \leq \sqrt{\frac{1-p}{p}}$, therefore,

$$\begin{aligned}
\mathbb{E}[|Z - \mathbb{E}[Z]|^i] &= p\sqrt{\frac{p}{1-p}}^i + (1-p)\sqrt{\frac{1-p}{p}}^i \\
&\leq \sqrt{\frac{1-p}{p}}\left(p\sqrt{\frac{p}{1-p}}^{i-1} + (1-p)\sqrt{\frac{1-p}{p}}^{i-1}\right) \\
&= \sqrt{\frac{1-p}{p}}\,\mathbb{E}[|Z - \mathbb{E}[Z]|^{i-1}]
\end{aligned}$$

therefore, we can take $L = \sqrt{\frac{1-p}{p}}$. ∎

For a given multilinear polynomial $f(x)$ on variables $x_1, \ldots, x_n$, we can naturally associate with it a hypergraph H on vertices $[n]$ and weighted hyperedges $E(H)$ where each $h \in E(H)$ corresponds to a distinct term of $f(x)$. Each hyperedge h is a subset $V(h)$ of vertices and has a real valued weight w_h which is the coefficient of that monomial in f. Therefore,

$$f(x) = \sum_{h \in E(H)} w_h \prod_{v \in V(h)} x_v$$

Assume f has degree d_p, then each hyperedge of H has at most d_p vertices.

Now, for a given collection of independent random variables $Y_1, \dots, Y_n$, a multilinear poynomial f with associated hypergraph H and weights w, and an integer $r \geq 0$, define

$$\mu_r(f, Y) = \max_{S \subseteq [n], |S| = r} \left(\sum_{h \in E(H), S \subseteq V(h)} |w_h| \prod_{v \in V(h) \setminus S} \mathbb{E}[|Y_v|] \right)$$

Lemma 2.8.6 ([166], Lemma 5.1)**.** *Given n independent central moment bounded random variables $Y_1, \dots, Y_n$ with the same parameter $L > 0$ and a degree d_p multilinear polynomial $f(x)$. Let $t \geq 2$ be an even integer, then*

$$\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^t] \leq \max \left\{ \left(\sqrt{t R_4^{d_p} \mathrm{Var}[f(Y)]} \right)^t, \max_{r \in [d_p]} (t^r R_4^{d_p} L^r \mu_r(f, Y))^t \right\}$$

where $R_4 \geq 1$ is some absolute constant.

In our setting, we can also bound the variance in terms of the μ_r as was shown in [166], which will simplify our calculations.

Lemma 2.8.7 ([166], Lemma 1.5)**.** *For the same setting as in Lemma 2.8.6,*

$$\mathrm{Var}[f(Y)] \leq 2 d_p 4^{d_p} \max_{r \in [d_p]} (\mu_0(f, Y) \mu_r(f, Y) 4^r L^r)$$

Proof of Lemma 2.8.1

We are ready to prove Lemma 2.8.1 which we restate for convenience.

Lemma 2.8.1. *For all even integers $t \geq 2$, if τ is a simple shape,*

$$\mathbb{E} \|\mathbf{M}_\tau\|_{2t}^{2t} \leq \left(n^{|V(\tau)|} (Ct)^{t|E(\tau)|} |V(\tau)|^{t|V(\tau)|} \right) \max_{U_\tau \cap V_\tau \subseteq S \subseteq V(\tau)} \left(\frac{1-p}{p} \right)^{t|E(S)|} n^{t(|V(\tau)| - |S|)}$$

for an absolute constant $C > 0$.

We will prove it the same way as Lemma 2.4.10, by bounding the schatten norm of each

diagonal block by an appropriate power of its Frobenius norm. In this case, to bound the expected power of the Frobenius norm, we use the scalar concentration inequality from the previous section.

Proof of Lemma 2.8.1. First, we note that $\mathbf{M}_\tau$ has a block diagonal structure indexed by the realizations of the set of common vertices $S_0 = U_{\tau_P} \cap V_{\tau_P}$. For $T \in [n]^{S_0}$, let $\mathbf{M}_{\tau,T}$ be the block of $\mathbf{M}_\tau$ with $\varphi(S_0) = T$. Then, $\mathbf{M}_{\tau,T}\mathbf{M}_{\tau,T'}^\intercal = \mathbf{M}_{\tau,T}^\intercal \mathbf{M}_{\tau,T'} = 0$ for $T \neq T'$ and so,

$$\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} = \sum_{T\in[n]^{S_0}} \mathbb{E}\,\left\|\mathbf{M}_{\tau,T}\right\|_{2t}^{2t} \leq \sum_{T\in[n]^{S_0}} \mathbb{E}(\left\|\mathbf{M}_{\tau,T}\right\|_2^2)^t$$

where we bounded the Schatten norm by a power of the Frobenius norm.

Fix $T \in [n]^{S_0}$ and consider $\mathbb{E}\left\|\mathbf{M}_{\tau,T}\right\|_2^2$. Let $\mathcal{R}$ be the set of realizations φ of τ such that $\varphi(S_0) = T$. Then, for $\varphi \in \mathcal{R}$ and $e \in E(S_0)$, the value of $\varphi(e)$ is fixed. Using this,

$$\begin{aligned}
\left\|\mathbf{M}_{\tau,T}\right\|_2^2 &= \sum_{\varphi\in\mathcal{R}} \prod_{e\in E(\tau)} G_{\varphi(e)}^2 \\
&= \prod_{e\in E(S_0)} G_{\varphi(e)}^2 \sum_{\varphi\in\mathcal{R}} \prod_{e\in E(\tau)\setminus E(S_0)} G_{\varphi(e)}^2 \\
&\leq L^{|E(S_0)|} \sum_{\varphi\in\mathcal{R}} \prod_{e\in E(\tau)\setminus E(S_0)} G_{\varphi(e)}^2
\end{aligned}$$

where $L = \frac{1-p}{p}$ is an upper bound on G_{ij}^2 for $p \leq \frac{1}{2}$. Define the quantity

$$A = \max_{S_0\subseteq S\subseteq V(\tau)} L^{|E(S)|} n^{|V(\tau)|-|S|}$$

Claim 2.8.8. $\mathbb{E}(\left\|\mathbf{M}_{\tau,T}\right\|_2)^t \leq (Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t$ *for an absolute constant* $C > 0$.

Using this claim, we have

$$\begin{aligned}\mathbb{E}\,\|\mathbf{M}_\tau\|_{2t}^{2t} &\le \sum_{T\in[n]^{S_0}} \mathbb{E}(\left\|\mathbf{M}_{\tau,T}\right\|_2)^t \\ &\le n^{|S_0|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t \\ &= n^{|V(\tau)|}(Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}\max_{U_\tau\cap V_\tau\subseteq S\subseteq V(\tau)}\left(\frac{1-p}{p}\right)^{t|E(S)|} n^{t(|V(\tau)|-|S|)}\end{aligned}$$

as required. ∎

It remains to prove the claim.

Proof of Claim 2.8.8. For $1 \le i,j \le n$, define the variables $Y_{ij} = G_{ij}^2$ with $\mathbb{E}[|Y_{ij}|] = 1$. Let $f(Y)$ be the polynomial $L^{|E(S_0)|}\sum_{\varphi\in\mathcal{R}}\prod_{e\in E(\tau)\setminus E(S_0)} Y_{\varphi(e)}$. It suffices to prove that $\mathbb{E}[f(Y)^t] \le (Ct)^{t|E_1|}A^t$.

We will first prove that $\mathbb{E}[(f(Y)-\mathbb{E}[f(Y)])^t] \le (C't)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t$ for a sufficiently large constant $C' > 0$.

f is a homogeneous multilinear polynomial of degree $|E(\tau)\setminus E(S_0)|$. If we had $E(\tau)\setminus E(S_0) = \emptyset$, then f is a constant and so, the inequality is obvious because $f(Y) = \mathbb{E}[f(Y)]$. Now, assume $E(\tau)\setminus E(S_0) \neq \emptyset$. We invoke Lemma 2.8.6. Let f have associated hypergraph H and weights w. Then,

$$\mathbb{E}[|f(Y)-\mathbb{E}[f(Y)]|^t] \le \max\left\{\left(\sqrt{tR_4^{|E(\tau)\setminus E(S_0)|}\mathrm{Var}[f(Y)]}\right)^t, \max_{r\in[|E(\tau)\setminus E(S_0)|]}(t^rR_4^{|E(\tau)\setminus E(S_0)|}L^r\mu_r(f,Y))^t\right\}$$

For all $r \ge 0$, we will prove that $L^r\mu_r(f,Y) \le |V(\tau)|^{|V(\tau)|}A$. By definition,

$$\mu_r(f,Y) = \max_{F\subseteq\binom{[n]}{2},|F|=r}\sum_{h\in E(H),F\subseteq V(h)}|w_h|$$

Consider any set of edge labels $F \subseteq \binom{[n]}{2}, |F| = r$. Then, $\sum_{h\in E(H),F\subseteq V(h)}|w_h|$ is at most $L^{|E(S_0)|}c$ where c is the number of realizations $\varphi\in\mathcal{R}$ such that $\varphi(E(\tau))$ contains F. Suppose

F contains v new labels apart from $\varphi(S_0) = T$. Then $c \leq |V(\tau)|^v n^{|V(\tau)|-|S_0|-v}$ because we can first choose and label the set of vertices that get these v labels and then label the remaining vertices freely, each of which has at most n choices.

Observe that $L^{|E(S_0)|}L^r n^{|V(\tau)|-|S_0|-v} \leq A$ because in the definition of S, we can set S to be the union of S and any valid choice of these v vertices. Putting this together, we get

$$\begin{aligned} L^r \mu_r(f, Y) &\leq L^r \max_{F \subseteq \binom{[n]}{2}, |F|=r} \sum_{h \in E(H), F \subseteq V(h)} |w_h| \\ &\leq |V(\tau)|^{|V(\tau)|} A \end{aligned}$$

which implies

$$\max_{r \in [|E(\tau) \setminus E(S_0)|]} (t^r R_4^{|E(\tau) \setminus E(S_0)|} L^r \mu_r(f, Y))^t \leq |V(\tau)|^{t|V(\tau)|} (R_4 t)^{t|E(\tau)|} A^t$$

and using Lemma 2.8.7,

$$\begin{aligned} \mathrm{Var}[f(Y)] &\leq 2|E(\tau)| 4^{|E(\tau)|} \max_{r \in [|E(\tau) \setminus E(S_0)|]} (\mu_0(f, Y) \mu_r(f, Y) 4^r L^r) \\ &\leq 2|E(\tau)| 16^{|E(\tau)|} |V(\tau)|^{2|V(\tau)|} A^2 \end{aligned}$$

Putting them together, we get

$$\begin{aligned} \mathbb{E}[(f(Y) - \mathbb{E}[f(Y)])^t] &\leq \max \left\{ \left(\sqrt{2t R_4^{|E(\tau)|} |E(\tau)| 16^{|E(\tau)|} |V(\tau)|^{2|V(\tau)|} A^2} \right)^t, |V(\tau)|^{t|V(\tau)|} (R_4 t)^{t|E(\tau)|} A^t \right\} \\ &\leq (C' t)^{t|E(\tau)|} |V(\tau)|^{t|V(\tau)|} A^t \end{aligned}$$

for an absolute constant $C' > 0$.

Finally, $\mathbb{E}[f(Y)] \leq L^{|E(S_0)|}|\mathcal{R}| \leq L^{|E(S_0)|}n^{|V(\tau)\setminus S_0|} \leq A$ which gives

$$\begin{aligned}\mathbb{E}[f(Y)^t] &\leq 2^t(\mathbb{E}[(f(Y) - \mathbb{E}[f(Y)])^t] + \mathbb{E}[f(Y)]^t) \\ &\leq 2^t((C't)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t + A^t) \\ &\leq (Ct)^{t|E(\tau)|}|V(\tau)|^{t|V(\tau)|}A^t\end{aligned}$$

for an absolute constant $C > 0$. ■

CHAPTER 3

THE SUM OF SQUARES HIERARCHY

In this chapter, we formally introduce the Sum of Squares (SoS) hierarchy. Then, we take a minor detour and define low-degree distinguishers and related concepts for hypothesis testing, which will set the stage for us to discuss SoS lower bounds. We then go back to SoS and discuss the heuristic known as pseudo-calibration, that will be a basic ingredient we use in our SoS lower bounds. We finally show a formal connection between pseudo-calibration and low-degree distinguishers.

3.1 The Sum of Squares hierarchy

We start by defining convex relaxations for polynomial optimization problems. The SoS hierarchy will then be a special family of convex relaxations. For a more detailed treatment, see e.g. [118, 17, 64].

3.1.1 Polynomial optimization and convex relaxations

In polynomial optimization, we are given multivariate polynomials $p, g_1, \ldots, g_m$ on n variables $x_1, \ldots, x_n$ taking real values, denoted collectively by x, and the task is to:

$$\text{maximize } p(x) \text{ such that } g_1(x) = 0, \ldots, g_m(x) = 0$$

In general, we could also allow inequality constraints, e.g., $g_i(x) \geq 0$. For technical convenience in our setup, we work only with equality constraints but much of the theory generalizes, with some modifications, when we have inequality constraints instead. An alternate approach is to replace each inequality $g_i(x) \geq 0$ by $g_i(x) = y^2$ where y is a new variable that we can introduce.

In this formulation, many optimization problems can be formulated as polynomial optimization problems.

Example 3.1.1 (Maximum Cut)**.** *Given a graph $G = (V, E)$, we would like to partition the set of vertices into two subsets such that the number of edges with endpoints in different subsets is maximized. To formulate this as a polynomial optimization problem, let the graph have n vertices and let $x_1, \ldots, x_n$ be variables, one for each vertex. We wish to enforce $x_i \in \{-1, 1\}$ where all vertices i with $x_i = -1$ form one subset and the rest form the other subset. We can enforce this set containment constraint via the polynomial constraint $x_i^2 = 1$. For any edge $(i, j) \in E$, it is cut if and only if $x_i x_j = -1$. Therefore, the total number of edges cut is $\sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j)$. The polynomial formulation therefore becomes*

$$\max_{x\in\mathbb{R}^n} \sum_{(i,j)\in E} \frac{1}{2}(1 - x_i x_j) \text{ such that}$$
$$x_i^2 = 1 \text{ for all } i \leq n$$

Example 3.1.2 (Maximum Clique)**.** *Given a graph $G = (V, E)$, we would like to find the maximize size subset of vertices that form a clique. Again, let $x_1, \ldots, x_n$ be variables, one for each vertex. This time, we wish to enforce $x_i \in \{0, 1\}$, which we can easily do so using the polynomial constraint $x_i^2 = x_i$, with the intent being that all vertices i with $x_i = 1$ form a clique. To enforce this clique constraint, we can add the polynomial constraint $x_i x_j = 0$ for all non-edges $(i, j) \notin E$. Finally, to maximize the size of the subset, we simply maximize $\sum_{i\leq n} x_i$. Therefore, the polynomial optimization is*

$$\max_{x\in\mathbb{R}^n} \sum_{i\leq n} x_i \text{ such that}$$
$$x_i x_j = 0 \text{ for all } (i, j) \notin E$$
$$x_i^2 = x_i \text{ for all } i \leq n$$

There can be other equivalent formulations for these problems. In general, many optimization problems can be stated in this manner, therefore generic polynomial optimization contains a large class of fundamental problems that appear in computer science.

Since exactly solving maximum cut or maximum clique is NP-hard [97], exactly solving these polynomial optimization problems is also NP-hard. Therefore, we turn to convex relaxations.

A convex relaxation of a polynomial optimization problem widens the search space of solution vectors x into a larger space that one can efficiently optimize over. We will describe one way to do this. We identify a convex space $\mathcal{C}$ that contains the space $\mathcal{S} = \{g_1(x) = 0, \ldots, g_m(x) = 0\}$ upto a map, that is, for each $x \in \mathcal{S}$, there exists a corresponding $y \in \mathcal{C}$ such that y is a representative of x. We also identify a convex function $\tilde{p}(y)$ such that if y is a representative of x, then $\tilde{p}(y) = p(x)$. Then, we simply optimize $\tilde{p}(y)$ over $\mathcal{C}$. There has been significant work on efficiently optimizing a convex function over a convex body, which is possible under reasonable assumptions (see e.g. [141]). It's clear that from the above properties, the solution we get is at least as large as the optimal solution (in the case of maximization), but it comes with the advantage that it is efficiently computable. It is desirable to design convex relaxations for problems that yield good approximations. The SoS hierarchy is a family of such convex relaxations.

3.1.2 Sum of Squares relaxations

The SoS hierarchy, sometimes referred to as the Lasserre hierarchy, was first independently studied by [144, 117, 169] and has been studied in other contexts by [135, 73, 74]. It is a family of convex relaxations for polynomial optimization, parameterized by an integer known as it's degree. As we increase the degree, we get progressively tighter relaxations, but requiring longer times to optimize over.

We now formally describe the Sum of Squares hierarchy, via the so-called pseudo-expectation

operator view.

Definition 3.1.3 (Pseudo-expectation values)**.** *Given multivariate polynomial constraints* $g_1 = 0, \ldots, g_m = 0$ *on* n *variables* $x_1, \ldots, x_n$*, degree* d *pseudo-expectation values are a linear map* $\widetilde{\mathbb{E}}$ *from polynomials of* $x_1, \ldots, x_n$ *of degree at most* d *to* $\mathbb{R}$ *satisfying the following conditions:*

1. $\widetilde{\mathbb{E}}[1] = 1$,
2. $\widetilde{\mathbb{E}}[f \cdot g_i] = 0$ *for every* $i \in [m]$ *and polynomial* f *such that* $\deg(f \cdot g_i) \leq d$.
3. $\widetilde{\mathbb{E}}[f^2] \geq 0$ *for every polynomial* f *such that* $\deg(f^2) \leq d$.

Any linear map $\widetilde{\mathbb{E}}$ satisfying the above properties is known as a degree d pseudoexpectation operator satisfying the constraints $g_1 = 0, \ldots, g_m = 0$.

Definition 3.1.4 (Degree d SoS)**.** *The degree* d *SoS relaxation for the polynomial optimization problem*

$$\textit{maximize } p(x) \textit{ such that } g_1(x) = 0, \ldots, g_m(x) = 0$$

is the program that maximizes $\widetilde{\mathbb{E}}[p(x)]$ *over all degree* d *pseudoexpectation operators* $\widetilde{\mathbb{E}}$ *satisfying the constraints* $g_1 = 0, \ldots, g_m = 0$.

The intuition behind pseudo-expectation values is that the conditions on the pseudo-expectation values are conditions that would be satisfied by any actual expected values over a distribution of solutions, so optimizing over pseudo-expectation values gives a relaxation of the problem.

The main observation is that the SoS relaxation can be efficiently solved! This is because the conditions on pseudo-expectation values can be captured by a semidefinite program. In particular, Item 3 in Definition 3.1.3 can be reexpressed in terms of a matrix called the moment matrix.

Definition 3.1.5 (Moment Matrix of $\widetilde{\mathbb{E}}$)**.** *Given a degree d pseudo-expectation operator $\widetilde{\mathbb{E}}$, define the associated moment matrix Λ to be a matrix with rows and columns indexed by monomials p and q such that the entry corresponding to row p and column q is*

$$\Lambda[p, q] := \widetilde{\mathbb{E}}\,[pq]\,.$$

It is easy to verify that Item 3 in Definition 3.1.3 equivalent to $\Lambda \succeq 0$. Therefore, solving the degree d SoS relaxation can be done via semidefinite programming, see for e.g. [181]. In general, for degree-d SoS, we can solve it in $n^{O(d)}$ time[1]. Therefore, constant degree SoS can be solved in polynomial time.

Analyzing degree 2 SoS for maximum clique

To illustrate the use of this technique, let's analyze the degree 2 SoS relaxation for the maximum clique problem on Erdős-Rényi random graphs $G_{n,1/2}$. We use the program from Example 3.1.2.

Let A be the adjacency matrix of a graph G sampled from $G_{n,1/2}$ and let J be the matrix with all 1s. Then, with high probability over the choice of G, from random matrix theory, we have $\lambda_{max}(A - J/2) = O(\sqrt{n})$ where $\lambda_{max}(.)$ denotes the maximum singular value. Now, suppose a set S of vertices form a clique and let $\mathbf{1}_S$ denote the indicator vector of the set S, then

$$\begin{aligned}\frac{k(k-1)}{2} &= \langle \mathbf{1}_S, (A - J/2)\mathbf{1}_S \rangle \\ &\leq \|\mathbf{1}_S\|^2 \cdot \lambda_{max}(A - J/2) \\ &\leq k \cdot O(\sqrt{n})\end{aligned}$$

1. This is not completely accurate due to issues of bit complexity [137] but this doesn't occur for most problems of interest [155]

which shows $k \leq O(\sqrt{n})$.

The crux of this simple argument is that this is a *low-degree proof*, more specifically degree 2 proof, that SoS can capture. That is, if we solve the degree 2 SoS relaxation, we will be able to show that $\widetilde{\mathbb{E}}[\sum x_i] = O(\sqrt{n})$ whp.

To see this formally, we start with the following inequality: $O(\sqrt{n})I - (A - J/2) \succeq 0$ whp. This implies

$$x^\intercal(O(\sqrt{n})I - (A - J/2))x = \sum p_i(x)^2$$

is a sum of squares of polynomials of degree at most 1. A simple computation yields

$$x^\intercal(A - J/2)x = \frac{1}{2}(\sum_{i=1}^{n} x_i)^2 - \sum_{i,j} x_i x_j \mathbf{1}_{(i,j)\notin E(G)}$$

For our program variables x, we have $x_i^2 = x_i$ and $x_i x_j \mathbf{1}_{(i,j)\notin E(G)} = 0$. Therefore,

$$\sum p_i(x)^2 = O(\sqrt{n})(\sum_{i=1}^{n} x_i) - \frac{1}{2}(\sum_{i=1}^{n} x_i)^2$$

Apply $\widetilde{\mathbb{E}}$ both sides. We finally use the fact that for a polynomial $p(x)$, we have $\widetilde{\mathbb{E}}[p(x)^2] \geq \widetilde{\mathbb{E}}[p(x)]^2$, which is true because this rearranges to $\widetilde{\mathbb{E}}[(p(x) - \widetilde{\mathbb{E}}[p(x)])^2] \geq 0$, which is true because the left hand side is the the pseudo-expectation of a square polynomial, which is nonnegative by definition. This simple fact is essentially saying that the pseudo-variance is

nonnegative. Using the linearity of $\widetilde{\mathbb{E}}$, we finally get

$$\begin{aligned} O(\sqrt{n})\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i] - \frac{1}{2}(\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i])^2 &\geq O(\sqrt{n})\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i] - \frac{1}{2}\widetilde{\mathbb{E}}[(\sum_{i=1}^{n} x_i)^2] \\ &= \widetilde{\mathbb{E}}[O(\sqrt{n})(\sum_{i=1}^{n} x_i) - \frac{1}{2}(\sum_{i=1}^{n} x_i)^2] \\ &= \widetilde{\mathbb{E}}[\sum p_i(x)^2] \\ &= \sum \widetilde{\mathbb{E}}[p_i(x)^2] \\ &\geq 0 \end{aligned}$$

Therefore, $\widetilde{\mathbb{E}}[\sum_{i=1}^{n} x_i] = O(\sqrt{n})$ like we wanted to show.

This shows that the degree 2 SoS relaxation certifies an upper bound of $O(\sqrt{n})$ whp on the size of the maximum clique of an Erdős-Rényi random graph. In contrast, the size of the true maximum clique is $(2 + o(1)) \log n$ [125]. And despite intense effort, polynomial time algorithms can only detect a planted k-clique when $k = \Omega(\sqrt{n})$. Therefore, SoS already achieves the best known guarantees for this problem upto constant factors. It was shown in [13] that higher degree SoS (upto degree $O(\log n)$) doesn't necessarily do much better, which is a SoS lower bound of the type we will study in this work.

Alternate viewpoints of SoS

In the polynomial optimization problem of maximizing $p(x)$ subject to the constraints $g_1(x) = 0, \ldots, g_m(x) = 0$, if there does not exist any degree d pseudo-expectation operator $\widetilde{\mathbb{E}}$ satisfying $g_1 = 0, \ldots, g_m = 0$ such that $\widetilde{\mathbb{E}}[p] > c$, then we say that degree d SoS certifies that $\widetilde{\mathbb{E}}[p(x)] \leq c$.

A degree d SoS proof that $p(x) \leq c$ given $g_1(x) = 0, \ldots, g_m(x) = 0$ is an expression of

the form

$$-1 = \sum_{i \leq m} g_i(x) q_i(x) + \sum_{i \leq a} s_i(x)^2 + (p(x) - c) \sum_{i \leq b} t_i(x)^2$$

where $q_1, \ldots, q_m, s_1, \ldots, s_a, t_1, \ldots, t_b$ are polynomials in x such that each term on the right hand side of the above expression has degree at most d. Indeed, the existence of such an expression automatically implies that $p(x) \leq c$ whenever $g_1(x) = 0, \ldots, g_m(x) = 0$.

When degree d SoS certifies that $\widetilde{\mathbb{E}}[p(x)] \leq c$, by duality, this will imply that there exists a degree d SoS proof that $p(x) \leq c$ given $g_1(x), \ldots, g_m(x) = 0$. The Positivstellensatz of Krivine and Stengle [112, 174] says that for any c, either there exists x such that $p(x) > c, g_1(x) = 0, \ldots, g_m(x) = 0$, or there is an SoS proof that $p(x) \leq c$ given $g_1(x) = 0, \ldots, g_m(x) = 0$.

For a fixed d, degree d SoS can indeed be construed as finding the best c so that there is a degree d SoS proof of $\widetilde{\mathbb{E}}[p(x)] \leq c$. This also intuitively explains why higher degree SoS gives tighter relaxations. For most programs stemming from combinatorial optimization problems, degree n SoS usually finds the optimal bound, where n is the number of variables. So, for instance, degree n SoS exactly outputs the size of the maximum clique of a graph. For efficient algorithms, we usually want constant degree SoS. For showing lower bounds, we would like to show for higher degrees. In this work, all our lower bounds are for degree n^{ε} SoS, which corresponds to subexponential time!

The viewpoint we have studied here is the dual view aka the search for simple proofs, which will suit our purposes. There is also the primal viewpoint where SoS can be viewed directly as a semi-definite programming relaxation of the program. This is sometimes useful for algorithm design.

Similar to the maximum clique application shown above, the SoS hierarchy has been shown formally to obtain the state-of-the art approximation guarantees for many fundamental problems both in the worst case and the average case setting. This includes constraint satisfaction problems [152], maximum cut [72], sparsest cut [6], tensor PCA [88], etc. There-

fore, it's natural to study the limits of SoS by studying SoS lower bounds.

Before we discuss SoS lower bounds, we introduce the framework of hypothesis testing problem in more detail, suited to our purposes.

3.2 Hypothesis testing

Let Ω be a sample space. Let ν, μ be probability distributions on Ω^n. The hypothesis testing problem is the problem of distinguishing ν, μ given access to a sample. Formally, input $x \sim \Omega^n$ is sampled from either

- H_0: $x \sim \mu$
- H_1: $x \sim \nu$.

Our objective is to determine which distribution it came from, with high probability. This is the hypothesis testing problem in general, where traditionally, H_0 is known as the null hypothesis and H_1 the alternate hypothesis. We abuse notation and use H_0, H_1 to also denote the probability distributions μ, ν respectively as well.

For example, H_0 could be the distribution of Erdős-Rényi random graphs and H_1 could be the distribution of Erdős-Rényi random graphs with a large planted clique. Given the graph, we would like to determine which of the two distributions it came from, or in other words, whether it contains a large clique.

A hypothesis test f is a function $f : \Omega^n \to \{0, 1\}$. Given the input x, if $f(x) = 0$, then we report that x came from the null distribution H_0 otherwise we report that x came from the alternate distribution H_1.

A successful hypothesis test is a test f such that when b is chosen uniformly at random from $\{0, 1\}$ and x is sampled from H_b, we have $\mathbb{E}_b \Pr_{x \sim H_b}[f(x) \neq b] \leq o(1)$. That is, test f has success probability $1 - o(1)$. Here, for simplicity, we don't distinguish type 1 and type 2 errors.

Indeed, for a test to be useful, it should be computable efficiently. But when computational efficiency is disregarded, the famous Neyman-Pearson lemma precisely characterizes the best hypothesis test. To define this test, we need the following standard definition.

Definition 3.2.1 (Likelihood ratio)**.** *For a given hypothesis testing problem, define the likelihood ratio of an input x to be $LR(x) = \frac{\Pr_{H_1}(x)}{\Pr_{H_0}(x)}$.*

Lemma 3.2.2 (Neyman-Pearson Lemma)**.** *For a given hypothesis testing problem, the test f that minimizes $\mathbb{E}_b \Pr_{x\sim H_b}[f(x) \neq b]$ is the likelihood ratio test*

$$f(x) = \begin{cases} 1 & \text{if } LR(x) > 1 \\ 0 & o.w. \end{cases}$$

In this work, our focus will be on efficiently computable tests f.

3.2.1 Low degree likelihood ratio

Given a hypothesis testing problem. We focus on a special class of efficiently computable hypothesis tests involving low degree multivariate polynomials. These are termed low-degree distinguishers. We give a brief treatment in this section and refer the readers to [91, 114] for a more detailed treatment.

In this section, for polynomials to be well-defined, assume $\Omega \subseteq \mathbb{R}$. Moreover, assume H_0 has finite moments. We will consider distinguishers that arise from multivariate polynomials $f : \mathbb{R}^n \to \mathbb{R}$. We say that the distinguisher has degree D if the degree of f is at most D. Since the output of a polynomial need not be boolean, we need an alternate definition of the success of this distinguisher. We use the following definition from [91].

Definition 3.2.3 (Degree D distinguisher)**.** *For a hypothesis testing problem, the multivariate polynomial f is a successful degree D distinguisher if*

- *(Low degree) f is a multivariate polynomial of degree at most D.*

- *(Normalization)* $\mathbb{E}_{x\sim H_0}[f(x)] = 0, \mathbb{E}_{x\sim H_0}[f(x)^2] = 1$
- *(Distinguishability)* $\lim_{n\to\infty} \mathbb{E}_{x\sim H_1}[f(x)] \to \infty$.

The normalization ensures appropriate scaling for the polynomial. Note that the normalization is over the null distribution. Informally, normalized f is a successful distinguisher if it attains unbounded values on the alternate distribution in the limit. Indeed, in applications, a hypothesis test may be obtained by appropriately thresholding on the value of the polynomial.

The limit on the degree imposes the kind of computational restrictions we wish to impose on our distinguishing algorithm. It's an active area of research trying to understand the power of such low-degree distinguishers for hypothesis testing problems. For instance, we could ask: If degree $O(\log n)$ distinguishers fail for a hypothesis testing problem with input size $n^{O(1)}$, is the problem hard for all polynomial time algorithms?

The first natural question is to ask what's the best degree D distinguisher for a given hypothesis testing problem. This has been answered in prior works and is simply the projection of the likelihood ratio $LR(x) = \frac{\Pr_{H_1}(x)}{\Pr_{H_0}(x)}$ to degree D polynomials.

To make this precise, for $f, g : \mathbb{R}^n \to \mathbb{R}$, define the inner product $\langle f, g\rangle = \mathbb{E}_{x\sim H_0} f(x)g(x)$. Then, we can canonically define the projection $f^{\leq D}$ of a function f to degree D polynomials via this inner product. Take an orthonormal basis $\chi_0 = 1, \chi_1, \ldots, \chi_t$ of multivariate polynomials of degree at most D where $\chi_0 = 1$ is the constant function. Then, $f^{\leq D}(x) = \sum_{i\leq t}\langle f, \chi_t\rangle\chi_t(x)$.

The following lemma is implicit in prior works (e.g. [90, 83]). We include a proof for completeness.

Lemma 3.2.4. *For a hypothesis testing problem, the optimal degree D test f that maximizes $\mathbb{E}_{x\sim H_1} f(x)$ is the normalized low-degree likelihood ratio $\frac{LR^{\leq D}-1}{\|LR^{\leq D}-1\|}$. And it's value is $\mathbb{E}_{x\sim H_1}[f(x)] = \left\|LR^{\leq D} - 1\right\|$.*

Proof. Let f be a normalized degree D polynomial with $f = \sum_{i=0}^{t} c_t \chi_t$. Then, $c_0 = \mathbb{E}[f] = 0$ and $\sum c_i^2 = \mathbb{E}[f^2] = 1$. Then,

$$\mathop{\mathbb{E}}_{x\sim H_1} f(x) = \sum_{1\le i\le t} c_i \mathop{\mathbb{E}}_{x\sim H_1} \chi_i \le \sqrt{(\sum_{1\le i\le t} c_i^2)(\sum_{1\le i\le t} (\mathop{\mathbb{E}}_{x\sim H_1} \chi_i)^2)} = \sqrt{\sum_{1\le i\le t} (\mathop{\mathbb{E}}_{x\sim H_1} \chi_i)^2}$$

On the other hand, equality is attained by the polynomial $g = \frac{LR^{\le D}-1}{\|LR^{\le D}-1\|}$. Indeed, we have $\mathbb{E}_{x\sim H_0}[g] = 0$ because $\mathbb{E}_{x\sim H_0}[LR^{\le D}(x)] = \mathbb{E}_{x\sim H_0}[LR(x)] = 1$ and trivially, $\mathbb{E}_{x\sim H_0}[g(x)^2] = 1$ since we scaled by the norm. Finally,

$$\mathop{\mathbb{E}}_{x\sim H_1} g(x) = \frac{1}{\left\|LR^{\le D} - 1\right\|} \sum_{1\le i\le t} \langle LR(x), \chi_i\rangle^2$$

We complete the proof by observing that $\langle LR(x), \chi_i\rangle = \mathbb{E}_{x\sim H_0}[LR(x)\chi_i(x)] = \mathbb{E}_{x\sim H_1}[\chi_i(x)]$. Computing the value is straightforward. ∎

The low-degree likelihood ratio hypothesis [85, 91, 116] hypothesizes that if H_0, H_1 are *sufficiently nice* distributions, then there is a successful hypothesis test with running itme $n^{O(D)}$ if and only if there exists a successful degree D distinguisher. In particular, based on the above discussion, if $\left\|LR^{\le D} - 1\right\| = O(1)$, then we expect that there is no $n^{O(D)}$ time successful hypothesis test.

A main contribution of this work is to provide strong evidence that this conjecture is true for many fundamental problems, by exhibiting strong SoS lower bounds. To see this connection a bit more formally, we will introduce pseudo-calibration and connect it with low-degree distinguishers.

3.3 Pseudo-calibration

Given an optimization problem we are trying to show SoS lower bounds for. To obtain SoS integrality gaps on random instances, we need to construct valid pseudo-expectation values

for a random input instance of the problem. Naturally, these pseudo-expectation values will depend on the input.

Psuedo-calibration is a heuristic introduced by [10] to construct such candidate pseudo-expectation values almost mechanically by considering a planted distribution supported on instances of the problem with large objective value and using this planted distribution as a guide to construct the pseudo-expectation values. Note here that, for historic reasons, we use the term random distribution instead of null distribution and the term planted distribution instead of alternative distribution.

Unfortunately, psuedo-calibration doesn't guarantee feasibility of these candidate pseudo-expectation values and the corresponding moment matrix and this has to be verified separately for different problems. This verification of feasibility is relatively easy except for the PSDness condition. This is where the main contribution of this work lies, where we analyze the behavior of the constructed random moment matrix.

Indeed for our applications, psuedocalibration is used to obtain a candidate pseudoexpectation operator $\widetilde{\mathbb{E}}$ and a corresponding moment matrix Λ from the random vs planted problem. This will be the starting point for all our applications. Pseudo-calibration gives lower bounds for many problems, such as the ones considered in the works [74, 164, 107, 41, 129], making it an intriguing but poorly understood technique.

Here, we do not attempt to motivate and describe pseudo-calibration in great detail. Instead, we will briefly describe the heuristic, the intuition behind it and show an example of how to use it. A detailed treatment can be found in [10].

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

in the planted and random distributions. That is,

$$\mathop{\mathbb{E}}_{v\sim\nu}[\widetilde{\mathbb{E}}[f(x)]g(v)] = \mathop{\mathbb{E}}_{(x,v)\sim\mu}[f(x)g(v)]$$

Here, the notation $(x, v) \sim \mu$ means that in the planted distribution μ, the input is v and x denotes the planted structure in that instance. For example, in planted clique, x would be the indicator vector of the clique. If there are multiple, pick an arbitrary one.

Let $\mathcal{F}$ denote the Fourier basis of polynomials for the input v. By choosing different basis functions from $\mathcal{F}$ as choices for g such that the degree is at most some truncation parameter D, we get all lower order Fourier coefficients for $\widetilde{\mathbb{E}}[f(x)]$ when considered as a function of v. Furthermore, the higher order coefficients are set to be 0 so that the candidate pseudoexpectation operator can be written as

$$\widetilde{\mathbb{E}}f(x) = \sum_{\substack{g\in\mathcal{F}\\ deg(g)\leq n^{\varepsilon}}} \mathop{\mathbb{E}}_{v\sim\nu}[\widetilde{\mathbb{E}}[f(x)]g(v)]g(v) = \sum_{\substack{g\in\mathcal{F}\\ deg(g)\leq n^{\varepsilon}}} \mathop{\mathbb{E}}_{(x,v)\sim\mu}[[f(x)]g(v)]g(v)$$

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

One intuition for pseudo-calibration is as follows. The planted distribution is usually chosen to be a maximum entropy distribution which still has the planted structure. This conforms to the philosophy that random instances are hard for SoS, such as the uniform Bernoulli distribution for planted clique or the Gaussian distribution for Tensor PCA. By conditioning on the lower order moments matching such a planted distribution, pseudo-calibration can be interpreted as sort of interpolating between the random and planted distributions by only looking at lower order Fourier characters. This intuition has proven to be successful, since pseudo-calibration been successfully exploited to construct SoS lower bounds for a wide variety of dense as well as sparse problems.

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

The condition $\widetilde{\mathbb{E}}[1] = 1$ is not automatically satisfied but in most applications, we usually require that $\widetilde{\mathbb{E}}[1] = 1 \pm o(1)$. 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 $\widetilde{\mathbb{E}}'$ defined as $\widetilde{\mathbb{E}}'[f(x)] = \widetilde{\mathbb{E}}[f(x)]/\widetilde{\mathbb{E}}[1]$.

We remark that the condition $\widetilde{\mathbb{E}}[1] = 1 \pm o(1)$ has been quite successful in predicting the right thresholds between approximability and inapproximability[85, 91, 116]. This will be crucial when we connect pseudo-calibration to low degree distinguishers.

Example: Planted Clique As an warmup, we review the pseudo-calibration calculation for planted clique. Here, the random distribution ν is $G(n, \frac{1}{2})$.

The planted distribution μ 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 $G \in \{-1, 1\}^{\binom{[n]}{2}}$ 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.

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

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

Using standard Fourier analysis, this implies that if we take

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

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

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

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

Here, we only set the Fourier coefficients for small E and set the other larger Fourier coefficients to 0. Usually, the choice of the truncation parameter is problem specific but there are some basic requirements [85]. We now outline our general strategy to show SoS lower bounds. We employ this in all our results.

3.3.1 Strategy to show SoS lower bounds

In this work, the general strategy to show SoS lower bounds can be summarized as follows.

- Given a random distribution, identify a suitable planted distribution
- Pseudocalibrate with respect the two distributions and obtain a candidate pseudoexpectation operator
- Show that the moment matrix satisfies the constraints

The most technically challenging part of this approach usually is to show that the moment matrix is positive semidefinite. Much of our contributions lies in this step, where we analyze the behavior of the random moment matrix thus obtained. Now, we connect pseudo-calibration to low-degree distinguishers.

3.3.2 Connection to Low-degree distinguishers

We are ready to connect psuedo-calibration to low-degree tests. Recall that in pseudo-calibration, we set the higher order Fourier coefficients to 0. This is known as truncation. In particular, we truncate so that the resulting pseudoexpectation has degree at most D in the input. By construction, $\mathbb{E}[\widetilde{\mathbb{E}}[1]] = 1$ and we would like to understand how much $\widetilde{\mathbb{E}}[1]$ deviates from 1. The following lemma says that the variance of $\widetilde{\mathbb{E}}[1]$ behaves like the squared value of the optimal degree-D distinguisher.

Lemma 3.3.1. *The pseudo-calibrated pseudo-expectation $\widetilde{\mathbb{E}}$, truncated to degree D, satisfies*

$$\mathrm{var}(\widetilde{\mathbb{E}}[1]) = \left\| LR^{\leq D} - 1 \right\|^2$$

Proof. Pseudocalibration sets $\mathbb{E}_{x\sim H_0}[\widetilde{\mathbb{E}}[1]\chi_i] = \mathbb{E}_{x\sim H_1}[\chi_i]$ for all $i \leq t$. Therefore, $\widetilde{\mathbb{E}}[1] = 1 + \sum_{1\leq i\leq t} \mathbb{E}_{x\sim H_1}[\chi_i]\chi_i$ giving $\mathrm{var}(\widetilde{\mathbb{E}}[1]) = \sum_{1\leq i\leq t}(\mathbb{E}_{x\sim H_1}\chi_i)^2 = \left\| LR^{\leq D} - 1 \right\|^2$. ∎

One of the essential steps in our SoS lower bound proofs is to verify, after pseudo-calibration, that $\widetilde{\mathbb{E}}[1]$ is well-behaved. In particular, for strong SoS lower bounds, we expect $\widetilde{\mathbb{E}}[1] = 1 + o(1)$. Although this is not formally necessary, it has often been the case in our applications and we expect it to be necessary for obtaining strong SoS lower bounds via this approach.

But when this is indeed the case and we exhibit SoS lower bounds, note that this is already strong evidence towards the low-degree likelihood ratio hypothesis. In more detail, because of Lemma 3.2.4 and Lemma 3.3.1, the best degree D distinguisher does not distinguish the

two distributions μ, ν. And our lower bounds affirm that the powerful SoS hierarchy cannot distinguish the two distributions as well, which is an important step towards the general hypothesis.

It's a fundamentally important open problem in this field to prove that for sufficiently nice distributions μ, ν, after pseudo-calibrating, $\widetilde{\mathbb{E}}[1] = 1 + o(1)$ implies the existence of strong SoS lower bounds.

CHAPTER 4

OUR MAIN RESULTS ON SUM-OF-SQUARES LOWER BOUNDS

In this chapter, we state formally the main Sum of Squares lower bounds that we prove in this thesis and put them in the context of prior works. The material in this chapter is adapted from [70, 149], where the results originally appeared.

4.1 The Sherrington-Kirkpatrick Hamiltonian

We first define the Gaussian Orthogonal Ensemble, GOE(n), a random matrix model for $n \times n$ matrices.

Definition 4.1.1. *The Gaussian Orthogonal Ensemble, denoted* GOE(n)*, is the distribution of* $\frac{1}{\sqrt{2}}(A + A^\intercal)$ *where* A *is a random* $n \times n$ *matrix with i.i.d. standard Gaussian entries.*

Equivalently, we could define GOE(n) to be a probability distribution over symmetric matrices W such that $W_{ii} \sim \mathcal{N}(0, 2)$ for $i \leq n$ and for $i \neq j$, $W_{ij} = W_{ji} \sim \mathcal{N}(0, 1)$ independently.

We consider the main optimization task

$$\mathrm{OPT}(W) := \max_{x \in \{\pm 1\}^n} x^\intercal W x, \tag{4.1}$$

where W is a random symmetric matrix in $\mathbb{R}^{n \times n}$. This is an important task that arises in computer science and statistical physics.

In computer science, a natural choice of W is to take it to be the Laplacian of a graph [80, Section 4]. Then, the problem is equivalent to the Maximum Cut problem, a well-known NP-hard problem in the worst case [98]. The equivalence is immediate by observing that $x \in \{\pm 1\}^n$ can be thought of as encoding a bipartition of $[n] = \{1, 2, \ldots, n\}$.

In particular, an interesting special case is when we consider sparse random graphs, sampled either from the Erdős-Rényi graphs $G(n, \frac{d}{n})$ with average degree d or a uniformly chosen d-regular graph, where $d \geq 3$ is a fixed integer. In this case, it is known that the true size of the maximum cut is asymptotically $n(\frac{d}{4} + f(d)\sqrt{d})$. Moreover, it was shown in [48] (originally conjectured in [187]) that $\lim_{d\to\infty} f(d) = \frac{1}{2}P^* \approx 0.382$, where

$$P^* := \frac{1}{2} \lim_{n\to\infty} \mathbb{E}_{W\sim \mathrm{GOE}(n)}[\frac{1}{n^{3/2}} \mathrm{OPT}(W)] \approx 0.7632$$

is referred to as the Parisi constant. This already strongly motivates the problem of studying Eq. (4.1) when $W \sim \mathrm{GOE}(n)$. But this problem is motivated for another fantastic reason.

In statistical physics, when $W \sim \mathrm{GOE}(n)$, our objective, upto scaling, is the Hamiltonian of the famous Sherrington-Kirkpatrick model. Here, x can be thought of as encoding spin values in a spin-glass model. $-W_{i,j}$ models the interaction between spin x_i and x_j (with $-W_{i,j} \geq 0$ being ferromagnetic and $-W_{i,j} < 0$ being anti-ferromagnetic). Then, the optimal value corresponds to the minimum-energy, or ground state of the system, upto sign. The works [142, 143, 45] predicted, using non-rigorous means, that $P^* \approx 0.7632$. This was eventually formalized in the works [176, 139, 75].

In this work, we will focus on this average case optimization problem when $W \sim \mathrm{GOE}(n)$. The first natural question is whether there exists a polynomial-time algorithm that given $W \sim \mathrm{GOE}(n)$ computes an x achieving close to $\mathrm{OPT}(W)$?" In a recent breakthrough work, Montanari [132] showed that, for any $\varepsilon > 0$, there exists a polynomial time algorithm that with high probability achieves a value of $(2P^* - \varepsilon)n^{3/2}$, assuming a widely believed conjecture.

Now we move onto certification: Is there an efficient algorithm to certify an upper bound on $\mathrm{OPT}(W)$ for any input W?

A simple algorithm will be the spectral algorithm where we just output the largest eigenvalue of W, upto scaling, for an upper bound. Note that $\mathrm{GOE}(n)$ is a particular kind of Wigner matrix ensemble, thereby satisfying the semicircle law, which in this case establishes

that the largest eigenvalue of W is $(2+\mathrm{o}_n(1))\cdot\sqrt{n}$ with probability $1-\mathrm{o}_n(1)$. Thus, a trivial spectral bound establishes $\mathrm{OPT}(W) \le (2+\mathrm{o}_n(1))\cdot n^{3/2}$ with probability $1-\mathrm{o}_n(1)$.

Now, we can ask if it's possible to beat this spectral algorithm for certification. In particular, we can ask how well SoS does as a certification algorithm. The natural upper bound of $(2+\mathrm{o}_n(1))\cdot n^{3/2}$ obtained via the spectral norm of W is also the value of the degree-2 SoS relaxation [133]. Two independent recent works of Mohanty–Raghavendra–Xu [129] and Kunisky–Bandeira [115] show that degree-4 SoS does not perform much better, and a heuristic argument from [9] suggests that even degree-$(n/\log n)$ SoS cannot certify anything stronger than the trivial spectral bound. Thus we ask,

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

In this work, we answer the question above negatively by showing that even at degree as large as n^δ, SoS cannot improve upon the basic spectral algorithm.

Theorem 4.1.2. *There exists a constant $\delta > 0$ such that, w.h.p. for $W \sim \mathrm{GOE}(n)$, there is a degree-n^δ SoS solution for the Sherrington–Kirkpatrick problem with value at least $(2-\mathrm{o}_n(1))\cdot n^{3/2}$.*

An independent and concurrent work by Kunisky [113] also showed a special case of the above theorem for degree-6 SoS, using different techniques.

We will present the proof of this theorem in Chapter 5. The above theorem and it's proof originally appeared in [70], from where the material here is adapted from. We now present the high level ideas behind the proof of this theorem.

4.1.1 Our approach

In order to prove Theorem 4.1.2, we first introduce a new average-case problem we call Planted Affine Planes (PAP) for which we directly prove a SoS lower bound. We then use

the PAP lower bound to prove a lower bound on the Sherrington–Kirkpatrick problem. The PAP problem can be informally described as follows (see Definition 5.1.1 for the formal definition).

Definition 4.1.3 (Informal statement of PAP)**.** *Given m random vectors $d_1, \ldots, d_m$ in $\mathbb{R}^n$, can we prove that there is no vector $v \in \mathbb{R}^n$ such that for all $u \in [m]$, $\langle v, d_u \rangle^2 = 1$? In other words, can we prove that m random vectors are not all contained in two parallel hyperplanes at equal distance from the origin?*

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

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

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

Theorem 4.1.4. *For both the Gaussian and Boolean settings, there exists a constant $c > 0$ such that for all $\varepsilon > 0$ and $\delta \leq c\varepsilon$, for $m \leq n^{3/2-\varepsilon}$, w.h.p. there is a feasible degree-n^δ SoS solution for Planted Affine Planes.*

It turns out that the Planted Affine Plane problem introduced above is closely related to the following "Boolean vector in a random subspace" problem, which we call the Planted Boolean Vector problem, introduced by [129] in the context of studying the performance of SoS on computing the Sherrington–Kirkpatrick Hamiltonian.

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

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

where V is a uniformly random p-dimensional subspace[1] of $\mathbb{R}^n$, and Π_V is the projector onto V. In brief, the relationship to the Planted Affine Plane problem is that the PAP vector v represents the coefficients on a linear combination for the vector b in the span of a basis of V.

An argument of [129] shows that, when $p \ll n$, w.h.p., $\mathrm{OPT}(V) \approx \frac{2}{\pi}$, whereas they also show that w.h.p. assuming $p \geq n^{0.99}$, there is a degree-4 SoS solution with value $1 - \mathrm{o}_n(1)$. They ask whether or not there is a polynomial time algorithm that can certify a tighter bound; we rule out SoS-based algorithms for a larger regime both in terms of SoS degree and the dimension p of the random subspace.

Theorem 4.1.5. *There exists a constant $c > 0$ such that, for all $\varepsilon > 0$ and $\delta \leq c\varepsilon$, for $p \geq n^{2/3+\varepsilon}$, w.h.p. over V there is a degree-n^δ SoS solution for Planted Boolean Vector of value* 1.

The bulk of our technical contribution lies in the SoS lower bound for the Planted Affine Planes problem, Theorem 4.1.4. We then show that Planted Affine Planes in the Gaussian setting is equivalent to the Planted Boolean Vector problem. The reduction from Sherrington-Kirkpatrick to the Planted Boolean Vector problem is due to Mohanty–Raghavendra–Xu [129].

As a starting point to the PAP lower bound, we employ pseudocalibration to produce a good candidate SoS solution $\widetilde{\mathbb{E}}$. The operator $\widetilde{\mathbb{E}}$ unfortunately does not exactly satisfy the PAP constraints "$\langle v, d_u \rangle^2 = 1$", it only satisfies them up to a tiny error. In the original work,

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

we use an interesting and rather generic approach to round $\widetilde{\mathbb{E}}$ to a nearby pseudoexpectation operator $\widetilde{\mathbb{E}}'$ which does exactly satisfy the constraints, We have omitted this in this thesis for the sake of brevity, but it can be found in the original work [70].

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

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

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

Similarly to [10], we decompose $\mathcal{M}$ as a linear combination of graph matrices, i.e., $\mathcal{M} = \sum_\alpha \lambda_\alpha \cdot M_\alpha$, where M_α is the graph matrix associated with shape α. In brief, each graph matrix aggregates all terms with shape α in the Fourier expansions of the entries of $\mathcal{M}$ – the shape α is informally a graph with labeled edges with size bounded by poly(D). A graph matrix decomposition of $\mathcal{M}$ is particularly handy in the PSD analysis since the operator norm of individual graph matrices M_α is (with high probability) determined by simple combinatorial properties of the graph α. One technical difference from [10] is that our graph matrices have two types of vertices $\square$ and $\bigcirc$; these graph matrices fall into the general framework developed by Ahn et al. in [2].

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

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

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

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

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

4.1.2 Related work

Degree-4 SoS lower bounds on the Sherrington-Kirkpatrick Hamiltonian problem were proved independently by Mohanty–Raghavendra–Xu [129] and Kunisky–Bandeira [115]. The concurrent and independent work by Kunisky [113] obtained degree 6 SoS lower bounds. In this

work, we prove an improved degree-n^δ SoS lower bound for some constant $\delta > 0$. Our result is obtained by reducing the Sherrington-Kirkpatrick problem to the "Boolean Vector in a Random Subspace" problem which is equivalent to our new Planted Affine Planes problem on the normal distribution. The reduction from Sherrington-Kirkpatrick problem to the "Boolean Vector in a Random Subspace" is due to Mohanty–Raghavendra–Xu [129]. The results of Mohanty–Raghavendra–Xu [129] and Kunisky–Bandeira [115] build on a degree-2 SoS lower bounds of Montanari and Sen [133].

Degree-4 SoS lower bounds on the "Boolean Vector in a Random Subspace" problem for $p \geq n^{0.99}$ were proved by Mohanty–Raghavendra–Xu in [129] where this problem was introduced. We improve the dependence on p to $p \geq n^{2/3+\varepsilon}$ for any $\varepsilon > 0$ and obtain a stronger degree-$n^{c\varepsilon}$ SoS lower bound for some absolute constant $c > 0$.

Interestingly, the recent work [186] exhibited a polynomial-time algorithm for the search variant of Planted Affine Planes for $m \geq n+1$, achieving statistical optimality. In particular, they beat prior known polynomial time algorithms, including SoS based ones, all of which required $m \gg n^2$. Their algorithm is a lattice-based method that uses the specific algebraic structure present in this problem. It's not clear if this algorithm can be used for certification.

4.2 Planted Slightly Denser 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 [97] and has been subject to a long line of work over the years (e.g. [59, 58, 101, 22, 23, 30, 124]). In this work, we consider the following certification variant of planted dense subgraph.

> *Given a random graph G sampled from the Erdős-Rényi model $G(n, \frac{1}{2})$, certify an upper bound on the edge density of the densest subgraph on k vertices.*

To apply the strategy from Section 3.3.1, 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 6.2, we compute the candidate moment matrix Λ obtained via pseudo-calibration. Our main theorem is as follows.

Theorem 4.2.1. *Let $C_p > 0$. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $k \leq n^{\frac{1}{2}-\varepsilon}$ and $p = \frac{1}{2} + \frac{n^{-C_p\varepsilon}}{2}$, then with high probability, the candidate moment matrix Λ given by pseudo-calibraton for degree $n^{C\varepsilon}$ Sum-of-Squares is PSD.*

Corollary 4.2.2. *Let $C_p > 0$. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $k \leq n^{\frac{1}{2}-\varepsilon}$ and $p = \frac{1}{2} + \frac{n^{-C_p\varepsilon}}{2}$, 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 $\approx p$.*

4.2.1 Related work

For many different parameter regimes of the random and planted distributions (an example being planting $G_{k,q}$ in $G_{n,p}$ for constants $p < q$), and when $k = o(\sqrt{n})$, the hardness of the easier distinguishing version of planted dense subgraph problem has been posed as formal conjecture (often referred to as the PDS conjecture) before in the literature (see e.g., [77, 40, 33, 34]). This has also led to many reductions to other problems [31], although it's

not clear if these reductions can be made in the SoS framework without loss in the parameter dependence.

In our case, we consider the slightly planted denser subgraph version where for $k \leq n^{\frac{1}{2}-\varepsilon}$, we plant a subgraph of density $\frac{1}{2} + \frac{1}{n^{O(\varepsilon)}}$, i.e. $p = \frac{1}{2}, q = \frac{1}{2} + \frac{1}{n^{O(\varepsilon)}}$. This has been widely believed to require sub-exponential time. Our work provides strong evidence towards this by exhibiting unconditional lower bounds against the powerful SoS hierarchy, even if we consider $n^{O(\varepsilon)}$ levels, which corresponds to $n^{n^{O(\varepsilon)}}$ running time! We expect this to lead to this problem being used as a natural starting point for reductions to show sub-exponential time hardness for various problems.

Within the SoS literature, [10] show that for $k \leq n^{\frac{1}{2}-\varepsilon}$ for a constant $\varepsilon > 0$, the degree $o(\log n)$ Sum-of-Squares cannot distinguish between a fully random graph sampled from $G(n, \frac{1}{2})$ 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 $k \leq n^{\frac{1}{2}-\varepsilon}$.

In Corollary 4.2.2, we show that for $k \leq n^{\frac{1}{2}-\varepsilon}$ for a constant $\varepsilon > 0$, degree $n^{\Omega(\varepsilon)}$ SoS cannot certify an edge density better than $\frac{1}{2} + \frac{1}{n^{O(\varepsilon)}}$. The degree of SoS in our setting, $n^{\Omega(\varepsilon)}$ is vastly higher than the earlier known result which uses degree $o(\log n)$. To the best of our knowledge, this is the first result that proves such a high degree lower bound.

We remark that when we take $k = n^{\frac{1}{2}-\varepsilon}$, the true edge density of the densest k-subgraph is $\frac{1}{2} + \frac{\sqrt{\log(n/k)}}{\sqrt{k}} + \mathrm{o}(\frac{1}{\sqrt{k}}) \approx \frac{1}{2} + \frac{1}{n^{1/4-\varepsilon/2}}$ as was shown in [68, Corollary 2] whereas, by Corollary 4.2.2, the SoS optimum is as large as $\frac{1}{2} + \frac{1}{n^{\varepsilon}}$. This highlights a significant difference in the optimum value.

4.3 Tensor PCA

The Tensor Principal Component Analysis problem, originally proposed by [160], is a variant 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, λ 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. [66, 36, 29, 14, 15, 24]). Tensor PCA is an average case version of the above problem and has been studied before in the literature [160, 89, 25, 85]. 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.*

Let $k \geq 2$ be an integer. We apply the strategy from Section 3.3.1 using the following distributions.

- Random distribution: Sample A from $\mathcal{N}(0, I_{[n]^k})$.
- 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 $\mathcal{N}(0, I_{[n]^k})$. Set $A = B + \lambda u^{\otimes k}$.

In Section 6.3, we compute the candidate moment matrix Λ obtained by using pseudo-calibration on this planted distribution. Our main theorem is as follows.

Theorem 4.3.1. *Let $k \geq 2$ be an integer. There exist constants $C, C_\Delta > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $\lambda \leq n^{\frac{k}{4}-\varepsilon}$ and $\Delta = n^{-C_\Delta \varepsilon}$ then with high probability, the candidate moment matrix Λ given by pseudo-calibration for degree $n^{C\varepsilon}$ Sum-of-Squares is PSD.*

Corollary 4.3.2. *Let $k \geq 2$ be an integer. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $\lambda \leq n^{\frac{k}{4}-\varepsilon}$, then with high probability, degree $n^{C\varepsilon}$ Sum-of-Squares cannot certify that for a random tensor A from $\mathcal{N}(0, I_{[n]^k})$, there is no vector u such that $\|u\| \approx 1$ and $\langle A, \underbrace{x \otimes \ldots \otimes x}_{k \text{ times}} \rangle \approx \lambda$.*

4.3.1 Related work

In [25], it was shown that $q \leq n$ levels of SoS certifies an upper bound of $\frac{2^{O(k)}(n \cdot \text{polylog}(n))^{k/4}}{q^{k/4-1/2}}$ for the Tensor PCA problem. When $q = n^{\varepsilon}$ for sufficiently small ε, this gives an upper bound of $n^{\frac{k}{4}-O(\varepsilon)}$. Corollary 4.3.2 shows that this is tight.

In [85], they state a theorem similar to Corollary 4.3.2 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 the more natural setting where the distribution is $\mathcal{N}(0, 1)$.

When $k = 2$, the maximum value of $\langle x^{\otimes k}, A \rangle$ 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 [178]. In contrast, by Corollary 4.3.2, the optimum value of the degree n^{ε} SoS is as large as $n^{\frac{k}{4}-O(\varepsilon)}$. This exhibits an integrality gap of $n^{\frac{k}{4}-\frac{1}{2}-O(\varepsilon)}$.

4.4 Sparse PCA

The Wishart model of Sparse PCA, also known as the Spiked Covariance model, was originally proposed by [94]. In this problem, we observe m vectors $v_1, \ldots, v_m \in \mathbb{R}^d$ 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 λ 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. [183, 134, 123, 177, 42, 3]). It's known that vanilla PCA does not yield good estimators in high dimensional settings [7, 146, 94]. A large volume of work has gone into studying Sparse PCA and it's variants, both from an algorithmic perspective (e.g. [4, 122, 111, 50, 184]) as well as from an inapproximability perspective (e.g. [20, 121, 51, 85, 31]).

Given the decades of research on this problem and how fundamental it is for a multitude of applications and disciplines, understanding the computational threshold behavior of the Wishart model of Sparse PCA is an extremely important research topic in statistics. In particular, prior works have explored statistical query lower bounds, SDP lower bounds, lower bounds by reductions from widely believed conjectures, etc. On the other hand, there have only been two prior works on lower bounds against SoS, specifically only for degree 2 and degree 4 SoS, which can be attributed to the difficulty in proving such lower bounds. In this paper, we vastly strengthen these lower bounds and show almost-tight lower bounds for the SoS hierarchy of degree d^{ε} which corresponds to a running time of $d^{d^{O(\varepsilon)}}$.

To apply the strategy from Section 3.3.1, 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 $\{-\frac{1}{\sqrt{k}}, 0, \frac{1}{\sqrt{k}}\}^d$ where the values are taken with probabilites $\frac{k}{2d}, 1 - \frac{k}{d}, \frac{k}{2d}$ respectively. Then sample $v_1, \ldots, v_m$ as follows. For each $i \in [m]$, with probability Δ, 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 $v_1, \ldots, v_m$.

In Section 6.4, we compute the candidate moment matrix Λ obtained by using pseudo-calibration on this planted distribution. We now state our main theorem.

Theorem 4.4.1. *There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}, m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, and there exists a constant A such that $0 < A < \frac{1}{4}$,*

$d^{4A} \leq k \leq d^{1-A\varepsilon}$, *and* $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$, *then with high probability, the candidate moment matrix* Λ *given by pseudo-calibration for degree* $d^{C\varepsilon}$ *Sum-of-Squares is PSD.*

Corollary 4.4.2. *There exists a constant* $C > 0$ *such that for all sufficiently small constants* $\varepsilon > 0$, *if* $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}, m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, *and there exists a constant* A *such that* $0 < A < \frac{1}{4}$, $d^{4A} \leq k \leq d^{1-A\varepsilon}$, *and* $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$, *then with high probability, the degree* $d^{C\varepsilon}$ *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* $\approx k$ *nonzero entries,* $\|u\| \approx 1$, *and* $\|Su\|^2 \approx m + m\Delta\lambda$.

4.4.1 Related work

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. In Fig. 4.1 and Fig. 4.2, we assign the different parameter regimes into the following categories.

- Diagonal thresholding: In this regime, Diagonal thresholding [94, 4] recovers the sparse vector. Covariance thresholding [111, 50] and SoS [54] can also be used in this regime. Covariance thresholding has better dependence on logarithmic factors and SoS works in the presence of adversarial errors.

- Vanilla PCA: Vanilla PCA can recover the vector, i.e. we do not need to use the fact that the vector is sparse (see e.g. [21, 54]).

- Spectral: An efficient spectral algorithm recovers the sparse vector (see e.g. [54]).

- Spectral*: A simple spectral algorithm distinguishes the planted distribution from the random distribution but it is information theoretically impossible to recover the sparse vector [54, Appendix E].

- Hard: A regime where it is conjectured to be hard to distinguish between the random and the planted distributions. We discuss this in more detail below.

In Fig. 4.1 and Fig. 4.2, the regimes corresponding to Diagonal thresholding, Vanilla PCA and Spectral are dark green, while the regimes corresponding to Spectral* and Hard are light green and red respectively. The hard regime is the one studied in this work.

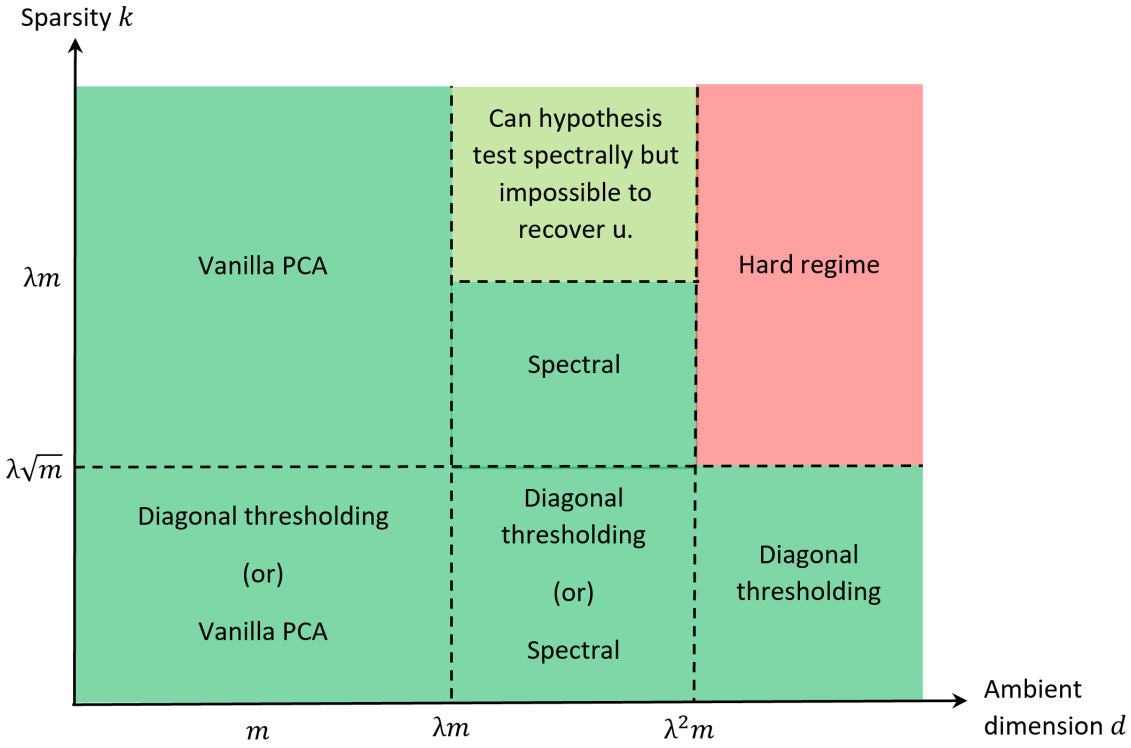

Figure 4.1: The computational barrier diagram when $\lambda \geq 1$

In the *Hard* parameter regime where $m \ll \frac{k^2}{\lambda^2}$ and $m \ll \frac{d}{\lambda^2}$, degree 2 and degree 4 SoS lower bounds have been shown in prior works, while we handle degree $d^{O(\varepsilon)}$. In particular, the works [111, 21] obtain degree 2 SoS lower bounds. [121] obtain degree 4 SoS lower bounds using an ad-hoc construction. It's not clear if their construction can be generalized for higher degrees. Moreover, the bounds they obtain are tight up to polylogarithmic factors when λ is a constant but are not tight when λ is not a constant, so we improve their bounds even in the degree 4 case. We subsume all these earlier known results in this work with Corollary 4.4.2. This is a vast improvement over prior known sum of squares lower bounds and provides compelling evidence for the hardness of Sparse PCA in this parameter range.

The work [85] considers the related but qualitatively different Wigner model of Sparse

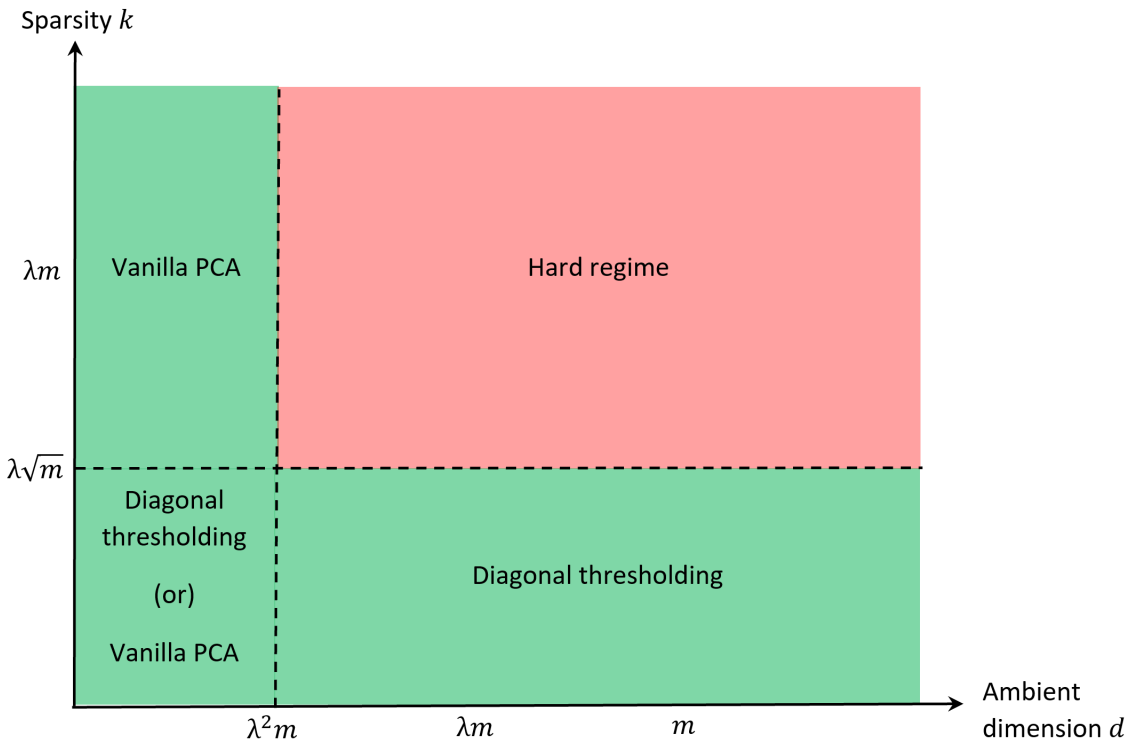

Figure 4.2: The computational barrier diagram when $\lambda < 1$

PCA and they state degree d^{ε} SoS lower bounds, without explicitly proving these bounds. The techniques in that work do not recover our results because the matrix formed by the random samples in the Wishart model is asymmetric, and handling it correctly is far from being a mere technicality. On the other hand, the machinery can recover the results on the Wigner model as well, though we only analyze the Wishart model in this paper.

In [54], 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^{ε} polynomials cannot distinguish the random and planted distributions. Corollary 4.4.2 says that under mildly stronger assumptions, degree n^{ε} 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.

There have also been direct reductions from planted clique to Sparse PCA [31], and it's natural to ask if these reductions can obtain SoS lower bounds on Sparse PCA from the

known SoS lower bounds on planted clique [10]. To the best of our knowledge, no such reduction is known and constructing such a reduction would be challenging as it would have to be captured by SoS and avoid losing too much in the parameters. Still, it may well be possible to construct such a reduction.

4.5 Our approach

Theorem 4.2.1, Theorem 4.3.1 and Theorem 4.4.1 all essentially boil down to showing that a large moment matrix Λ is PSD. All three results are obtained via applications of one main theorem, which we call the machinery. In this work, we state and use the machinery, whose proof can be found in the original work where it appeared [149]. To show PSDness, the machinery constructs certain *coefficient matrices* from Λ and gives conditions on these coefficient matrices which are sufficient to guarantee that Λ is PSD with high probability. In this section, we give an informal sketch of the machinery and how it generalizes the techniques used to prove the SoS lower bound for planted clique [10]. We also motivate some of the conditions that arise in the machinery.

Shapes and graph matrices

Before we can describe how the machinery works, we need to describe shapes and graph matrices, which were originally introduced by [10, 126] and later generalized in [2]. Both the planted clique analysis and our analysis use shapes and graph matrices.

Shapes α are graphs that contain extra information about the vertices. Corresponding to each shape α, there is a matrix-valued function (i.e. a matrix whose entries depend on the input) M_α that we call a graph matrix. Graph matrices are analogous to a Fourier basis, but for matrix-valued functions that exhibit a certain kind of symmetry. In our setting, Λ will be such a matrix-valued function, so we can decompose Λ as a linear combination of graph matrices.

Shapes and graph matrices have several properties which make them very useful to work with. First, $\|M_\alpha\|$ can be bounded with high probability in terms of simple combinatorial properties of the shape α. Second, if two shapes α and β match up in a certain way, we can combine them to form a larger shape $\alpha \circ \beta$. We call this operation shape composition. Third, each shape α has a canonical decomposition into three shapes, the left, middle and right parts of α, which we call σ, τ, and σ'^T. For this canonical decomposition, we have that $\alpha = \sigma \circ \tau \circ \sigma'^T$ and $M_\alpha \approx M_\sigma M_\tau M_{\sigma'^T}$ [2]. This decomposition turns out to be crucial for both the planted clique analysis and our analysis.

Summary of the SoS lower bound for planted clique and the machinery

We now give a brief summary of the techniques for the SoS lower bound for planted clique and for the machinery.

For planted clique, the SoS lower bound analysis works as follows

1. Using the technique of pseudo-calibration, construct a candidate moment matrix Λ.

2. Decompose the moment matrix Λ as a linear combination $\Lambda = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$ of graph matrices M_α.

3. For each shape α, decompose α into a left part σ, a middle part τ, and a right part σ'^T. We then have that $M_\alpha \approx M_\sigma M_\tau M_{\sigma'^T}$.

4. Using the approximate decompositions $M_\alpha \approx M_\sigma M_\tau M_{\sigma'^T}$, give an approximate decomposition $\Lambda \approx LQL^T$ of M where $Q \succeq 0$ with high probability.

5. Show that with high probability, $\Lambda = LQL^T - (LQL^T - M) \succeq 0$ by carefully analyzing the difference $LQL^T - M$ using similar techniques.

2. Actually, due to a technical issue related to automorphism groups, this equation is off by a multiplicative constant. For details, see [149].

The machinery uses a similar framework. The key innovation of the machinery is that it introduces coefficient matrices (step 4) and carry out the analysis in terms of these coefficient matrices.

1. Construct a candidate moment matrix Λ. This can be done either using pseudo-calibration or in a more ad-hoc manner.

2. Decompose the moment matrix Λ as a linear combination $\Lambda = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$ of graph matrices M_α.

3. For each shape α, decompose α into a left part σ, a middle part τ, and a right part σ'^T.

4. Based on the coefficients λ_α and the decompositions of the shapes α into left, middle, and right parts, construct coefficient matrices H_{Id_U} and H_τ.

5. Based on the coefficient matrices H_{Id_U} and H_τ, obtain an approximate PSD decomposition of Λ.

6. Show that the error terms (which we call intersection terms) can be bounded by the approximate PSD decomposition of Λ.

We show that this analysis will succeed as long as three conditions on the coefficient matrices are satisfied. Thus, in order to use the machinery to prove sum of squares lower bounds, it is sufficient to do the following.

1. Construct a candidate moment matrix Λ.

2. Decompose the moment matrix Λ as a linear combination $\Lambda = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$ of graph matrices M_α (akin to Fourier decomposition) and find the corresponding coefficient matrices.

3. Verify the required conditions on the coefficient matrices.

A sketch of the intuition behind the machinery conditions

Giving an approximate PSD factorization As discussed above, we decompose the moment matrix Λ as a linear combination $\Lambda = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$ of graph matrices M_α. We then decompose each α into left, middle, and right parts σ, τ, and ${\sigma'}^T$. We now have that

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

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

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

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

which shows that the contribution from these terms is positive semidefinite. In fact, this turns out to be the case for the planted clique analysis. However, this may not hold in general. To handle this, we note that the existence of v_σ can be relaxed as follows: Let H be the matrix with rows and columns indexed by left shapes σ such that $H(\sigma, \sigma') = \lambda_{\sigma \circ {\sigma'}^T}$. Up to scaling, H will be one of our coefficient matrices. 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 the main theorem in the machinery.

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 τ is non-trivial. Our strategy will be to charge these terms to other terms.

For the sake of simplicity, we will describe how to handle one term. A starting point is the following inequality. For a left shape σ, a middle shape τ, a right shape ${\sigma'}^T$, and real

numbers a, b,

$$(aM_\sigma - bM_{\sigma'}M_{\tau^T})(aM_\sigma - bM_{\sigma'}M_{\tau^T})^T \succeq 0$$

which rearranges to

$$\begin{aligned} ab(M_\sigma M_\tau M_{\sigma'^T} + (M_\sigma M_\tau M_{\sigma'^T})^T) &\preceq a^2 M_\sigma M_{\sigma^T} + b^2 M_{\sigma'} M_{\tau^T} M_\tau M_{\sigma'^T} \\ &\preceq a^2 M_\sigma M_{\sigma^T} + b^2 \|M_\tau\|^2 M_{\sigma'} M_{\sigma'^T} \end{aligned}$$

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

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

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. Instead, the machinery generalizes this inequality to work with the entire set of shapes σ, σ' for a fixed τ. This will lead us to the second condition of the main theorem of the machinery.

Handing intersection terms There's one important technicality in the above heuristic calculations. Whenever we decompose α into left, middle, and right parts σ, τ, and σ'^T, $M_\sigma M_\tau M_{\sigma'^T}$ is only approximately equal to $M_\alpha = M_{\sigma\circ\tau\circ\sigma'^T}$. All the other error terms have to be carefully handled in the analysis. We call these terms intersection terms.

These intersection terms themselves turn out to be graph matrices and the strategy is to now recursively decompose them into $\sigma_2 \circ \tau_2 \circ \sigma_2'^T$ and apply the previous ideas. To do this methodically, the machinery employs several ideas such as the notion of intersection patterns and the generalized intersection tradeoff lemma. Properly handling the intersection terms is one of the most technically intensive parts of [149]. This analysis leads us to the third

condition of the main theorem of the machinery.

Applying the machinery To apply the machinery to our problems of interest, we verify the spectral conditions that our coefficients should satisfy and then we can use the main theorem. The Planted slightly denser subgraph application is straightforward and will serve as a good warmup to understand the 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 the machinery. We do this in this work.

4.6 Related work on Sum-of-Squares Lower Bounds for Certification Problems

[107] 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 they don't state it in this manner, the pseudo-expectation values used by [107] can also be derived using pseudo-calibration [156, 35]. 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.

[129] 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.

[113] 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.

4.7 Organization of the proofs

We prove the Sherrington-Kirkpatrick lower bound, Theorem 4.1.2, in Chapter 5. The proofs for planted slightly denser subgraph, tensor PCA and sparse PCA, namely Theorem 4.2.1, Theorem 4.3.1 and Theorem 4.4.1, are split between Chapter 6 and Chapter 7. The latter proofs are split into qualitative and quantitative versions. Qualitative theorem statements capture the essence of the inequalities we prove, and serve to illustrate the main forms of the bounds we desire, without getting lost in the details. Quantitative theorems on the other hand build on their qualitative counterparts by stating the precise bounds that are needed. In Chapter 6, we introduce the machinery and and in Section 6.2, Section 6.3 and Section 6.4, 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 Chapter 7, where we fully verify the conditions of the machinery in Section 7.1, Section 7.2 and Section 7.3.

CHAPTER 5

THE SHERRINGTON-KIRKPATRICK HAMILTONIAN

In this chapter, we will prove SoS lower bounds for the certification problem of the Sherrington-Kirkpatrick Hamiltonian, in particular Theorem 4.1.2. The material in this chapter is adapted from [70], where this work originally appeared.

5.1 Technical preliminaries

In this section we record formal problem statements, then define and discuss one of the main objects in our SoS lower bound: graph matrices.

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

5.1.1 Problem statements

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

Definition 5.1.1 (Planted Affine Planes (PAP) problem)**.** *Given $d_1, \ldots, d_m \sim \mathcal{D}$ where each d_u is a vector in $\mathbb{R}^n$, determine whether there exists $v \in \{\pm \frac{1}{\sqrt{n}}\}^n$ such that*

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

for every $u \in [m]$.

Our results hold for the Gaussian setting $\mathcal{D} = \mathcal{N}(0, I)$ and the boolean setting where $\mathcal{D}$ is uniformly sampled from $\{\pm 1\}^n$, though we conjecture in Section 8.2.2 that similar SoS bounds hold under more general conditions on $\mathcal{D}$.

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

As we saw in Chapter 4, the Sherrington–Kirkpatrick (SK) problem comes from the spin-glass model in statistical physics [168].

Definition 5.1.2 (Sherrington-Kirkpatrick problem)**.** *Given $W \sim \mathrm{GOE}(n)$, compute*

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

The Planted Boolean Vector problem was introduced by Mohanty–Raghavendra–Xu [129], where it was called the "Boolean Vector in a Random Subspace".

Definition 5.1.3 (Planted Boolean Vector problem)**.** *Given a uniformly random p-dimensional subspace V of $\mathbb{R}^n$ in the form of a projector Π_V onto V, compute*

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

5.1.2 Graph matrices

To study $\mathcal{M}$, we decompose it using the framework of *graph matrices*. Originally developed in the context of the planted clique problem, graph matrices are random matrices whose entries are symmetric functions of an underlying random object – in our case, the set of vectors $d_1, \ldots, d_m$. We take the general presentation and results from [2]. For our purposes, the following definitions are sufficient.

The graphs that we study have two types of vertices, circles ◯ and squares □. We let $\mathcal{C}_m$ be a set of m circles labeled 1 through m, which we denote by $\textcircled{1}, \textcircled{2}, \ldots, \textcircled{m}$, and let $\mathcal{S}_n$ be a set of n squares labeled 1 through n, which we denote by $\boxed{1}, \boxed{2}, \ldots, \boxed{n}$. We will work with bipartite graphs with edges between circles and squares, which have positive integer labels on

the edges. When there are no multiedges (the graph is simple), such graphs are in one-to-one correspondence with Fourier characters on the vectors d_u. An edge between $\textcircled{u}$ and $\boxed{i}$ with label l represents $h_l(d_{u,i})$ where $\{h_k\}$ is the Fourier basis (e.g. Hermite polynomials).

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

An example of a Fourier polynomial as a graph with labeled edges is given in Fig. 5.1. Unlabeled edges are implicitly labeled 1.

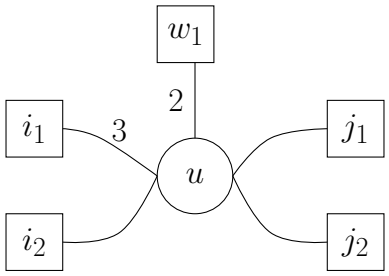

Figure 5.1: The Fourier polynomial $h_3(d_{u,i_1})h_1(d_{u,i_2})h_2(d_{u,w_1})h_1(d_{u,j_1})h_1(d_{u,j_2})$ represented as a graph.

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

Definition 5.1.4 (Proper)**.** *We say an edge-labeled graph is proper if it has no multiedges.*

The definitions allow for "improper" edge-labeled multigraphs which simplify multiplying graph matrices (Section 5.4.2).

Definition 5.1.5 (Matrix indices)**.** *A matrix index is a set A of elements from $\mathcal{C}_m \cup \mathcal{S}_n$.*

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

Definition 5.1.6 (Ribbons)**.** *A ribbon is an undirected, edge-labeled graph $R = (V(R), E(R), A_R, B_R)$, where $V(R) \subseteq \mathcal{C}_m \cup \mathcal{S}_n$ and A_R, B_R are two matrix indices (possibly not disjoint) with $A_R, B_R \subseteq V(R)$, representing two distinguished sets of vertices. Furthermore, all edges in $E(R)$ go between squares and circles.*

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

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

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

Definition 5.1.7 (Matrix for a ribbon)**.** *The matrix M_R has rows and columns indexed by subsets of $\mathcal{C}_m \cup \mathcal{S}_n$, with a single nonzero entry defined by*

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

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

Definition 5.1.8 (Index shapes)**.** *An index shape is a set U of formal variables. Furthermore, each variable is labeled as either a "circle" or a "square".*

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

Definition 5.1.9 (Shapes)**.** *A shape is an undirected, edge-labeled graph $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ where $V(\alpha)$ is a set of formal variables, each of which is labeled as either a "circle" or a "square". U_α and V_α are index shapes (possibly with variables in common) such that $U_\alpha, V_\alpha \subseteq V(\alpha)$. The edge set $E(\alpha)$ must only contain edges between the circle variables and the square variables.*

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

Remark 5.1.10. *We will abuse notation and use $\boxed{i}, \boxed{j}, \textcircled{u}, \textcircled{v}, \ldots$ for both the vertices of ribbons and the vertices of shapes. If they are ribbon vertices, then the vertices are elements of $\mathcal{C}_m \cup \mathcal{S}_n$ and if they are shape vertices, then they correspond to formal variables with the appropriate type.*

Definition 5.1.11 (Trivial shape)**.** *Define a shape α to be trivial if $U_\alpha = V_\alpha$, $W_\alpha = \emptyset$ and $E(\alpha) = \emptyset$.*

Definition 5.1.12 (Transpose of a shape)**.** *The transpose of a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is defined to be the shape $\alpha^\intercal = (V(\alpha), E(\alpha), V_\alpha, U_\alpha)$.*

For a shape α and an injective map $\sigma : V(\alpha) \to \mathcal{C}_m \cup \mathcal{S}_n$, we define the realization $\sigma(\alpha)$ as a ribbon in the natural way, by labeling all the variables using the map σ. We also require σ to be type-preserving i.e. it takes square variables to $\mathcal{S}_n$ and circle variables to $\mathcal{C}_m$. The ribbons that result are referred to as *ribbons of shape α*; notice that this partitions the set of all ribbons according to their shape[1][2].

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

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

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

Definition 5.1.13 (Graph matrices)**.** *Given a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$, the graph matrix M_α is*

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

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

5.1.3 Norm bounds

Similar to the norm bounds for graph matrices with only a single type of vertex (see Chapter 2), the spectral norm of a graph matrix in our setting is determined, up to logarithmic factors, by relatively simple combinatorial properties of the graph. For a subset $S \subseteq \mathcal{C}_m \cup \mathcal{S}_n$, we define the weight $w(S) := (\#\text{ circles in } S) \cdot \log_n(m) + (\#\text{ squares in } S)$. Observe that $n^{w(S)} = m^{\#\text{ circles in } S} \cdot n^{\#\text{ squares in } S}$.

Definition 5.1.14 (Minimum vertex separator)**.** *For a shape α, a set $S_{\min}$ is a minimum vertex separator if all paths from U_α to V_α pass through $S_{\min}$ and $w(S_{\min})$ is minimized over all such separating sets.*

Let W_{iso} denote the set of isolated vertices in W_α. Then essentially the following norm bound holds for all shapes α with high probability (a formal statement can be found in Section 5.6.1):

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

In fact, the only probabilistic property required of the inputs $d_1, \ldots, d_m$ by our proof is that the above norm bounds hold for all shapes that arise in the analysis. We henceforth assume that the norm bounds in Lemma 5.6.3 (for the Gaussian case) and Lemma 5.6.1 (for the boolean case) hold.

5.2 Proof Strategy

Now we explain in more detail the ideas for the Planted Affine Planes lower bound. Towards the proof of Theorem 4.1.4, fix a constant $\varepsilon > 0$ and a random instance $d_1, \ldots, d_m$ with $n \le m \le n^{3/2-\varepsilon}$. We will construct a pseudoexpectation operator and show that it is PSD up to degree $D = 2 \cdot n^{\delta}$ with high probability.

We start by pseudocalibrating to obtain a pseudoexpectation operator $\widetilde{\mathbb{E}}$. The operator $\widetilde{\mathbb{E}}$ will exactly satisfy the "booleanity" constraints "$v_i^2 = \frac{1}{n}$" though it may not exactly satisfy the constraints "$\langle v, d_u \rangle^2 = 1$" due to truncation error in the pseudocalibration. Taking the truncation parameter n^{τ} to be larger than the degree D of the SoS solution, i.e., $\delta \ll \tau$, the truncation error is small enough that we can round $\widetilde{\mathbb{E}}$ to a nearby $\widetilde{\mathbb{E}}'$ that exactly satisfies the constraints. This is formally accomplished by viewing $\widetilde{\mathbb{E}} \in \mathbb{R}^{\binom{[n]}{\le D}}$ as a vector and expressing the constraints as a matrix Q such that $\widetilde{\mathbb{E}}$ satisfies the constraints iff it lies in the null space of Q. The choice of $\widetilde{\mathbb{E}}'$ is then the projection of $\widetilde{\mathbb{E}}$ to $\mathrm{Null}(Q)$. The end result is that we construct a moment matrix $\mathcal{M}_{fix} = \mathcal{M} + \mathcal{E}$ that exactly satisfies the constraints such that $\|\mathcal{E}\|$ is tiny. For the sake of brevity, we omit this technicality in this work, see [70] for the details.

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

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

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

In this decomposition of $\mathcal{M}$, the trivial shapes will be the dominant terms which we will use to bound the other terms. Recall that a shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is trivial if $U_\alpha = V_\alpha$, $W_\alpha = \emptyset$ and $E(\alpha) = \emptyset$. These shapes contribute scaled identity matrices on

different blocks of the main diagonal of $\mathcal{M}$, with trivial shape α contributing an identity matrix with coefficient $n^{-|U_\alpha|}$. Two trivial shapes are illustrated in Fig. 5.2.

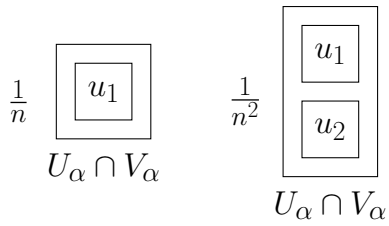

Figure 5.2: Two examples of trivial shapes.

Let $\mathcal{M}_{\text{triv}}$ be this diagonal matrix of trivial shapes in the above decomposition of $\mathcal{M}$. To prove that $\mathcal{M} \succeq 0$, we attempt the simple strategy of showing that the norm of all other terms can be "charged" against this diagonal matrix $\mathcal{M}_{\text{triv}}$. For several shapes this strategy is indeed viable. To illustrate, let's consider one such shape α depicted in Fig. 5.3.

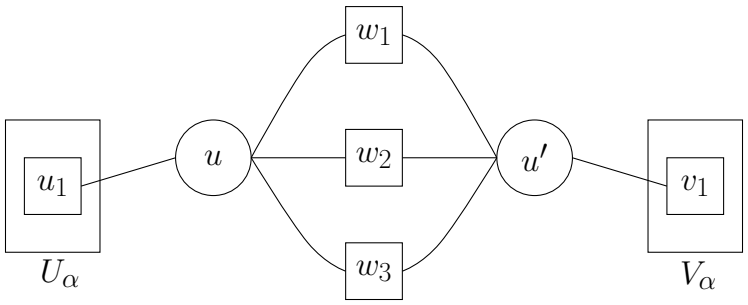

Figure 5.3: Picture of basic non-spider shape α.

This graph matrix has $|\lambda_\alpha| = \Theta(\frac{1}{n^5})$. Using the graph matrix norm bounds, with high probability the norm of this graph matrix is $\tilde{O}(n^2 m)$: there are four square vertices and two circle vertices which are not in the minimum vertex separator. Thus, for this shape α, with high probability $|\lambda_\alpha| \, \|M_\alpha\|$ is $\tilde{O}\left(\frac{m}{n^3}\right)$ and thus $\lambda_\alpha M_\alpha \preceq \frac{1}{n} Id$ (which is the multiple of the identity appearing in the corresponding block of $\mathcal{M}_{\text{triv}}$).

Unfortunately, some shapes α that appear in the decomposition have $\|\lambda_\alpha M_\alpha\|$ too large to be charged against $\mathcal{M}_{\text{triv}}$. These are shapes with a certain substructure (actually the same structure that appears in the matrix Q used to project the pseudoexpectation operator!)

whose norms cannot be handled by the preceding argument, and which we denote *spiders*. The following graph depicts one such *spider* shape (and also motivates this terminology):

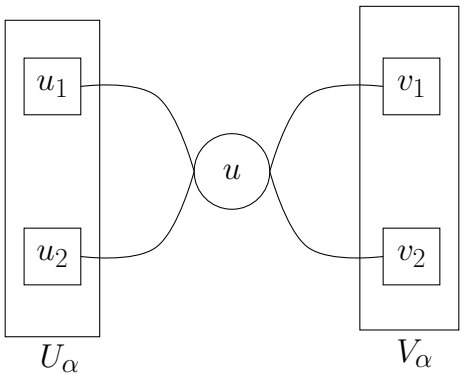

Figure 5.4: Picture of basic spider shape α.

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

To skirt this and other spiders, we restrict ourselves to vectors $x \perp \text{Null}(M)$, and observe that this spider α satisfies $x^\intercal M_\alpha \approx 0$. To be more precise, consider the following argument. Consider the two shapes in Fig. 5.5, β_1 and β_2 (take note of the label 2 on the edge in β_2).

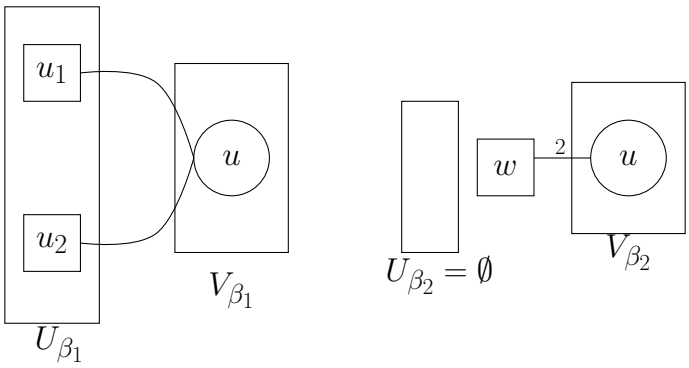

Figure 5.5: Picture of shapes β_1 and β_2.

We claim that every column of the matrix $2M_{\beta_1} + \frac{1}{n}M_{\beta_2}$ is in the null space of $\mathcal{M}$. There are m nonzero columns indexed by assignments to V, which can be a single circle

$\textcircled{1}, \textcircled{2}, \ldots, \textcircled{m}$. The nonzero rows are $\emptyset$ in β_2 and $\{\boxed{i}, \boxed{j}\}$ for $i \neq j$ in β_1. Fixing $I \subseteq [n]$, entry $(I, \textcircled{u})$ of the product matrix $\mathcal{M}(2M_{\beta_1} + \frac{1}{n} M_{\beta_2})$ is

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

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

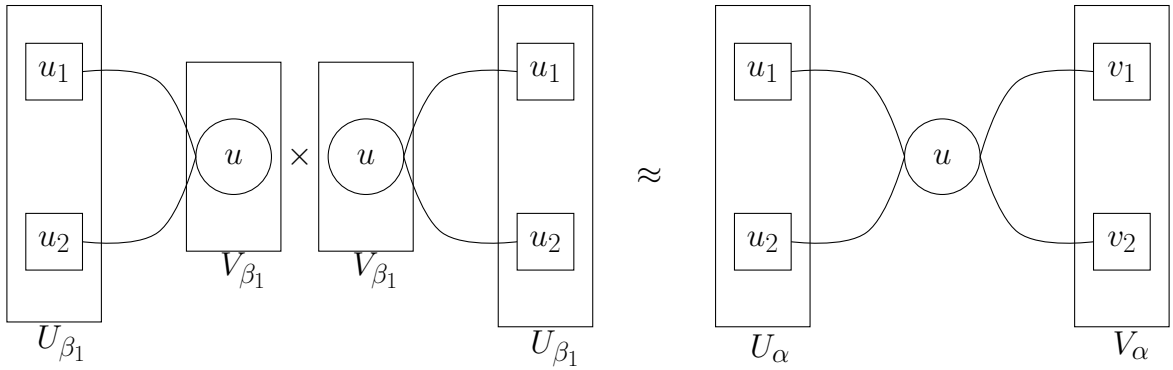

Figure 5.6: Approximation $\beta_1 \times \beta_1^\intercal \approx \alpha$.

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

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

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

Fact 5.2.1. *If $x \perp \text{Null}(\mathcal{M})$ and $\mathcal{M}A = 0$, then $x^\intercal(AB+\mathcal{M})x = x^\intercal(B^\intercal A^\intercal+\mathcal{M})x = x^\intercal\mathcal{M}x$.*

Using the fact, we can freely add multiples of A to $\mathcal{M}$ without changing the action of $\mathcal{M}$ on Null$(\mathcal{M})^\perp$. A judicious choice is to subtract $\lambda_\alpha A$ which will "kill" the spider from $\mathcal{M}$. Doing this for all spiders, we produce a matrix whose action is equivalent on Null$(\mathcal{M})^\perp$, and which has high minimum eigenvalue by virtue of the fact that it has no spiders, showing that $\mathcal{M}$ is PSD. The catch is two-fold: first, the coefficients c_β may contribute to the coefficients on the non-spiders; second, the further intersection terms M_β may themselves be spiders (though they will always have fewer square vertices than α). Thus we must recursively kill these spiders, until there are no spiders remaining in the decomposition of $\mathcal{M}$. The resulting matrix has some new coefficients on the non-spiders

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

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

5.3 Pseudocalibration

As we saw in Chapter 3, to be able to apply the pseudocalibration technique to an average-case feasibility problem, in our case the PAP problem, one needs to design a planted distribution supported on feasible instances. This is done in Section 5.3.1. In Section 5.3.2, we recall the precise details in applying pseudocalibration. Then we pseudocalibrate in the Gaussian (Section 5.3.3) and boolean (Section 5.3.4) settings.

5.3.1 PAP planted distribution

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

Definition 5.3.1 (Gaussian PAP distributions)**.** *The Gaussian PAP distributions are as follows.*

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

For the boolean setting, we have the following distributions.

Definition 5.3.2 (Boolean PAP distributions)**.** *The boolean PAP distributions are as follows*

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

5.3.2 Pseudocalibration technique

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

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

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

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

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

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

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

Fact 5.3.3 (Proof in [70])**.** *If $p(v)$ is a polynomial which is uniformly zero on the planted distribution, then $\widetilde{\mathbb{E}}[p]$ is the zero function. If $p(v, d)$ is a polynomial which is uniformly zero on the planted distribution, then the only nonzero Fourier coefficients of $\widetilde{\mathbb{E}}[p]$ are those with size between $n^\tau \pm \deg_d(p)$.*

Truncation introduces a tiny error in the constraints, which we are able to handle in [70],

omitted in this work for brevity.

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

Pseudocalibration also by default does not enforce the condition $\widetilde{\mathbb{E}}[1] = 1$. However, this is easily fixed by dividing the operator by $\widetilde{\mathbb{E}}[1]$. As will be pointed out in Remark 5.4.9, w.h.p. in the unnormalized pseudocalibration, $\widetilde{\mathbb{E}}[1] = 1 + \mathrm{o}_n(1)$ and so the error introduced does not impact the statement of any lemmas.

5.3.3 Gaussian setting pseudocalibration

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

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

Lemma 5.3.4. *For any $I \subseteq [n]$, the pseudocalibration value is*

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

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

Proof. The truncated pseudocalibrated value is defined to be

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

So we set about to compute the planted moments. For this computation, the following lemma is crucial. Here, we give a short proof of this lemma using generating functions. For a different combinatorial proof, see [70].

Lemma 5.3.5. *Let $\alpha \in \mathbb{N}^n$. When v is fixed and b is fixed (not necessarily ± 1) and $d \sim N(0, I)$ conditioned on $\langle v, d \rangle = b \|v\|$,*

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

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

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

The Hermite polynomial generating function is

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

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

Completing the square,

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

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

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

$$\begin{aligned}
\mathbb{E}_{\text{pl}}[h_\alpha(d_1, \ldots, d_m)v^I] &= \mathbb{E}_{v,b_u} v^I \prod_{u=1}^{m} \mathbb{E}_d \left[h_{\alpha_u}(d_u) \mid v, b_u\right] \\
&= \mathbb{E}_{v,b_u} v^I \cdot \prod_{u=1}^{m} \frac{v^{\alpha_u}}{\|v\|^{|\alpha_u|}} \cdot h_{|\alpha_u|}(b_u) \qquad \text{(Lemma 5.3.5)} \\
&= \left(\mathbb{E}_v \frac{v^{I+\sum_{u=1}^m \alpha_u}}{\|v\|^{\sum_{u=1}^m |\alpha_u|}}\right) \cdot \left(\prod_{u=1}^{m} \mathbb{E}_{b_u} h_{|\alpha_u|}(b_u)\right)
\end{aligned}$$

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

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

Definition 5.3.6. *Let $\mathcal{L}$ be the set of all proper shapes α with the following properties*

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

Remark 5.3.7. *Note that the shapes in $\mathcal{L}$ can have isolated vertices in $U_\alpha \cap V_\alpha$.*

Remark 5.3.8. *$\mathcal{L}$ captures all the shapes that have nonzero coefficient when we write $\mathcal{M}$ in terms of graph matrices. The constraint $\deg(\textcircled{u}) \ge 4$ arises because pseudocalibration gives us that $\deg(\textcircled{u})$ is even, $\textcircled{u}$ cannot be isolated, and $h_2(1) = 0$.*

For a shape α, we define

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

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

Definition 5.3.9. *For any shape α, if $\alpha \in \mathcal{L}$, define*

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

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

Corollary 5.3.10. *Modulo the footnote*[3]*, $\mathcal{M} = \sum_{\text{shapes } \alpha} \lambda_\alpha M_\alpha$.*

3. Technically, the graph matrices M_α have rows and columns indexed by all subsets of $\mathcal{C}_m \cup \mathcal{S}_n$. The submatrix with rows and columns from $\binom{\mathcal{S}_n}{\le D/2}$ equals the moment matrix for $\widetilde{\mathbb{E}}$.

5.3.4 Boolean setting pseudocalibration

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

Definition 5.3.11 (Slice)**.** *Let $v \in \{\pm 1\}^n$ and $\theta \in \mathbb{Z}$. The slice $\mathcal{S}_v(\theta)$ is defined as*

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

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

Remark 5.3.12. *With our notation for the slice, the planted distribution in the boolean setting can be equivalently described as*

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

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

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

Definition 5.3.13. $e(k) \coloneqq \mathbb{E}_{x \in_R \mathcal{S}(\sqrt{n})} [x_1 \cdots x_k]$.

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

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

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

Lemma 5.3.14. *We have*

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

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

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

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

Computing the RHS above yields

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

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

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

Definition 5.3.15. *Let $\mathcal{L}_{bool}$ be the set of shapes in $\mathcal{L}$ from Definition 5.3.6 in which the edge labels are all 1.*

Remark 5.3.16. *$\mathcal{L}_{bool}$ captures all the shapes that have nonzero coefficient when we write $\mathcal{M}$ in terms of graph matrices. Similar to Remark 5.3.8, since $e(2) = 0$ (see Claim 5.6.5), we have the same condition $deg(\textcircled{u}) \geq 4$ for shapes in $\mathcal{L}_{bool}$.*

Definition 5.3.17. *For all shapes α, if $\alpha \in \mathcal{L}_{bool}$ define*

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

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

Corollary 5.3.18. $\mathcal{M} = \sum_{shapes\ \alpha} \lambda_\alpha M_\alpha$

Unifying the analysis

It turns out that the analysis of the boolean setting mostly follows from the analysis in the Gaussian setting. Initially, the boolean pseudocalibration is essentially equal to the Gaussian pseudocalibration in which we have removed all shapes containing at least one edge with a label $k \geq 2$. The coefficients on the graph matrices will actually be slightly different, but they both admit an upper bound that is sufficient for our purposes (see Proposition 5.4.13 for the precise statement).

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

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

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

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

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

5.4 Proving PSD-ness

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

Definition 5.4.1 (Block)**.** *For $k, l \in \{0, 1, \ldots, D/2\}$, the (k, l) block of $\mathcal{M}$ is the submatrix with rows from $\binom{[n]}{k}$ and columns from $\binom{[n]}{l}$. Note that when $\mathcal{M}$ is expressed as a sum of graph matrices, this exactly restricts $\mathcal{M}$ to shapes α with $|U_\alpha| = k$ and $|V_\alpha| = l$.*

We define the parameter $\eta := 1/\sqrt{n}$. The trivial shapes live in the diagonal blocks of $\mathcal{M}$, and on the (k, k) block contribute a factor of $\frac{1}{n^k} = \eta^{2k}$ on the diagonal. In principle, we

could make η as small as we like[4] by considering the moments of a rescaling of v rather than v itself. Counterintuitively, it will turn out that the scaling helps us prove PSD-ness (see [70] for more details). It turns out that pseudocalibrating v as a unit vector (equivalently, using $\eta = 1/\sqrt{n}$) is sufficient for our analysis.

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

Lemma 5.4.2. *Suppose a symmetric matrix $\mathcal{A} \in \mathbb{R}^{\binom{[n]}{\leq D} \times \binom{[n]}{\leq D}}$ satisfies, for some parameter $\eta \in (0,1)$,*

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

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

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

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

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

■

We start by defining spiders, which are special shapes α that we will handle separately in the decomposition of $\mathcal{M}$. Informally, these contain special substructures which allow their

4. Though pseudocalibration truncation errors may become nonnegligible for extremely tiny η.

norm bounds not to be negligible with respect to the identity matrix. We then show that shapes which are not spiders have bounded norms.

Definition 5.4.3 (Left Spider)**.** *A left spider is a proper shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ with the property that there exist two distinct square vertices $\boxed{i}, \boxed{j} \in U_\alpha$ of degree 1 and a circle vertex $\textcircled{u} \in V(\alpha)$ such that $E(\alpha)$ contains the edges $(\boxed{i}, \textcircled{u})$ and $(\boxed{j}, \textcircled{u})$ (these are necessarily the only edges incident to $\boxed{i}$ and $\boxed{j}$).*

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

Definition 5.4.4 (Right spider)**.** *A shape $\alpha = (V(\alpha), E(\alpha), U_\alpha, V_\alpha)$ is a right spider if $\alpha^\intercal = (V(\alpha), E(\alpha), V_\alpha, U_\alpha)$ is a left spider. The end vertices of $\alpha^\intercal$ are also called the end vertices of α.*

Definition 5.4.5 (Spider)**.** *A shape α is a spider if it is either a left spider or a right spider.*

Remark 5.4.6. *A spider can have many pairs of end vertices. For each possible spider shape, we single out a pair of end vertices, so that in what follows we can discuss "the" end vertices of the spider.*

5.4.1 Non-spiders are negligible

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

Lemma 5.4.7. *If $\alpha \in \mathcal{L}$ is not a trivial shape and not a spider, then*

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

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

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

Remark 5.4.8. *Instead of using the minimum vertex separator, we will actually use a set S of square vertices such that $w(S) \leq w(S_{\min})$. For details, see the actual distribution scheme below.*

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

> Attempt 1: Take each edge and assign a factor of $\frac{1}{\sqrt[4]{n}}$ to its square endpoint and a factor of $\frac{1}{\sqrt[8]{m}}$ to its circle endpoint.
>
> With this distribution scheme, since each circle vertex has degree at least 4, each circle vertex is assigned a factor of $\frac{1}{\sqrt{m}}$ or smaller. Since each square vertex in W_α has degree at least 2, each square vertex in W_α is assigned a factor of $\frac{1}{\sqrt{n}}$ or smaller. However, square vertices in $U_\alpha \cup V_\alpha$ may only have degree 1 in which case they are assigned a factor of $\frac{1}{\sqrt[4]{n}}$ which is not small enough.
>
> To fix this issue, we can have all of the edges which are incident to a square vertex in $U_\alpha \cup V_\alpha$ give their entire factor of $\frac{1}{\sqrt{n}}$ to the square vertex.
>
> **Remark 5.4.9.** *For analyzing $\widetilde{\mathbb{E}}[1]$, this first attempt works as $U_\alpha = V_\alpha = \emptyset$. Thus, as long as $m \leq n^{2-\varepsilon}$, with high probability $\widetilde{\mathbb{E}}[1] = 1 \pm \mathrm{o}_n(1)$.*

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

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

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

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

1. Choose a set of square vertices $S \subseteq U_\alpha \cup V_\alpha$ as follows. Start with $S = U_\alpha \cap V_\alpha$. Whenever we have a circle vertex which is adjacent to both a square vertex in $U_\alpha \setminus V_\alpha$ and a square vertex in $V_\alpha \setminus U_\alpha$, put one of these two square vertices in S (this choice is arbitrary). Observe that $w(S) \leq w(S_{\min})$
2. For each edge which is incident to a square vertex in S, assign a factor of $\frac{1}{\sqrt[3]{m}}$ to its circle endpoint and nothing to this square.
3. For each edge which is incident to a square vertex in $(U_\alpha \cup V_\alpha) \setminus S$, assign a factor of $\frac{1}{\sqrt{n}}$ to the square vertex and nothing to the circle vertex.

4. For all other edges, assign a factor of $\frac{1}{\sqrt[4]{n}}$ to its square endpoint and a factor of $\frac{1}{\sqrt[6]{m}}$ to its circle endpoint.

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

We now make this argument formal.

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

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

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

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

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

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

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

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

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

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

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

Using this, we get

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

■

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

Corollary 5.4.10. *If $\alpha \in \mathcal{L}_{bool}$ is not a trivial shape and not a spider, then*

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

Corollary 5.4.11. *If $\alpha \in \mathcal{L}$ is not a trivial shape and not a spider, then w.h.p.*

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

Proof. Using the norm bounds in Lemma 5.6.3, we have

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

We have $W_{iso}=\emptyset$. Observe that since there are no degree 0 vertices in $V_{rel}(\alpha)$, we have that $|V_{rel}(\alpha)|\leq 2|E(\alpha)|$ and since we also have $|V(\alpha)|\cdot(1+|E(\alpha)|)\cdot\log n \leq n^{O(\tau)}$, the factor $2\cdot(|V(\alpha)|\cdot(1+|E(\alpha)|)\cdot\log(n))^{C\cdot(|V_{rel}(\alpha)|+|E(\alpha)|)}$ can be absorbed into $\frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$. The result follows from Lemma 5.4.7. ∎

This says that nontrivial non-spider shapes have $o_n(1)$ norm (ignoring the extra factor η for the moment). We now demonstrate how to use this norm bound to control the total norm of all non-spiders in a block of $\mathcal{M}$, Corollary 5.4.14. We will first need a couple propositions which will also be of use to us later after we kill the spiders.

Proposition 5.4.12. *The number of proper shapes with at most L vertices and exactly k edges is at most $L^{8(k+1)}$.*

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

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

∎

Proposition 5.4.13. $|\lambda_\alpha|\leq \eta^{|U_\alpha|+|V_\alpha|}\cdot\frac{|E(\alpha)|^{3\cdot|E(\alpha)|}}{n^{|E(\alpha)|/2}}$ *where we assume by convention that* $0^0=1$.

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

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

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

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

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

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

Using Corollary 5.6.16, we have that $|e(k)| \leq k^{3k} \cdot n^{-k/2}$. Thus,

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

■

Corollary 5.4.14. *For $k, l \in \{0, 1, \ldots, D/2\}$, let $\mathcal{B}_{k,l} \subseteq \mathcal{L}$ denote the set of nontrivial, non-spiders $\alpha \in \mathcal{L}$ on the (k, l) block i.e. $|U_\alpha| = k, |V_\alpha| = l$. The total norm of the non-spiders in $\mathcal{B}_{k,l}$ satisfies*

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

Proof.

$$\begin{aligned}
\sum_{\alpha\in\mathcal{B}_{k,l}} |\lambda_\alpha| \, \|M_\alpha\| &\leq \sum_{\alpha\in\mathcal{B}_{k,l}} \eta^{k+l} \cdot \frac{|E(\alpha)|^{3|E(\alpha)|}}{n^{|E(\alpha)|/2}} \, \|M_\alpha\| && \text{(Proposition 5.4.13)} \\
&\leq \eta^{k+l} \cdot \sum_{\alpha\in\mathcal{B}_{k,l}} \left(\frac{|E(\alpha)|^3}{n^{\Omega(\varepsilon)}} \right)^{|E(\alpha)|} && \text{(Corollary 5.4.11)} \\
&\leq \eta^{k+l} \cdot \sum_{\alpha\in\mathcal{B}_{k,l}} \left(\frac{n^{3\tau}}{n^{\Omega(\varepsilon)}} \right)^{|E(\alpha)|} && (\alpha \in \mathcal{L}) \\
&\leq \eta^{k+l} \cdot \sum_{\alpha\in\mathcal{B}_{k,l}} \frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}} \\
&\leq \eta^{k+l} \cdot \sum_{i=1}^{\infty} \frac{n^{O(\tau i)}}{n^{\Omega(\varepsilon i)}} && \text{(Proposition 5.4.12 and } |E(\alpha)| \geq 1 \text{ for } \alpha \in \mathcal{B}_{k,l}) \\
&= \eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}} \qquad \blacksquare
\end{aligned}$$

5.4.2 Killing a single spider

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

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

Definition 5.4.15 (Improper collapse)**.** *Let α be a shape and let $\boxed{i}, \boxed{j}$ be two distinct square vertices in $V(\alpha)$. We define the improper collapse of $\boxed{i}, \boxed{j}$ by:*

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

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

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

Definition 5.4.16 (Collapsing a shape)**.** *Let α be a shape with two distinct square vertices $\boxed{i}, \boxed{j}$. We say that β is a (proper) collapse of $\boxed{i}, \boxed{j}$ if β appears in the expansion of the improper collapse of $\boxed{i}, \boxed{j}$.*

Remark 5.4.17. *If $l_1, \ldots, l_k$ are the labels of a set of parallel edges, then the product $h_{l_1}(z) \cdots h_{l_k}(z)$ is even/odd depending on the parity of $l_1 + \cdots + l_k$. Thus the nonzero Fourier coefficients will be the terms of matching parity. Therefore, in both the boolean and Gaussian cases, the shapes that are proper collapses of a given improper collapse are formed by replacing each set of parallel edges by a single edge e such that $l(e) \leq l_1 + \ldots + l_k$ and $l(e) \equiv l_1 + \cdots + l_k \pmod 2$.*

Remark 5.4.18. *Looking at the definition and in light of the previous remark, we have the following.*

1. *The number of circle vertices does not change by collapsing a shape but the number of square vertices decreases by* 1.

2. $\alpha \in \mathcal{L}$ *has the property that the vertices have odd degree if and only if they are in* $(U_\alpha \cup V_\alpha) \setminus (U_\alpha \cap V_\alpha)$. *When* α *collapses, this property is preserved.*

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

Definition 5.4.19. *For* $k \geq 2$ *define the shape* ℓ_k *on* $\{\boxed{1}, \ldots, \boxed{k}, \textcircled{1}\}$ *with two edges* $\{\{\boxed{1}, \textcircled{1}\}, \{\boxed{2}, \textcircled{1}\}\}$. *The left side of* ℓ_k *consists of* $U_{\ell_k} = \{\boxed{1}, \ldots, \boxed{k}\}$. *The right side consists of* $V_{\ell_k} = \{\boxed{3}, \ldots, \boxed{k}, \textcircled{1}\}$.

Definition 5.4.20. *Define the "completed" version* L_k *of* ℓ_k *to be the matrix which is the sum of* $c_\beta M_\beta$ *for* β *being the following shapes with coefficients:*

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

For a pictorial representation of the ribbons/shapes, see Fig. 5.7 below.

Lemma 5.4.21. $\mathcal{M}_{fix} L_k = 0$

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

$j_1, j_2 \in [n]$. That is, exactly the following equality holds,

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

To demonstrate how the coefficients arise, we analyze the ribbons R which L_k is composed of and see how they contribute to the output. For pictures of the ribbons/shapes, see Fig. 5.7 below. Let the ribbon be partially realized as $\textcircled{u}$ and $S = \{\boxed{j_3}, \ldots, \boxed{j_k}\}$. Let $(M_{fix} L_k)_{I(u,S)}$ denote the terms in $(M_{fix} L_k)_I$ with this partial realization. In this notation we want to show

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

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

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

   This ribbon must "cover" both ordered pairs (j_1, j_2) and (j_2, j_1), so we want each such ribbon R to appear with a coefficient of 2 in L_k.

2. If we take a ribbon R with $A_R = \{\boxed{j_1}, \ldots, \boxed{j_k}\} \setminus \{\boxed{j_1}, \boxed{j_3}\}$, $B_R = \{\boxed{j_3}, \ldots, \boxed{j_k}\} \cup \{\textcircled{u}\}$

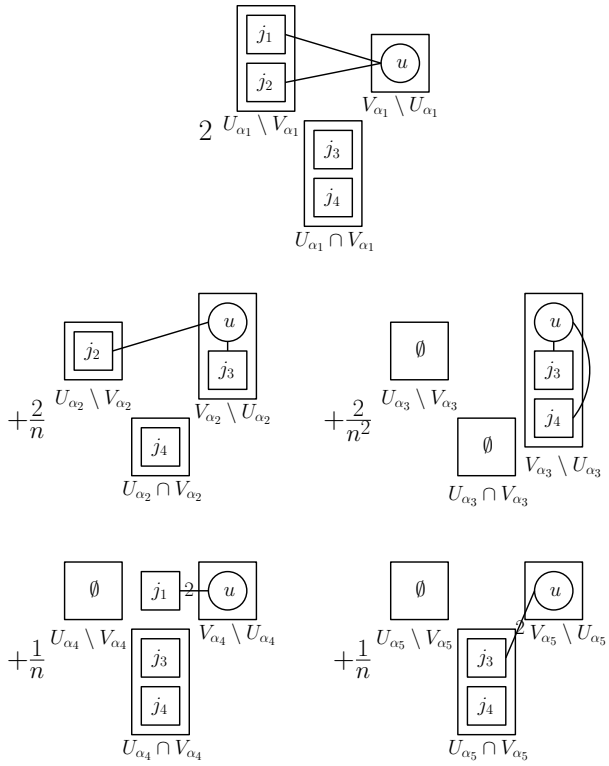

Figure 5.7: The five shapes that make up L_4.

and $E(R) = \{\{\boxed{j_3}, \textcircled{u}\}, \{\boxed{j_2}, \textcircled{u}\}\}$ where $j_1 = j_3 \in S$ then

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

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

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

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

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

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

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

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

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

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

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

■

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

Definition 5.4.22. *Say that shapes α and β are composable if V_α and U_β have the same number of square and circle vertices. We say a shape γ is a gluing of α and β, if the graph of γ is the disjoint union of the graphs of α and β, followed by identifying V_α and U_β under some type-preserving bijection, and if $U_\gamma = U_\alpha$ and $V_\gamma = V_\beta$.*

Proposition 5.4.23. *Let α, β be composable shapes. Assume that $V(\alpha) \setminus V_\alpha$ has only square vertices. Let $\{\gamma_i\}$ be the distinct gluings of α and β, and let $\widetilde{\mathcal{I}}$ be the set of improper collapses of any number of squares (possibly zero) in $V(\alpha) \setminus V_\alpha$ with distinct squares in $V(\beta) \setminus U_\beta$ in*

any gluing γ_i. Then there are coefficients c_γ for $\gamma \in \widetilde{\mathcal{I}}$ such that

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

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

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

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

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

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

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

vertices $V(T) \setminus V(R)$ must come from S, as must B_R; pick a subset of $V(R) \setminus B_R$ to include in $2^{|V(\alpha) \setminus V_\alpha|}$ ways. ∎

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

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

where we have defined,

Definition 5.4.24. *Let α be a left spider with end vertices $\boxed{i}, \boxed{j}$. Define* $\text{body}(\alpha)$ *as the shape whose graph is α with $\boxed{i}$ and $\boxed{j}$ deleted and with $U_{\text{body}(\alpha)} = U_\alpha \cup \{\textcircled{u}\} \setminus \{\boxed{i}, \boxed{j}\}, V_{\text{body}(\alpha)} = V_\alpha$. The definition is analogous for right spiders.*

Due to Lemma 5.4.21, the right-hand side of the approximation is in the null space of $\mathcal{M}$. We now formalize this approximate factorization.

Definition 5.4.25. *Let α be a spider with end vertices $\boxed{i}, \boxed{j}$. Define $\widetilde{\mathcal{I}}_\alpha$ to be the set of shapes that can be obtained from α by performing at least one of the following steps:*

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

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

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

Definition 5.4.26. *Let $\mathcal{I}_\alpha^{(1)}$ be the set of shapes that can be obtained from α by performing at least one of the following steps:*

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

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

Lemma 5.4.27. *Let α be a left spider with end vertices $\boxed{i}$, $\boxed{j}$. There are coefficients c_β for $\beta \in \widetilde{\mathcal{I}}_\alpha$ such that*

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

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

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

The shapes in $\widetilde{\mathcal{I}}_\alpha$ are definitionally the intersection terms that appear in this graph matrix product, and furthermore the shapes in $\widetilde{\mathcal{I}}_\alpha$ are definitionally the intersection terms for the ℓ_k term. Using Proposition 5.4.23, for each of the five shapes in $L_{|U_\alpha|}$ the coefficient it contributes is bounded by $4\,|V(\alpha)|^3$. The coefficient on ℓ_k is 2, so the coefficients for $\widetilde{\mathcal{I}}_\alpha^{(1)}$ are at most $8\,|V(\alpha)|^3$. The maximum coefficient of the other four shapes in $L_{|U_\alpha|}$ is $\frac{2}{n}$, so their total contribution to coefficients on $\widetilde{\mathcal{I}}_\alpha^{(2)}$ is at most $\frac{32|V(\alpha)|^3}{n}$. ∎

We now want to turn our improper shapes into proper ones from $\mathcal{I}_\alpha$. Unfortunately it is not quite true that to expand an improper shape, one can just expand each edge individually (though this is true for improper ribbons). There is an additional difficulty that arises due to ribbon symmetries. To see the difficulty, consider the example given in Fig. 5.8 below.

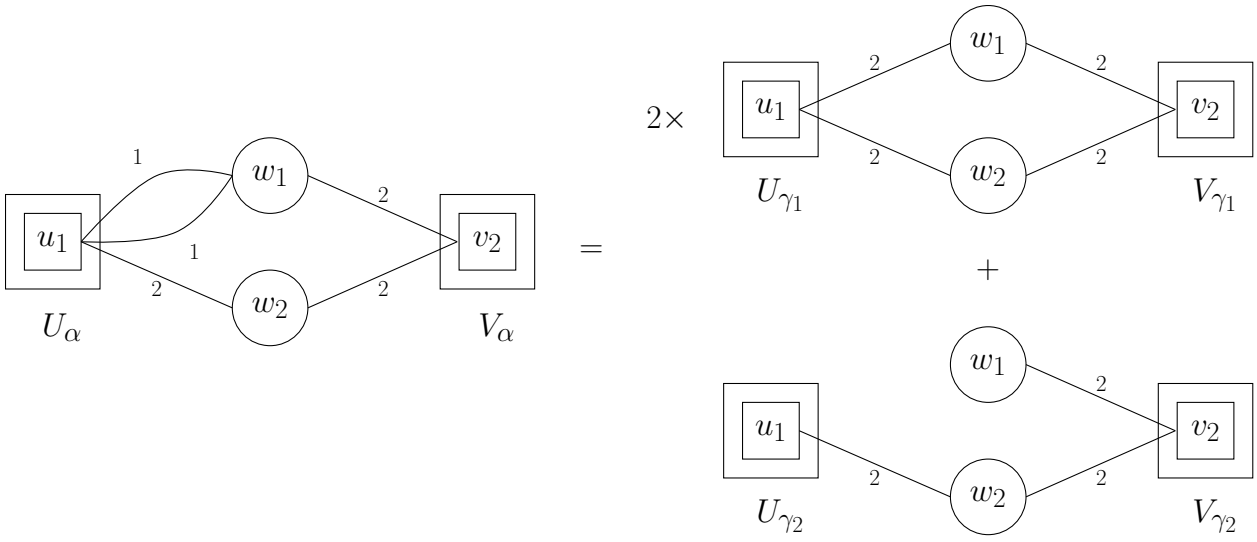

Figure 5.8: A surprising equality of graph matrices.

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

Definition 5.4.28. *An automorphism of a shape α is a function $\varphi : V(\alpha) \to V(\alpha)$ that preserves the sets U_α, V_α and is an automorphism of the underlying edge-labeled graph. Let* $\mathrm{Aut}(\alpha)$ *denote the automorphism group of α.*

Proposition 5.4.29. *Let α be an improper shape, and let $\mathcal{P}$ be the set of proper shapes that can be obtained by expanding α. Then there are coefficients $\left|c_\gamma\right| \leq C_{Fourier} \cdot C_{Aut}$ such that*

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

where $C_{Fourier}$ is a bound on the magnitude of Fourier coefficients in the expansion and

$C_{Aut} = \max_{\gamma \in \mathcal{P}} \frac{|\mathrm{Aut}(\gamma)|}{|\mathrm{Aut}(\alpha)|}$.

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

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

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

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

■

Proposition 5.4.30. *Let $l_1 \leq \cdots \leq l_k \in \mathbb{N}$ and let $L = l_1 + \cdots + l_k$. Assume $L \geq 1$. In the Fourier expansion of $h_{l_1}(z) \cdots h_{l_k}(z)$, the maximum coefficient is bounded in magnitude by $(2L)^{L - l_k}$.*

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

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

A formula from, e.g., [163, Example G (Continued)] shows that $\mathbb{E}[h_{l_1} \cdots h_{l_k} \cdot h_p]$ equals the number of "block perfect matchings": perfect matchings on $l_1 + \cdots + l_k + p$ elements divided

into blocks of size l_i or p such that no two elements from the same block are matched. Bound the number of block perfect matchings by:

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

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

Proposition 5.4.31. *For a shape α, let $\alpha \pm e$ denote the shape with edge e added or deleted. Then*

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

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

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

Lemma 5.4.32. *If α is a left spider, there are coefficients c_β for each $\beta \in \mathcal{I}_\alpha$ such that*

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

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

Proof. We express each $M_\beta, \beta \in \widetilde{\mathcal{I}}_\alpha$ in Lemma 5.4.27 in terms of proper shapes. We apply Proposition 5.4.29 using the following bounds on $C_{Fourier}$ and C_{Aut}. The only improperness in β comes from collapsing (at most) the two end vertices, which have a single incident

edge each. Therefore the set of labels of any parallel edges is either $\{1, k\}$ or $\{1, 1, k\}$, for some $k \leq |E(\alpha)|$. By Proposition 5.4.30, we have $C_{Fourier} \leq 4\,|E(\alpha)|^2$. There are at most two extra parallel edges in β, so we have $C_{Aut} \leq |V(\alpha)|^4$ using Proposition 5.4.31. Therefore the coefficients increase by at most $C_{Fourier} \cdot C_{Aut} \leq 4\,|E(\alpha)|^2\,|V(\alpha)|^4$. ∎

Corollary 5.4.33. *If α is a right spider, there are coefficients c_β with the same bounds given in Lemma 5.4.32 such that*

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

Corollary 5.4.34. *If $x \perp \mathrm{Null}(\mathcal{M}_{fix})$ and α is a spider, then for some c_β with the same bounds given in Lemma 5.4.32,*

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

Proof. For a left spider, since

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

we are in position to use Fact 5.2.1. For a right spider, the proof is analogous. ∎

5.4.3 Killing all the spiders

The strategy is to start with the moment matrix $\mathcal{M}$ and apply Corollary 5.4.34 repeatedly until we end up with no spiders in our decomposition. For each spider, killing it via Corollary 5.4.34 leaves only intersection terms. Some of those intersection terms may themselves be smaller spiders, in which case we will apply the corollary again and again until only non-spiders remain. The difficulty during this procedure is to bound the total coefficient

accumulated on each non-spider. To capture this process, we define the web of a spider α, which will be a directed acyclic graph that will capture the spider killing process. For the sake of distinction, we will call the vertices of this graph "nodes".

Definition 5.4.35 (Web of α)**.** *The web $W(\alpha)$ of a spider α is a rooted directed acyclic graph (DAG) whose nodes are shapes and whose root is α. Each spider node γ has edges to nodes β for each shape $\beta \in \mathcal{I}_\gamma$. The non-spider nodes are leaves/sinks of the DAG.*

Remark 5.4.36. *The DAG structure arises because each shape in $\mathcal{I}_\gamma$ has strictly fewer square vertices than γ for any spider γ. As a consequence, the height of a web $W(\alpha)$ is at most $|V(\alpha)|$.*

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

- Initially, set $v_\alpha = 1$ and for all other γ, set $v_\gamma = 0$.
- Starting from the root and in topological order, each spider node γ adds $v_\gamma c_\beta$ to v_β for each child $\beta \in \mathcal{I}_\gamma$, where the c_β are the coefficients from Corollary 5.4.34.

Proposition 5.4.37. *If $x \perp \mathrm{Null}(\mathcal{M}_{fix})$, then*

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

Proof. Start with the equation $x^\intercal M_\alpha x = x^\intercal v_\alpha M_\alpha x$. In each step, we take the topologically first spider γ, which in this case means the spider closest to the root of $W(\alpha)$, that is present in the right hand side of our equation and using Corollary 5.4.34, we replace $v_\gamma M_\gamma$ by $\sum_{\beta \in \mathrm{children}(\gamma)} v_\gamma c_\beta M_\beta$. Precisely by the definition of the v_γ, this process ends with the equation

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

■

Proposition 5.4.38. *For any node β in $W(\alpha)$, $|\text{parents}(\beta)| \leq 4\,|V(\alpha)|^3 \cdot |E(\alpha)|^2$ where parents(β) is the set of nodes γ in $W(\alpha)$ such that $\beta \in \mathcal{I}_\gamma$.*

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

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

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

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

Proposition 5.4.39. *Let p be a path in $W(\alpha)$ with $\#_1(p)$ type 1 edges and $\#_2(p)$ type 2 edges. Then $\#_1(p) \leq |E(\alpha)| + 2\#_2(p)$.*

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

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

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

Now we bound this expression for α. From the definition of $\mathcal{L}$, Definition 5.3.6, for spiders appearing in the pseudocalibration, the square vertices in W_α, $U_\alpha \setminus (U_\alpha \cap V_\alpha)$ and $V_\alpha \setminus (U_\alpha \cap V_\alpha)$ have degree at least 1 and can only be connected to circle vertices. Therefore their number is bounded by $|E(\alpha)|$. Hence, initially $|\mathcal{S}_\alpha \cap W_\alpha| + |U_\alpha \setminus (U_\alpha \cap V_\alpha)| + |V_\alpha \setminus (U_\alpha \cap V_\alpha)| \leq |E(\alpha)|$.

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

Corollary 5.4.40. $\#_2(p) \geq \frac{|p|}{3} - \frac{|E(\alpha)|}{3}$.

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

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

Lemma 5.4.41. *There are absolute constants C_1, C_2 so that for all leaves γ of $W(\alpha)$,*

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

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

Remark 5.4.42. *Here it is important that type 2 edges have stronger bounds on their coefficients $\left|c_\beta\right| \leq C \cdot (|V(\alpha)|\,|E(\alpha)|)^{O(1)}/n \ll 1$.*

Before we proceed with the proof we establish some convenient notation and recall some facts. For consecutive shapes β_{i-1}, β_i (i.e., β_i is a child of β_{i-1}), we denote by c_{β_i} the coefficient from Corollary 5.4.34 applied on β_{i-1}. By Proposition 5.4.38, the in-degree of $W(\alpha)$ can be bounded as $B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}$ for some constants B_1, B_2. Thus, the number of paths of length r ending at γ is at most $(B_1\,|V(\alpha)|\,|E(\alpha)|)^{B_2 r}$. Using Corollary 5.4.34, set B_1, B_2 large enough so that c_{β_i} is at most $B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}$ for a type 1 edge (resp. $B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}/n$ for a type 2 edge).

$$
\begin{aligned}
|v_\gamma| &\le \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \prod_{i=1}^{r} |c_{\beta_i}| \\
&\le \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{\#_1(p)} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}/n\right)^{\#_2(p)} && \text{(Corollary 5.4.34)} \\
&\le \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{|E(\alpha)|+2\#_2(p)} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}/n\right)^{\#_2(p)} && \text{(Proposition 5.4.39)} \\
&= \sum_{r=0}^{\infty} \sum_{\substack{p=(\beta_0=\alpha,\ldots,\beta_r=\gamma)\\ \text{path from } \alpha \text{ to } \gamma \text{ in } W(\alpha)}} \left(B_1 \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2}\right)^{|E(\alpha)|} \left(B_1' \cdot (|V(\alpha)|\,|E(\alpha)|)^{B_2'}/n\right)^{\#_2(p)}
\end{aligned}
$$

for some constants B_1', B_2'. We split the above sum into two sums, $r \le 3|E(\alpha)|$ and $r > 3|E(\alpha)|$. For $r \le 3\,|E(\alpha)|$, upper bounding the $\#_2(p)$ term by 1 and upper bounding the number of paths by $(B_1\,|V(\alpha)|\,|E(\alpha)|)^{B_2 r}$ gives a bound of $(B_1''\,|V(\alpha)|\,|E(\alpha)|)^{B_2''|E(\alpha)|}$ for some constants B_1'', B_2''. For larger r, we lower bound $\#_2(p) \ge r/9 = |E(\alpha)|\,/3$ using Corollary 5.4.40. Applying the same bound on the number of paths, the total contribution of the terms corresponding to larger r is bounded by 1 using the power of n in the denominator (assuming δ, τ are small enough). ∎

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

Definition 5.4.43. *Define the matrix $\mathcal{M}^+$ as the result of killing all the spiders,*

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

5.4.4 Finishing the proof

The final step of the proof is to argue that, after the spider killing process is completed, the newly created non-spider terms in $\mathcal{M}^+$ also have small norm. Towards this, we would like to prove a statement similar to Corollary 5.4.11. In that proof, we used special structural properties of the non-spiders in $\mathcal{L}$ to prove that non-spiders in the pseudocalibration were negligible. But now, the non-spiders in $\mathcal{M}^+$ need not have the properties of $\mathcal{L}$ – for instance, there could be circle vertices of degree 2 or isolated vertices. To handle the potentially larger norms, we will use that the coefficients of these new non-spider terms β come with the coefficients λ_α of the spider terms α in whose web they lie. Since α has more vertices/edges than β, the power of $\frac{1}{n}$ in λ_α is larger than the "expected pseudocalibration" coefficient of $\eta^{|U_\beta|+|V_\beta|} \cdot \frac{1}{n^{|E(\beta)|/2}}$. We prove that these extra factors of $\frac{1}{n}$ are enough to overpower isolated vertices or a smaller vertex separator using a careful charging argument.

Lemma 5.4.44. *If β is a nontrivial non-spider and $\beta \in W(\alpha)$ for some spider $\alpha \in \mathcal{L}$, then*

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

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

Proof. We start by giving the idea of the proof. Suppose we try to use the same distribution scheme as in the proof of Lemma 5.4.7. It doesn't work for two reasons. Firstly, the circle vertices in β still have even degree, which follows from Remark 5.4.18, but now, they could have degrees 0 or 2. For the previous distribution scheme to go through, we needed them

to have degree at least 4 which gave the necessary edge decay to handle the norm bounds. Secondly, the square vertices can now have degree 0 hence getting no decay from the edges.

The first issue is relatively easy to handle. Since β was obtained by collapsing α, the circle vertices of degrees 0 or 2 in β must have had degree at least 4 in α to begin with. Hence, we can fix a particular sequence of collapses from α to β and then assume for the sake of analysis that the removed edges are still present. In this case, the same charging argument as in Lemma 5.4.7 would go through. This is made formal by looking at the sequence of improper collapses of this chain of collapses.

To handle the second issue, let's analyze more carefully how degree 0 square vertices appear. Fix a sequence of collapses from α to β and consider a specific step where γ collapsed to γ' and a square vertex of degree 0 was formed. Let the two square vertices that collapsed in γ be $\boxed{i}, \boxed{j}$ and let the square vertex of degree 0 that formed in γ' be $\boxed{k}$. In light of Remark 5.4.18, since $\boxed{k}$ has degree 0, it must not be in $(U_{\gamma'} \cup V_{\gamma'}) \setminus (U_{\gamma'} \cap V_{\gamma'})$ and hence, $U_{\gamma'}(\boxed{k}) = V_{\gamma'}(\boxed{k}) = 0$ or $U_{\gamma'}(\boxed{k}) = V_{\gamma'}(\boxed{k}) = 1$. But in the latter case, this vertex does not contribute to norm bounds since it's in $U_{\gamma'} \cap V_{\gamma'}$ so it can be safely disregarded. Note that it doesn't have to stay in this set since future collapses might collapse this vertex, but this is not a problem as we can charge for this collapse if it happens.

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

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

square vertex of degree 0.

We now make these ideas formal.

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

Because of parity constraints, Remark 5.4.18, and because there are no circle vertices in $U_\beta \cup V_\beta$, all circle vertices have even degree in β. Let c_0 be the number of degree 0 circle vertices in β. Let $c_2, c_{\geq 4}$ be the number of degree 2 circle vertices and the number of circle vertices of degree at least 4 in $V(\beta) \setminus S_{min}$ respectively. Then, we have

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

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

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

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

Lemma 5.4.45. $|U_\alpha| + |V_\alpha| - |U_\beta| - |V_\beta| \geq 2s_0$

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

■

We will now prove a structural lemma.

Lemma 5.4.46. *Any vertex* $\textcircled{u}$ *that has degree at least* 2 *in* $V(\beta) \setminus S_{min}$ *is adjacent to at most* 1 *vertex of* P'.

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

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

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

Lemma 5.4.47. $|E(\alpha)| \geq |P'| + s_{\geq 2} + (1.5 - \varepsilon)(c_2 + c_{\geq 4}) + 2(1.5 - \varepsilon)c_0 + \Omega(\varepsilon\,|E(\alpha)|)$

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

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

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

Suppose the degree of $\textcircled{u}$ in $V(\beta)$ is 2. Then, it is incident on at least 2 phantom edges. By Lemma 5.4.46, it is also adjacent to at most one vertex of P' and so, must have been allocated a value of at least $2(1-\frac{\varepsilon}{10})+(deg_{\widetilde{\beta}}(\textcircled{u})-3)(\frac{1}{2}-\frac{\varepsilon}{10})$. This is at least $1.5-\varepsilon+\frac{\varepsilon}{10}$.

Suppose the degree of $\textcircled{u}$ in $V(\beta)$ is at least 4. By Lemma 5.4.46, it is adjacent to at most one vertex of P'. Then it must have been allocated a value of at least $(deg_{\widetilde{\beta}}(\textcircled{u})-1)(\frac{1}{2}-\frac{\varepsilon}{10})$. Using $deg_{\widetilde{\beta}}(\textcircled{u})\geq 4$, this is at least $1.5-\varepsilon+\frac{\varepsilon}{10}$.

This implies

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

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

Adding Lemma 5.4.45 and Lemma 5.4.47, we get the result. ∎

Corollary 5.4.48. *If β is a nontrivial non-spider and $\beta\in W(\alpha)$ for some spider $\alpha\in\mathcal{L}$, then*

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

Proof. From Lemma 5.6.3, we have

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

We have $|V(\beta)|\cdot(1+|E(\beta)|)\cdot\log(n)\leq n^{O(\tau)}$. Also, $|V_{rel}(\beta)|\leq 2(|E(\alpha)|+|E(\beta)|)$ since all the degree 0 vertices in $V_{rel}(\beta)$ would have had vertices of $V_{rel}(\alpha)$ collapse into it in the chain of collapses and there are no degree 0 vertices in $V_{rel}(\alpha)$. Finally, since $|E(\alpha)|\geq|E(\beta)|$, the factor $2\cdot(|V(\beta)|\cdot(1+|E(\beta)|)\cdot\log(n))^{C\cdot(|V_{rel}(\beta)|+|E(\beta)|)}$ can be absorbed into $\frac{1}{n^{\Omega(\varepsilon|E(\alpha)|)}}$. The result follows from Lemma 5.4.44. ∎

Proposition 5.4.49. *If β is a trivial shape, $\lambda_\beta^+=\lambda_\beta$.*

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

Lemma 5.4.50. *For $k, l \in \{0, 1, \ldots, D/2\}$, let $\mathcal{B}_{k,l}$ denote the set of nontrivial non-spiders on block (k, l). Then*

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

Proof.

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

To bound the first term, we checked previously in Corollary 5.4.14 that the total norm of nontrivial non-spiders appearing in the pseudocalibration (i.e. this term) is $\eta^{k+l} o_n(1)$. For the second term, via Lemma 5.4.41 we have a bound on the accumulations v_γ of one spider on one non-spider, so it is at most

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

Use the bound on the coefficients $|\lambda_\alpha|$, Proposition 5.4.13,

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

Invoking the norm bound for non-spiders which are collapses, Corollary 5.4.48,

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

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

Bound the sum over all spiders by the sum over all shapes. By Proposition 5.4.12, the number of shapes with i edges is $n^{O(\tau(i+1))}$. Summing by the number of edges, observe that $|E(\alpha)| \geq \max(|E(\beta)|, 2)$ since spiders always have at least 2 edges.

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

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

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

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

Corollary 5.4.51. *For $k \in \{0, \ldots, D/2\}$, the (k,k) block of $\mathcal{M}^+$ has minimum singular value at least $\eta^{2k}(1 - \frac{1}{n^{\Omega(\varepsilon)}})$, and for $k, l \in \{0, \ldots, D/2\}, l \neq k$, the (k,l) off-diagonal block has norm at most $\eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$.*

Proof. By Proposition 5.4.49 the identity matrix appears on the (k,k) blocks with coefficient η^{2k}. By construction, $\mathcal{M}^+$ has no spider shapes. By Lemma 5.4.50, the total norm of the non-spider shapes on the (k,l) block is at most $\eta^{k+l} \cdot \frac{1}{n^{\Omega(\varepsilon)}}$. $\blacksquare$

Theorem 5.4.52. *W.h.p. $\mathcal{M}_{fix} \succeq 0$.*

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

with $\|x\|_2 = 1$,

$$\begin{aligned} x^\intercal \mathcal{M}_{fix} x &= x^\intercal (\mathcal{M} + \mathcal{E}) x \\ &= x^\intercal \mathcal{M}^+ x + x^\intercal \left(\sum_{\text{spiders } \alpha} \lambda_\alpha \left(\mathcal{M}_\alpha - \sum_{\text{leaves } \gamma \text{ of } W(\alpha)} v_\gamma M_\gamma \right) \right) x \\ &\quad + x^\intercal \mathcal{E} x \\ &= x^\intercal (\mathcal{M}^+ + \mathcal{E}) x \qquad \text{(Proposition 5.4.37)} \end{aligned}$$

Because the norm bound on $\mathcal{E}$ is significantly less than $\eta^D = n^{-n^\delta}$ (see [70]), the bound on the norm of each block of $\mathcal{M}^+$ in Corollary 5.4.51 also applies to the blocks of $\mathcal{M}^+ + \mathcal{E}$. Therefore, we use Lemma 5.4.2 to conclude $\mathcal{M}^+ + \mathcal{E} \succeq 0$ and the above expression is nonnegative. ∎

5.5 Sherrington-Kirkpatrick Lower Bounds

Here, we prove Theorem 4.1.5 and Theorem 4.1.2.

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

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

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

Theorem 4.1.5. *There exists a constant $c > 0$ such that, for all $\varepsilon > 0$ and $\delta \le c\varepsilon$, for $p \ge n^{2/3+\varepsilon}$, w.h.p. over V there is a degree-n^δ SoS solution for Planted Boolean Vector of value* 1.

Proof. We wish to produce an SoS solution $\widetilde{\mathbb{E}}$ on boolean variables $b_1, \dots, b_n$ such that $\widetilde{\mathbb{E}}[b^\intercal \Pi_V b] = n$. Instead of sampling a uniformly random p-dimensional subspace V of $\mathbb{R}^n$, we first sample $d_1, \dots, d_n$ i.i.d. p-dimensional Gaussian vectors from $\mathcal{N}(0, I)$, then form an

n-by-p matrix A with rows $d_1, \ldots, d_n$, and finally take V to be the span of the columns of A. Since the columns of A are isotropic i.i.d. random Gaussian vectors, we have that V is a uniform p-dimensional subspace[5] of $\mathbb{R}^n$.

We will consider V as the input for the Planted Boolean Vector problem while the vectors $d_1, \ldots, d_n$ will be used to construct a pseudoexpectation operator for the Planted Affine Planes problem[6]. Since $n \leq p^{3/2-\Omega(\varepsilon)}$, by Theorem 4.1.4, for all $\delta \leq c\varepsilon$ for a constant $c > 0$, w.h.p., there exists a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}'$ on formal variables $v = (v_1, \ldots, v_p)$ such that $\widetilde{\mathbb{E}}'[\langle v, d_u\rangle^2] = 1$ for every $u \in [n]$.

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

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

■

Now we prove lower bounds for the Sherrington-Kirkpatrick problem, using a reduction and proof due to [129]. We include it here for completeness. Recall that the SK problem is to compute

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

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

Theorem 4.1.2. *There exists a constant $\delta > 0$ such that, w.h.p. for $W \sim \mathrm{GOE}(n)$, there is a degree-n^δ SoS solution for the Sherrington–Kirkpatrick problem with value at least $(2 - o_n(1)) \cdot n^{3/2}$.*

5. Except for a zero measure event.

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

We will use the following standard results from random matrix theory of GOE(n).

Fact 5.5.1. *Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of $W \sim \mathrm{GOE}(n)$ with corresponding normalized eigenvectors $w_1, \ldots, w_n$. Then,*

1. *For every $p \in [n]$, the span of $w_1, \ldots, w_p$ is a uniformly random p-dimensional subspace of $\mathbb{R}^n$ (see e.g. [138, Section 2]).*
2. *W.h.p., $\lambda_{n^{0.67}} \geq (2 - \mathrm{o}(1))\sqrt{n}$ (Corollary of Wigner's semicircle law [185])*

Proof of Theorem 4.1.2: Let $p = n^{0.67}$ and $W \sim \mathrm{GOE}(n)$. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of W with corresponding orthonormal set of eigenvectors $w_1, \ldots, w_n$. By Fact 5.5.1, we have that $\lambda_p \geq (2 - \mathrm{o}(1))\sqrt{n}$ and that $w_1, \ldots, w_p$ span a uniformly random p-dimensional subspace V of $\mathbb{R}^n$.

We consider V as the input of the Boolean Planted Vector problem and by Theorem 4.1.5, for some constant $\delta > 0$, w.h.p. there exists a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}$ such that $\widetilde{\mathbb{E}}[x_i^2] = 1$ and $\widetilde{\mathbb{E}}[\sum_{i=1}^p \langle x, w_i\rangle^2] = \widetilde{\mathbb{E}}[x^\intercal \Pi_V x] = n$. Now,

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

■

Remark 5.5.2. *Using the same proof as above, we can obtain Theorem 4.1.2 even if we were only able to prove SoS lower bounds for Planted Affine Planes for some $m = \omega(n)$.*

So, pushing the value of m up to $n^{3/2-\varepsilon}$, which is Theorem 4.1.4, offers only a modest improvement.

5.6 Omitted technical details

5.6.1 Norm Bounds

The precise norm bounds we use come from applying the trace power method in [2], but qualitatively, the bounds from Chapter 2 also work. The paper [2] uses a slightly different definition of matrix index. They define a *matrix index piece* as a tuple of distinct elements from either $\mathcal{C}_m$ or $\mathcal{S}_n$ along with a fixed integer denoting multiplicity. A matrix index is then a set of matrix index pieces. Our graph matrix M_α appears as a submatrix of those matrices: for a given set of square vertices, order the squares in increasing order in a tuple, and assign it multiplicity 1. Hence the same norm bounds apply.

Boolean norm bounds:

Lemma 5.6.1. *Let $V_{rel}(\alpha) := V(\alpha) \setminus (U_\alpha \cap V_\alpha)$. There is a universal constant C such that the following norm bound holds for all proper shapes α w.h.p.:*

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

Proof. From Corollary 8.13 of [2], with probability at least $1 - \varepsilon$ for a fixed shape α,

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

Letting N_k be the number of distinct shapes on k vertices (either circles or squares), we apply the corollary with $\varepsilon = 1/(mnN_{|V(\alpha)|})$. Union bounding, the failure probability across all shapes of size k is at most $1/mn$, and since the number of vertices in a shape is at most

$m+n \le 2m$, we have a bound that holds with high probability for all shapes. It remains to simplify the exact bound.

Proposition 5.6.2. $N_k \le 8^k 2^{k^2}$

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

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

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

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

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

Lemma 5.6.3. *Let $V_{rel}(\alpha) := V(\alpha) \setminus (U_\alpha \cap V_\alpha)$ as sets. There is a universal constant C such that the following norm bound holds for all proper shapes α with total size at most n w.h.p.:*

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

The proof performs the same calculation starting from [2, Corollary 8.15]. Note that in our notation, $l(\alpha) = |E(\alpha)|$. There is a further difference which is that [2] uses normalized Hermite polynomials whereas we use unnormalized Hermite polynomials; this contributes

the additional term $\prod_{e \in E(\alpha)} l(e)! \leq (1 + |E(\alpha)|)^{|E(\alpha)|}$. We must replace Proposition 5.6.2 with the following:

Proposition 5.6.4. *The number of Hermite shapes with total size k is at most $k2^k(k+1)^{2k+k^2}$.*

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

5.6.2 Properties of $e(k)$

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

Claim 5.6.5. $e(2) = 0$.

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

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

concluding the proof. ∎

Definition 5.6.6. *We say that a tuple $\lambda = (\lambda_1, \dots, \lambda_k)$ of non-negative integers is a partition of k provided $\sum_{i=1}^k \lambda_i = k$ and $\lambda_1 \geq \cdots \geq \lambda_k$. We use the notation $\lambda \vdash k$ to denote a partition of k. We refer to λ_i as a row/part of λ.*

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

Definition 5.6.7. *The transpose of partition $\lambda = (\lambda_1, \dots, \lambda_k)$ is denoted λ^t and defined as $\lambda_i^t = \big|\{j \in [k] \mid \lambda_j \geq i\}\big|$.*

Remark 5.6.8. *For a partition $\lambda \vdash k$, λ_1^t is the number of rows/parts of λ.*

Definition 5.6.9. *The automorphism group of a partition $\mathrm{Aut}(\lambda) \leq S_{\lambda_1^t}$ is the group generated by transpositions (i,j) of rows $\lambda_i = \lambda_j$.*

Remark 5.6.10. *Let $\lambda \vdash k$ and $p_1(\lambda), \ldots, p_k(\lambda)$ be such that $p_i(\lambda) = \left|\{j \in [\lambda_1^t] \mid \lambda_j = i\}\right|$. Then $\mathrm{Aut}(\lambda) \simeq S_{p_1} \times \cdots \times S_{p_k}$.*

Lemma 5.6.11. *We have*

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

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

Claim 5.6.12. *Let $\lambda \vdash k$. We have*

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

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

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

2. Case $\lambda_1 = 2$: Let λ' be the partition obtained from λ by removing λ_1. Note that

$\lambda_1^t = (\lambda')_1^t + 1 \le k-2$. By the induction hypothesis, we have

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

3. Case $\lambda_1 = 1$: To bound $(n)_k \cdot \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right]$, we use Lemma 5.6.11 and the two preceding cases. Let $p(k)$ be the partition function, i.e., $p(k) = |\{\lambda \vdash k\}|$. We deduce that

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

as desired.

■

Claim 5.6.13. *Suppose $k < \sqrt{n}/2$. We have*

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

Proof. Follows from Claim 5.6.12 and the bound on k. ■

Remark 5.6.14. *In Claim 5.6.13, the factor 3^{k^3} is too lossy to allow a meaningful bound with $k = n^{\varepsilon}$, where $\varepsilon > 0$ is a constant.*

Refining the ideas of Claim 5.6.12, we prove a stronger lemma below which will imply a tighter bound on $e(k)$ sufficient for our application.

Lemma 5.6.15. *There exists an universal constant $C \geq 1$ such that*

$$\sum_{\lambda \vdash k} \frac{\lambda!}{\lambda_1! \cdots \lambda_k!} \cdot \frac{(n)_{\lambda_1^t}}{|\mathrm{Aut}(\lambda)|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_k^{\lambda_k} \right] \right| \leq k^{C \cdot k} \cdot n^{k/2}. \tag{5.1}$$

In particular, for $n \geq 6$, Eq. (5.1) holds with $C = 2$.

Proof. We induct on k. For $k = 1$, we have $n \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} x_1 \right| \leq \sqrt{n}$ as desired. Using $e(2) = 0$ from Claim 5.6.5 and the case $k = 1$ of Eq. (5.1), we get that Lemma 5.6.15 also holds for $k = 2$. Now, consider $k \geq 3$. Let $\Lambda_1 = \{\lambda \vdash k \mid \lambda_1 = 1\}$, $\Lambda_2 = \{\lambda \vdash k \mid \lambda_1 = 2\}$ and $\Lambda_{\geq 3} = \{\lambda \vdash k \mid \lambda_1 = 3\}$. Note that $\Lambda_1 \sqcup \Lambda_2 \sqcup \Lambda_{\geq 3} = \{\lambda \vdash k\}$ and $|\Lambda_1| = 1$.

For convenience define a_λ to be the term associated to $\lambda \vdash k$ on the LHS of Eq. (5.1), i.e.,

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

First we bound the contribution of the terms associated to partitions from $\Lambda_{\geq 3}$ in the LHS of Eq. (5.1). Let λ' be the partition obtained from λ by removing two boxes from λ_1. Note that $\lambda_1^t = (\lambda')_1^t \leq k - 2$ and $\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda'_1} \dots x_{k-2}^{\lambda'_{k-2}} \right] = \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1^{\lambda_1} \dots x_{k-2}^{\lambda_{k-2}} \right]$. Thus,

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

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

$$\sum_{\lambda \in \Lambda_{\geq 3}} a_\lambda \leq k \cdot \sum_{\lambda' \vdash k-2} k^3 \cdot a_{\lambda'} \leq k^4 \cdot k^{C \cdot (k-2)} \cdot n^{(k-2)/2}, \tag{5.2}$$

where the last equality follows from the induction hypothesis.

Now we bound the contribution of the terms a_λ associated to partitions λ from Λ_2 in the LHS of Eq. (5.1). Let $i \geq 1$ be the number of parts of size two of λ and let λ' be the partition obtained from λ by removing these i parts of size two. Note that $\lambda_1^t = (\lambda')_1^t + i \leq k-1$. We have

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

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

$$\begin{aligned} \sum_{\lambda \in \Lambda_2} a_\lambda &\leq \sum_{i=1}^{\lfloor k/2 \rfloor} n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot \left(\frac{\lambda'!}{\lambda_1'! \cdots \lambda_k'!} \cdot \frac{(n)_{(\lambda')_1^t}}{|\mathrm{Aut}(\lambda')|} \cdot \left| \mathbb{E}_{x \in \mathcal{S}(\sqrt{n})} \left[x_1 \dots x_{k-2i} \right] \right| \right) \\ &\leq \sum_{i=1}^{\lfloor k/2 \rfloor} n^i \cdot \frac{(k)_i}{2^i} \cdot \frac{1}{i!} \cdot k^{C \cdot (k-2i)} \cdot n^{(k-2i)/2} \\ &\leq k^{C \cdot (k-1)} \cdot n^{k/2} \cdot \sum_{i=0}^{\infty} k^{-C \cdot i} \leq \frac{3}{2} \cdot k^{C \cdot (k-1)} \cdot n^{k/2}, \end{aligned}$$

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

Finally, we consider the case $\lambda_1 = 1$. To bound a_λ, we use Lemma 5.6.11 and the two

preceding cases. We deduce that

$$a_\lambda \le n^{k/2} + \sum_{\mu\in\Lambda_2} a_\mu + \sum_{\mu\in\Lambda_{\ge 3}} a_\mu \le n^{k/2} + k^4 \cdot k^{C\cdot(k-2)} \cdot n^{(k-2)/2} + \frac{3}{2}\cdot k^{C\cdot(k-1)}\cdot n^{k/2}$$
$$= k^{C\cdot k}\cdot n^{k/2}\left(\frac{1}{k^{C\cdot k}} + \frac{k^4}{n\cdot k^{2\cdot C}} + \frac{3}{2\cdot k^C}\right).$$

We can bound the LHS of Eq. (5.1) as

$$\sum_{\mu\in\Lambda_1} a_\mu + \sum_{\mu\in\Lambda_2} a_\mu + \sum_{\mu\in\Lambda_{\ge 3}} a_\mu \le k^{C\cdot k}\cdot n^{k/2}\left(\frac{1}{k^{C\cdot k}} + \frac{2\cdot k^4}{n\cdot k^{2\cdot C}} + \frac{3}{k^C}\right)$$
$$\le k^{C\cdot k}\cdot n^{k/2},$$

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

Corollary 5.6.16. *We have*

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

Proof. Suppose $k \le \sqrt{n}$. Note that Lemma 5.6.15 implies that for $\lambda \vdash k$ with λ_1 there exists a constant $C > 0$ such that

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

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

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

Furthermore, for $n \geq 6$, Lemma 5.6.15 allows us to choose $C = 2$. Since $\left|\mathbb{E}_{x \in \mathcal{S}(\sqrt{n})}[x_1]\right| = 1/\sqrt{n}$, the simpler bound applies for all values of k

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

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

CHAPTER 6

THE MACHINERY AND QUALITATIVE BOUNDS

In this chapter, we state the main machinery that we use to prove our results. The complete proof of the machinery can be found in [149]. After that, we exhibit the qualitiative bounds for applying the machinery to our problems of interest. The material in this chapter is adapted from [149].

6.1 Statement of the machinery

6.2 Qualitative bounds for Planted slightly denser subgraph

6.2.1 Pseudo-calibration

We will pseudo-calibrate with respect to the following pair of random and planted distributions which we denote ν and μ 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 $\frac{k}{n}$. 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.

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 χ_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 6.2.1. *Let $I \subseteq [n], E \subseteq \binom{[n]}{2}$. Then,*

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

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

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

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 $\frac{1}{2}$, 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 $Pr_{(G,S)\sim\mu}[I \cup V(E) \subseteq S] = \left(\frac{k}{n}\right)^{|I \cup V(E)|}$ and

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

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

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

Definition 6.2.2. *Define the degree of SoS to be $D_{sos} = n^{C_{sos}\varepsilon}$ for some constant $C_{sos} > 0$ that we choose later.*

Definition 6.2.3 (Truncation parameter)**.** *Define the truncation parameter to be $D_V = n^{C_V \varepsilon}$ for some constant $C_V > 0$.*

Remark 6.2.4 (Choice of parameters)**.** *We first set ε 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 Λ.

Definition 6.2.5. *If a shape α satisfies the following properties:*

- *α is proper,*
- *α satisfies the truncation parameter D_{sos}, D_V.*

then define

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

Corollary 6.2.6. $\Lambda = \sum \lambda_\alpha M_\alpha$.

6.2.2 Proving positivity - Qualitative bounds

We use the canonical definition of H'_γ from **??**. In this section, we will prove the main qualitative bounds Lemma 6.2.7, Lemma 6.2.9 and Lemma 6.2.11.

Lemma 6.2.7. *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 τ to the Fourier coefficients.

Definition 6.2.8. *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 6.2.9. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

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

We define the following quantity to capture the contribution of the vertices within γ to the Fourier coefficients.

Definition 6.2.10. *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 6.2.11. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

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

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

Definition 6.2.12. *For a shape $\sigma \in \mathcal{L}$, define*

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

Definition 6.2.13. *For $U \in \mathcal{I}_{mid}$, 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 6.2.14. *For all $U \in \mathcal{I}_{mid}, \rho \in \mathcal{P}_U$, $H_{Id_U} = \frac{1}{|Aut(U)|} v_U v_U^T$.*

Proof. This follows by verifying the conditions of Definition 6.2.5. ∎

This immediately implies that for all $U \in \mathcal{I}_{mid}$, $H_{Id_U} \succeq 0$, which is Lemma 6.2.7.

Proposition 6.2.15. *For any $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, $H_\tau = \frac{1}{|Aut(U)|^2} S(\tau) v_U v_U^T$.*

Proof. This follows by a straightforward verification of the conditions of Definition 6.2.5. ∎

We can now prove Lemma 6.2.9 and Lemma 6.2.11.

Proof of Lemma 6.2.9.

$$\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} = \begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} v_U v_U^T & \frac{S(\tau)}{|Aut(U)|^2} v_U v_U^T \\ \frac{S(\tau)}{|Aut(U)|^2} v_U v_U^T & \frac{S(\tau)}{|Aut(U)|} v_U v_U^T \end{bmatrix} \succeq 0$$

∎

Proof of Lemma 6.2.11. Fix $\sigma, \sigma' \in \mathcal{L}_U$ such that $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$. Note that $|V(\sigma)| - \frac{|V_\sigma|}{2} + |V(\sigma')| - \frac{|V_{\sigma'}|}{2} + 2(|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}) = |V(\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T)|$. Using Definition 6.2.5, we can easily verify that $\lambda_{\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T} = T(\sigma) T(\sigma') S(\gamma)^2$. Therefore, $H_{Id_V}^{-\gamma,\gamma}(\sigma, \sigma') = \frac{|Aut(U)|}{|Aut(V)|} S(\gamma)^2 H_{Id_U}(\sigma, \sigma')$. Since $H'_\gamma(\sigma, \sigma') = H_{Id_U}(\sigma, \sigma')$ whenever $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$, this completes the proof. ∎

6.3 Qualitative bounds for Tensor PCA

6.3.1 Pseudo-calibration

Definition 6.3.1 (Slack parameter)**.** *Define the slack parameter to be $\Delta = n^{-C_\Delta \varepsilon}$ for a constant $C_\Delta > 0$.*

- Random distribution: Sample A from $\mathcal{N}(0, I_{[n]^k})$.
- 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 $\mathcal{N}(0, I_{[n]^k})$. 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 $h_a(A) := \prod_{e \in [n]^k} h_e(A_e)$. We will work with this Hermite basis.

Lemma 6.3.2. *Let $I \in \mathbb{N}^n, a \in \mathbb{N}^{[n]^k}$. For $i \in [n]$, let $d_i = \sum_{i \in e \in [n]^k} a_e$. Let c be the number of i such that $I_i + d_i$ is nonzero. Then, if $I_i + d_i$ are all even, we have*

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

Else, $\mathbb{E}_{\mu}[u^I h_a(v)] = 0$.

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

$$\mathop{\mathbb{E}}_{(u_i, w_e) \sim \mu}[u^I h_a(A)] = \mathop{\mathbb{E}}_{(u_i) \sim \mu}[u^I \prod_{e \in [n]^k} (\lambda \prod_{i \leq k} u_{e_i})^{a_e}] = \mathop{\mathbb{E}}_{(u_i) \sim \mu}[\prod_{i \in [n]} u_i^{I_i + d_i}] \prod_{e \in [n]^k} \lambda^{a_e}$$

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

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

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

Definition 6.3.4 (Truncation parameters)**.** *Define the truncation parameters to be $D_V = n^{C_V \varepsilon}, D_E = n^{C_E \varepsilon}$ for some constants $C_V, C_E > 0$.*

Remark 6.3.5 (Choice of parameters)**.** *We first set ε to be a sufficiently small constant. Based on the choice of ε, 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_{sos} to be sufficiently small to satisfy the inequalities we use.*

Remark 6.3.6. *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 6.3.7. *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 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 that U corresponds to.*
- *For a shape α and vertex i in α, let $deg^\alpha(i) = \sum_{i \in e \in E(\alpha)} l_e$.*
- *For any shape α, let $deg(\alpha) = deg(U_\alpha) + deg(V_\alpha)$.*

We will now describe the decomposition of the moment matrix Λ.

Definition 6.3.8. *If a shape α satisfies the following properties:*

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

then define

$$\lambda_\alpha = \Delta^{|V(\alpha)|} \left(\frac{1}{\sqrt{\Delta n}}\right)^{deg(\alpha)} \prod_{e \in E(\alpha)} \left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e}$$

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

Corollary 6.3.9. $\Lambda = \sum \lambda_\alpha M_\alpha$.

6.3.2 Proving positivity - Qualitative bounds

We use the canonical definition of H'_γ from **??**. In this section, we will prove the following qualitative bounds.

Lemma 6.3.10. *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 τ to the Fourier coefficients.

Definition 6.3.11. *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*

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

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

Lemma 6.3.12. *For all* $U \in \mathcal{I}_{mid}$ *and* $\tau \in \mathcal{M}_U$,

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

We define the following quantity to capture the contribution of the vertices within γ to the Fourier coefficients.

Definition 6.3.13. *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*

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

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

Lemma 6.3.14. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

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

Proof of Lemma 6.3.10

When we compose shapes σ, σ', from Definition 6.3.8, 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 6.3.15. *Define a parity vector ρ to be a vector whose entries are in $\{0, 1\}$.*

Definition 6.3.16. *For $U \in \mathcal{I}_{mid}$, define $\mathcal{P}_U$ to be the set of parity vectors ρ whose coordinates are indexed by U.*

Definition 6.3.17. *For a left shape σ, define $\rho_\sigma \in \mathcal{P}_{V_\sigma}$, called the parity vector of σ, to be the parity vector such that for each vertex $i \in V_\sigma$, the i-th entry of ρ_σ is the parity of $deg^{U_\sigma}(i) + deg^{\sigma}(i)$, that is $(\rho_\sigma)_i \equiv deg^{U_\sigma}(i) + deg^{\sigma}(i) \pmod 2$.*

Definition 6.3.18. *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 ρ.*

Definition 6.3.19. *For a shape τ, for a τ coefficient matrix H_τ and parity vectors $\rho \in \mathcal{P}_{U_\tau}, \rho' \in \mathcal{P}_{V_\tau}$, define the τ-coefficient matrix $H_{\tau,\rho,\rho'}$ as $H_{\tau,\rho,\rho'}(\sigma, \sigma') = H_\tau(\sigma, \sigma')$ if $\sigma \in \mathcal{L}_{U_\tau,\rho}, \sigma' \in \mathcal{L}_{V_\tau,\rho'}$ and 0 otherwise.*

Proposition 6.3.20. *For any shape τ and τ-coefficient matrix H_τ, $H_\tau = \sum_{\rho \in \mathcal{P}_{U_\tau}, \rho' \in \mathcal{P}_{V_\tau}} H_{\tau,\rho,\rho'}$*

Proposition 6.3.21. *For any $U \in \mathcal{I}_{mid}$, $H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_U,\rho,\rho}$*

Proof. For any $\sigma, \sigma' \in \mathcal{L}_U$, using Definition 6.3.8, 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 σ to the Fourier coefficients.

Definition 6.3.22. *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*

$$T(\sigma) = \Delta^{|V(\sigma)| - \frac{|V_\sigma|}{2}} \left(\frac{1}{\sqrt{\Delta n}}\right)^{deg(U_\sigma)} \prod_{e \in E(\sigma)} \left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e}$$

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

Definition 6.3.23. *For $U \in \mathcal{I}_{mid}$ and $\rho \in \mathcal{P}_U$, define v_ρ 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 6.3.24. *For all $U \in \mathcal{I}_{mid}, \rho \in \mathcal{P}_U$, $H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|} v_\rho v_\rho^T$.*

Proof. This follows by verifying the conditions of Definition 6.3.8. ∎

Lemma 6.3.10. *For all $U \in \mathcal{I}_{mid}$, $H_{Id_U} \succeq 0$*

Proof. We have $H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|} \sum_{\rho \in \mathcal{P}_U} v_\rho v_\rho^T \succeq 0$. ∎

Proof of Lemma 6.3.12

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

Proposition 6.3.25. *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 $\pi : P_\tau \to Q_\tau$ such that $H_\tau = \sum_{\rho \in P_\tau} H_{\tau,\rho,\pi(\rho)}$.*

Proof. Using Definition 6.3.8, in order for $H_\tau(\sigma, \sigma')$ to be nonzero, in $\sigma \circ \tau \circ \sigma'$, we must have that for all $i \in U_\tau \cup V_\tau$, $deg^{U_\sigma}(i) + deg^{U_{\sigma'}}(i) + deg^{\sigma \circ \tau \circ \sigma'^T}(i)$ 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 $\sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_U$ with $H_\tau(\sigma, \sigma')$ nonzero, then the parity of σ' is ρ'. Also, observe that ρ' determines ρ. We then take P_τ to be the set of ρ such that ρ' exists, Q_τ to be the set of ρ' and in this case, we define $\pi(\rho) = \rho'$. ∎

We restate Definition 6.3.11 for convenience.

Definition 6.3.11. *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*

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

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

Proposition 6.3.26. *For any $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, suppose we take $\rho \in P_\tau$. Let π be the bijection from Proposition 6.3.25 so that $\pi(\rho) \in Q_\tau$. Then, $H_{\tau,\rho,\pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) v_\rho v_{\pi(\rho)}^T$.*

Proof. This follows by a straightforward verification of the conditions of Definition 6.3.8. ∎

Lemma 6.3.12. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

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

Proof. Let P_τ, Q_τ, π be from Proposition 6.3.25. For $\rho, \rho' \in \mathcal{P}_U$, let $W_{\rho,\rho'} = v_\rho (v_{\rho'})^T$. Then, $H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|} \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho}$ and $H_\tau = \sum_{\rho \in P_\tau} H_{\tau,\rho,\pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) \sum_{\rho \in P_\tau} W_{\rho,\pi(\rho)}$. We have

$$\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|}H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|}H_{Id_U} \end{bmatrix} = \frac{S(\tau)}{|Aut(U)|^2}\begin{bmatrix} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} \end{bmatrix}$$

Since $\frac{S(\tau)}{|Aut(U)|^2} \geq 0$, it suffices to prove that $\begin{bmatrix} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} \succeq 0$. Consider

$$\begin{bmatrix} \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in\mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} = \begin{bmatrix} \sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\rho} & 0 \\ 0 & \sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\rho} \end{bmatrix} + \begin{bmatrix} \sum_{\rho\in P_\tau} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix}$$

We have $\sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\rho} = \sum_{\rho\in\mathcal{P}_U\setminus P_\tau} v_\rho v_\rho^T \succeq 0$. Similarly, $\sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\rho} \succeq 0$ and so, the first term in the above expression, $\begin{bmatrix} \sum_{\rho\in\mathcal{P}_U\setminus P_\tau} W_{\rho,\rho} & 0 \\ 0 & \sum_{\rho\in\mathcal{P}_U\setminus Q_\tau} W_{\rho,\rho} \end{bmatrix}$ is positive semidefinite. For the second term,

$$\begin{aligned} \begin{bmatrix} \sum_{\rho\in P_\tau} W_{\rho,\rho} & \sum_{\rho\in P_\tau} W_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} W^T_{\rho,\pi(\rho)} & \sum_{\rho\in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix} &= \sum_{\rho\in P_\tau} \begin{bmatrix} W_{\rho,\rho} & W_{\rho,\pi(\rho)} \\ W^T_{\rho,\pi(\rho)} & W_{\pi(\rho),\pi(\rho)} \end{bmatrix} \\ &= \sum_{\rho\in P_\tau} \begin{bmatrix} v_\rho v_\rho^T & v_\rho (v_{\pi(\rho)})^T \\ v_{\pi(\rho)}(v_\rho)^T & v_{\pi(\rho)}(v_{\pi(\rho)})^T \end{bmatrix} \\ &= \sum_{\rho\in P_\tau} \begin{bmatrix} v_\rho \\ v_{\pi(\rho)} \end{bmatrix} \begin{bmatrix} v_\rho & v_{\pi(\rho)} \end{bmatrix} \\ &\succeq 0 \end{aligned}$$

■

Proof of Lemma 6.3.14

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

Definition 6.3.27. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for $\gamma \in \Gamma_{U,V}$ and parity vectors $\rho, \rho' \in \mathcal{P}_U$, define the $\gamma \circ \gamma^T$-coefficient matrix $H^{-\gamma,\gamma}_{Id_V,\rho,\rho'}$ as $H^{-\gamma,\gamma}_{Id_V,\rho,\rho'}(\sigma,\sigma') = H^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma')$ if $\sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_{U,\rho'}$ and 0 otherwise.*

Proposition 6.3.28. *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 $P_\gamma \subseteq \mathcal{P}_U$ such that*

$$H^{-\gamma,\gamma}_{Id_V} = \sum_{\rho\in P_\gamma} H^{-\gamma,\gamma}_{Id_V,\rho,\rho}$$

Proof. Take any $\rho \in \mathcal{P}_U$. For $\sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_U$, since $H^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma') = \frac{\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}}{|Aut(V)|}$, $H^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma')$ is nonzero precisely when $\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}$ is nonzero. For this quantity to be nonzero, using Definition 6.3.8, we get that it is necessary, but not sufficient, that the parity vector of σ' must also be ρ. And also observe that there exists a set P_γ of parity vectors ρ for which $H^{-\gamma,\gamma}_{Id_V,\rho,\rho}$ is nonzero and their sum is precisely $H^{-\gamma,\gamma}_{Id_V}$. ■

Definition 6.3.29. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for all $\gamma \in \Gamma_{U,V}$ and parity vector $\rho \in \mathcal{P}_U$, define the matrix $H'_{\gamma,\rho,\rho}$ as $H'_{\gamma,\rho,\rho}(\sigma,\sigma') = H'_\gamma(\sigma,\sigma')$ if $\sigma, \sigma' \in \mathcal{L}_{U,\rho}$ and 0 otherwise.*

Proposition 6.3.30. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for $\gamma \in \Gamma_{U,V}$, $H'_\gamma = \sum_{\rho\in P_\gamma} H'_{\gamma,\rho,\rho}$.*

We restate Definition 6.3.13 for convenience.

Definition 6.3.13. *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*

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

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

Proposition 6.3.31. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for all $\gamma \in \Gamma_{U,V}$ and $\rho \in P_\gamma$,*

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

Proof. Fix $\sigma, \sigma' \in \mathcal{L}_{U,\rho}$ such that $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$. Note that $|V(\sigma)| - \frac{|V_\sigma|}{2} + |V(\sigma')| - \frac{|V_{\sigma'}|}{2} + 2(|V(\gamma)| - \frac{|U_\gamma| + |V_\gamma|}{2}) = |V(\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T)|$. Using Definition 6.3.8, we can easily verify that $\lambda_{\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T} = T(\sigma)T(\sigma')S(\gamma)^2$. Therefore, $H^{-\gamma,\gamma}_{Id_V,\rho,\rho}(\sigma, \sigma') = \frac{|Aut(U)|}{|Aut(V)|} S(\gamma)^2 H_{Id_U,\rho,\rho}(\sigma, \sigma')$. Since $H'_{\gamma,\rho,\rho}(\sigma, \sigma') = H_{Id_U,\rho,\rho}(\sigma, \sigma')$ whenever $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$, this completes the proof. ∎

Lemma 6.3.14. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

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

Proof. We have

$$\begin{aligned}\frac{|Aut(V)|}{|Aut(U)|}\cdot\frac{1}{S(\gamma)^2}H^{-\gamma,\gamma}_{Id_V} &= \sum_{\rho\in P_\gamma}\frac{|Aut(V)|}{|Aut(U)|}\cdot\frac{1}{S(\gamma)^2}H^{-\gamma,\gamma}_{Id_V,\rho,\rho}\\ &= \sum_{\rho\in P_\gamma}H'_{\gamma,\rho,\rho}\\ &\preceq \sum_{\rho\in\mathcal{P}_U}H'_{\gamma,\rho,\rho}\\ &= H'_\gamma\end{aligned}$$

where we used the fact that for all $\rho\in\mathcal{P}_U$, we have $H'_{\gamma,\rho,\rho}\succeq 0$ which can be proved the same way as the proof of Lemma 6.3.10. ∎

6.4 Qualitative bounds for Sparse PCA

6.4.1 Pseudo-calibration

Definition 6.4.1 (Slack parameter)**.** *Define the slack parameter to be* $\Delta = d^{-C_\Delta\varepsilon}$ *for a constant* $C_\Delta > 0$.

We will pseudo-calibrate with respect the following pair of random and planted distributions which we denote ν and μ respectively.

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

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

Definition 6.4.2. *For a nonnegative integer t, define* $t!! = \begin{cases} \frac{(2t)!}{t!2^t} = 1 \times 3 \times \ldots \times t, & \text{if } t \text{ is odd} \\ 0, & \text{otherwise} \end{cases}$

Lemma 6.4.3. *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*

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

Else, $\mathbb{E}_\mu[u^I h_a(v)] = 0$.

Proof. $v_1, \ldots, v_m \sim \mu$ can be written as $v_i = g_i + \sqrt{\lambda} b_i l_i u$ 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 Δ.

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

$$\mathop{\mathbb{E}}_{(u,l_i,b_i,g_i)\sim\mu}[u^I h_a(v)] = \mathop{\mathbb{E}}_{(u,l_i,b_i)\sim\mu}[u^I \prod_{i,j} (\sqrt{\lambda} b_i l_i u_j)^{a_{ij}}] = \mathop{\mathbb{E}}_{(u,l_i,b_i)\sim\mu}[\prod_{i\in[m]} (b_i l_i)^{e_i} \prod_{j\in[d]} u_j^{f_j}] \prod_{i,j} \sqrt{\lambda}^{a_{ij}}$$

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 Δ. 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 ν, we need e_i, f_j to be even. The expectation then becomes

$$\Delta^{c_1} \left(\frac{k}{d}\right)^{c_2} \mathop{\mathbb{E}}_{(u,l_i)\sim\mu}[\prod_{i\in[m]} l_i^{e_i} \prod_{j\in[d]} u_j^{f_j}] \prod_{i,j} \sqrt{\lambda}^{a_{ij}} = \left(\frac{1}{\sqrt{k}}\right)^{|I|} \left(\frac{k}{d}\right)^{c_2} \Delta^{c_1} \prod_{i\in[m]} (e_i - 1)!! \prod_{i,j} \frac{\sqrt{\lambda}^{a_{ij}}}{\sqrt{k}^{a_{ij}}}$$

The last equality follows because, for each j such that u_j is nonzero, we have $u_j^t = (\frac{1}{\sqrt{k}})^t$ and $\mathbb{E}_{g\sim\mathcal{N}(0,1)}[g^t] = (t-1)!!$ if t is even. ■

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

Definition 6.4.4. *Define the degree of SoS to be $D_{sos} = d^{C_{sos}\varepsilon}$ for some constant $C_{sos} > 0$ that we choose later.*

Definition 6.4.5 (Truncation parameters)**.** *Define the truncation parameters to be $D_V = d^{C_V\varepsilon}, D_E = d^{C_E\varepsilon}$ for some constants $C_V, C_E > 0$.*

Remark 6.4.6 (Choice of parameters)**.** *We first set $\varepsilon > 0$ to be a sufficiently small constant. Based on the choice of ε, 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_{sos} to be sufficiently small to satisfy the inequalities we use.*

Remark 6.4.7. *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 6.4.8. *For the analysis of Sparse PCA, we will use the following notation.*

- *For a shape α 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 α and vertex i in α, let $deg^{\alpha}(i) = \sum_{i \in e \in E(\alpha)} l_e$.*
- *For any shape α, let $deg(\alpha) = deg(U_\alpha) + deg(V_\alpha)$.*
- *For an index shape $U \in \mathcal{I}_{mid}$ 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 $U \in \mathcal{I}_{mid}$. Also, denote by $|U|_t$ the length of the tuple U_t.*

We will now describe the decomposition of the moment matrix Λ.

Definition 6.4.9. *If a shape α satisfies the following properties:*

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

then define

$$\lambda_\alpha = \left(\frac{1}{\sqrt{k}}\right)^{deg(\alpha)} \left(\frac{k}{d}\right)^{|\alpha|_1} \Delta^{|\alpha|_2} \prod_{j \in V_2(\alpha)} (deg^{\alpha}(j) - 1)!! \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

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

Corollary 6.4.10. $\Lambda = \sum \lambda_\alpha M_\alpha$.

6.4.2 Proving positivity - Qualitative bounds

We use the canonical definition of H'_γ from **??**. In this section, we will prove the following qualitative bounds.

Lemma 6.4.11. *For all $U \in \mathcal{I}_{mid}$, $H_{Id_U} \succeq 0$*

For technical reasons, it will be convenient to discretize the Normal distribution. The following fact follows from standard results on Gaussian quadrature, see for e.g. [51, Lemma 4.3].

Fact 6.4.12 (Discretizing the Normal distribution)**.** *There is an absolute constant C_{disc} such that, for any positive integer D, there exists a distribution $\mathcal{E}$ 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*
- $\mathbb{E}_{g\sim\mathcal{E}}[g^t] = \mathbb{E}_{g\sim\mathcal{N}(0,1)}[g^t]$ *for all* $t = 0, 1, \ldots, 2D-1$

We define the following quantity to capture the contribution of the vertices within τ to the Fourier coefficients.

Definition 6.4.13. *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*

$$S(\tau) = \left(\frac{k}{d}\right)^{|\tau|_1 - |U_\tau|_1} \Delta^{|\tau|_2 - |U_\tau|_2} \prod_{j \in V_2(\tau)\setminus U_\tau \setminus V_\tau} (deg^\tau(j) - 1)!! \prod_{e \in E(\tau)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

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

Definition 6.4.14. *For any shape τ, suppose $U' = (U_\tau)_2, V' = (V_\tau)_2$ are the type 2 vertices in U_τ, V_τ respectively. Define*

$$R(\tau) = (C_{disc}\sqrt{D_E})^{\sum_{j \in U' \cup V'} deg^\tau(j)}$$

where C_{disc} is the constant from Fact 6.4.12.

Lemma 6.4.15. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

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

We define the following quantity to capture the contribution of the vertices within γ to the Fourier coefficients.

Definition 6.4.16. *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*

$$S(\gamma) = \left(\frac{k}{d}\right)^{|\gamma|_1 - \frac{|U_\gamma|_1 + |V_\gamma|_1}{2}} \Delta^{|\gamma|_2 - \frac{|U_\gamma|_2 + |V_\gamma|_2}{2}} \prod_{j \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!! \prod_{e \in E(\gamma)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

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

Lemma 6.4.17. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

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

Proof of Lemma 6.4.11

When we compose shapes σ, σ', from Definition 6.4.9, 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 6.4.18. *Define a parity vector ρ to be a vector whose entries are in $\{0, 1\}$.*

Definition 6.4.19. *For $U \in \mathcal{I}_{mid}$, define $\mathcal{P}_U$ to be the set of parity vectors ρ whose coordinates are indexed by U_1 followed by U_2.*

Definition 6.4.20. *For a left shape σ, define $\rho_\sigma \in \mathcal{P}_{V_\sigma}$, called the parity vector of σ, to be the parity vector such that for each vertex $i \in V_\sigma$, the i-th entry of ρ_σ is the parity of $deg^{U_\sigma}(i) + deg^{\sigma}(i)$, that is, $(\rho_\sigma)_i \equiv deg^{U_\sigma}(i) + deg^{\sigma}(i) \pmod 2$.*

Definition 6.4.21. *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 ρ.*

Definition 6.4.22. *For a shape τ, for a τ coefficient matrix H_τ and parity vectors $\rho \in \mathcal{P}_{U_\tau}, \rho' \in \mathcal{P}_{V_\tau}$, define the τ-coefficient matrix $H_{\tau,\rho,\rho'}$ as $H_{\tau,\rho,\rho'}(\sigma,\sigma') = H_\tau(\sigma,\sigma')$ if $\sigma \in \mathcal{L}_{U_\tau,\rho}, \sigma' \in \mathcal{L}_{V_\tau,\rho'}$ and 0 otherwise.*

Proposition 6.4.23. *For any shape τ and τ-coefficient matrix H_τ, $H_\tau = \sum_{\rho \in \mathcal{P}_{U_\tau}, \rho' \in \mathcal{P}_{V_\tau}} H_{\tau,\rho,\rho'}$*

Proposition 6.4.24. *For any $U \in \mathcal{I}_{mid}$, $H_{Id_U} = \sum_{\rho \in \mathcal{P}_U} H_{Id_U,\rho,\rho}$*

Proof. For any $\sigma, \sigma' \in \mathcal{L}_U$, using Definition 6.4.9, note that in order for $H_{Id_U}(\sigma,\sigma')$ 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 6.4.25. *Let $\mathcal{D}$ be a distribution over the real numbers obtained by setting $D = D_E$ in Fact 6.4.12. So, in particular, for any x sampled from $\mathcal{D}$, we have $|x| \le C_{disc}\sqrt{D_E}$ and for $t \le 2D_E - 1$, $\mathbb{E}_{x\sim\mathcal{D}}[x^t] = (t-1)!!$.*

We define the following quantity to capture the contribution of the vertices within σ to the Fourier coefficients.

Definition 6.4.26. *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*

$$T(\sigma) = \left(\frac{1}{\sqrt{k}}\right)^{deg(U_\sigma)} \left(\frac{k}{d}\right)^{|\sigma|_1 - \frac{|V_\sigma|_1}{2}} \Delta^{|\sigma|_2 - \frac{|V_\sigma|_2}{2}} \prod_{j \in V_2(\sigma)\setminus V_\sigma} (deg^{\sigma}(j) - 1)!! \prod_{e \in E(\sigma)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

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

Definition 6.4.27. *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 σth entry is $T(\sigma)\prod_{i\in U_2} x_i^{deg^\sigma(i)}$ if $\sigma \in \mathcal{L}_{U,\rho}$ and 0 otherwise.*

Proposition 6.4.28. *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 $\mathcal{D}$. Then,*

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

Proof. Observe that for $\sigma, \sigma' \in \mathcal{L}_{U,\rho}$ and $t \in \{1,2\}$, $(|\sigma|_t - \frac{|V_\sigma|_t}{2}) + (|\sigma'|_t - \frac{|V_{\sigma'}|_t}{2}) = |\sigma \circ \sigma'|_t$. The result follows by verifying the conditions of Definition 6.4.9 and using Definition 6.4.25. ∎

Lemma 6.4.11. *For all $U \in \mathcal{I}_{mid}$, $H_{Id_U} \succeq 0$*

Proof. We have $H_{Id_U} = \sum_{\rho\in\mathcal{P}_U} H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|}\sum_{\rho\in\mathcal{P}_U} \mathbb{E}_{x_{U_2}\sim\mathcal{D}^{U_2}}[v_{\rho,x_{U_2}} v_{\rho,x_{U_2}}^T] \succeq 0$. ∎

Proof of Lemma 6.4.15

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

Proposition 6.4.29. *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 $\pi : P_\tau \to Q_\tau$ such that $H_\tau = \sum_{\rho\in P_\tau} H_{\tau,\rho,\pi(\rho)}$.*

Proof. Using Definition 6.4.9, in order for $H_\tau(\sigma,\sigma')$ to be nonzero, we must have that, in $\sigma \circ \tau \circ \sigma'$, for all $i \in U_\tau \cup V_\tau$, $deg^{U_\sigma}(i) + deg^{U_{\sigma'}}(i) + deg^{\sigma\circ\tau\circ\sigma'^T}(i)$ 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 $\sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_U$ with $H_\tau(\sigma,\sigma')$ nonzero, then the parity of σ' is ρ'. Also, observe that ρ' determines ρ. We

then take P_τ to be the set of ρ such that ρ' exists, Q_τ to be the set of ρ' and in this case, we define $\pi(\rho) = \rho'$. ∎

We restate Definition 6.4.13 for convenience.

Definition 6.4.13. *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*

$$S(\tau) = \left(\frac{k}{d}\right)^{|\tau|_1 - |U_\tau|_1} \Delta^{|\tau|_2 - |U_\tau|_2} \prod_{j \in V_2(\tau) \setminus U_\tau \setminus V_\tau} (deg^\tau(j) - 1)!! \prod_{e \in E(\tau)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

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

Proposition 6.4.30. *For any $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, suppose we take $\rho \in P_\tau$. Let π be the bijection from Proposition 6.4.29 so that $\pi(\rho) \in Q_\tau$. Let $U' = (U_\tau)_2, V' = (V_\tau)_2$ be the type 2 vertices in U_τ, V_τ respectively. Let x_i for $i \in U' \cup V'$ be random variables independently sampled from $\mathcal{D}$. Define $x_{U'}$ (resp. $x_{V'}$) to be the subset of variables x_i for $i \in U'$ (resp. $i \in V'$). Then,*

$$H_{\tau,\rho,\pi(\rho)} = \frac{1}{|Aut(U)|^2} S(\tau) \mathop{\mathbb{E}}_{x} \left[v_{\rho, x_{U'}} \left(\prod_{i \in U' \cup V'} x_i^{deg^\tau(i)} \right) v^T_{\pi(\rho), x_{V'}} \right]$$

Proof. For $\sigma \in L_{U,\rho}, \sigma' \in \mathcal{L}_{U,\pi(\rho)}$ and $t \in \{1, 2\}$, we have $(|\tau|_t - |U_\tau|_t) + (|\sigma|_t - \frac{|V_\sigma|_t}{2}) + (|\sigma'|_t - \frac{|V_{\sigma'}|_t}{2}) = |\sigma \circ \tau \circ \sigma'|_t$. The result then follows by a straightforward verification of the conditions of Definition 6.4.9 using Definition 6.4.25. ∎

Lemma 6.4.15. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

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

Proof. Let P_τ, Q_τ, π be from Proposition 6.4.29. Let $U' = (U_\tau)_2, V' = (V_\tau)_2$ be the type 2 vertices in U_τ, V_τ respectively. Let x_i for $i \in U' \cup V'$ be random variables independently sampled from $\mathcal{D}$. Define $x_{U'}$ (resp. $x_{V'}$) to be the subset of variables x_i for $i \in U'$ (resp. $i \in V'$).

For $\rho \in \mathcal{P}_U$, define $W_{\rho,\rho} = \mathbb{E}_{y_{U_2} \sim \mathcal{D}^{U_2}}[v_{\rho,y_{U_2}} v^T_{\rho,y_{U_2}}]$ so that $H_{Id_U,\rho,\rho} = \frac{1}{|Aut(U)|} W_{\rho,\rho}$. Observe that $W_{\rho,\rho} = \mathbb{E}[v_{\rho,x_{U'}} v^T_{\rho,x_{U'}}] = \mathbb{E}[v_{\rho,x_{V'}} v^T_{\rho,x_{V'}}]$ because $x_{U'}$ and $x_{V'}$ are also sets of variables sampled from $\mathcal{D}$ and, U', V' have the same size as U_2 because $U_\tau = V_\tau = U$.

For $\rho, \rho' \in \mathcal{P}_U$, define $Y_{\rho,\rho'} = \mathbb{E}\left[v_{\rho,x_{U'}} \left(\prod_{i \in U' \cup V'} x_i^{deg^\tau(i)}\right) v^T_{\pi(\rho),x_{V'}}\right]$. 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

$$\begin{bmatrix} \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} = \frac{S(\tau)}{|Aut(U)|^2} \begin{bmatrix} R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y^T_{\rho,\pi(\rho)} & R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} \end{bmatrix}$$

Since $\frac{S(\tau)}{|Aut(U)|^2} \geq 0$, it suffices to prove that $\begin{bmatrix} R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y^T_{\rho,\pi(\rho)} & R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} \succeq 0$. Consider

$$\begin{aligned} \begin{bmatrix} R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y^T_{\rho,\pi(\rho)} & R(\tau) \sum_{\rho \in \mathcal{P}_U} W_{\rho,\rho} \end{bmatrix} =& R(\tau) \begin{bmatrix} \sum_{\rho \in \mathcal{P}_U \setminus P_\tau} W_{\rho,\rho} & 0 \\ 0 & \sum_{\rho \in \mathcal{P}_U \setminus Q_\tau} W_{\rho,\rho} \end{bmatrix} \\ &+ \begin{bmatrix} R(\tau) \sum_{\rho \in P_\tau} W_{\rho,\rho} & \sum_{\rho \in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho \in P_\tau} Y^T_{\rho,\pi(\rho)} & R(\tau) \sum_{\rho \in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix} \end{aligned}$$

We have $\sum_{\rho \in \mathcal{P}_U \setminus P_\tau} W_{\rho,\rho} = \sum_{\rho \in \mathcal{P}_U \setminus P_\tau} \mathbb{E}[v_{\rho,x_{U'}} v^T_{\rho,x_{U'}}] \succeq 0$. Similarly, $\sum_{\rho \in \mathcal{P}_U \setminus Q_\tau} W_{\rho,\rho} \succeq 0$. Also, $R(\tau) \geq 0$ and so, the first term in the above expression, $R(\tau) \begin{bmatrix} \sum_{\rho \in \mathcal{P}_U \setminus P_\tau} W_{\rho,\rho} & 0 \\ 0 & \sum_{\rho \in \mathcal{P}_U \setminus Q_\tau} W_{\rho,\rho} \end{bmatrix}$

is positive semidefinite. For the second term,

$$\begin{bmatrix} R(\tau)\sum_{\rho\in P_\tau} W_{\rho,\rho} & \sum_{\rho\in P_\tau} Y_{\rho,\pi(\rho)} \\ \sum_{\rho\in P_\tau} Y^T_{\rho,\pi(\rho)} & R(\tau)\sum_{\rho\in P_\tau} W_{\pi(\rho),\pi(\rho)} \end{bmatrix}$$

$$= \sum_{\rho\in P_\tau} \begin{bmatrix} R(\tau)\,\mathbb{E}[v_{\rho,x_{U'}} v^T_{\rho,x_{U'}}] & \mathbb{E}\left[v_{\rho,x_{U'}} \left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right) v^T_{\pi(\rho),x_{V'}}\right] \\ \mathbb{E}\left[v^T_{\rho,x_{U'}} \left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right) v_{\pi(\rho),x_{V'}}\right] & R(\tau)\,\mathbb{E}[v_{\pi(\rho),x_{V'}} v^T_{\pi(\rho),x_{V'}}] \end{bmatrix}$$

$$= \sum_{\rho\in P_\tau} \mathbb{E}\begin{bmatrix} R(\tau) v_{\rho,x_{U'}} v^T_{\rho,x_{U'}} & v_{\rho,x_{U'}} \left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right) v^T_{\pi(\rho),x_{V'}} \\ v^T_{\rho,x_{U'}} \left(\prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}\right) v_{\pi(\rho),x_{V'}} & R(\tau) v_{\pi(\rho),x_{V'}} v^T_{\pi(\rho),x_{V'}} \end{bmatrix}$$

We will prove that the term inside the expectation is positive semidefinite for each $\rho \in P_\tau$ and each sampling of the x_i from $\mathcal{D}$, which will complete the proof. Fix $\rho \in P_\tau$ and any sampling of the x_i from $\mathcal{D}$. Let $w_1 = v_{\rho,X_{U'}}, w_2 = v_{\pi(\rho),x_{V'}}$. Let $E = \prod_{i\in U'\cup V'} x_i^{deg^\tau(i)}$. We would like to prove that $\begin{bmatrix} R(\tau) w_1 w_1^T & E w_1 w_2^T \\ E w_1^T w_2 & R(\tau) w_2 w_2^T \end{bmatrix} \succeq 0$. For all y sampled from $\mathcal{D}$, $|y| \le C_{disc}\sqrt{D_E}$ and so, $|E| \le (C_{disc}\sqrt{D_E})^{\sum_{j\in U'\cup V'} deg^\tau(j)} = R(\tau)$.

If $E \ge 0$, then

$$\begin{bmatrix} R(\tau) w_1 w_1^T & E w_1 w_2^T \\ E w_1^T w_2 & R(\tau) w_2 w_2^T \end{bmatrix} = (R(\tau) - E)\begin{bmatrix} w_1 w_1^T & 0 \\ 0 & w_2 w_2^T \end{bmatrix} + E\begin{bmatrix} w_1 w_1^T & w_1 w_2^T \\ w_1^T w_2 & w_2 w_2^T \end{bmatrix}$$

$$= (R(\tau) - E)\left(\begin{bmatrix} w_1 \\ 0 \end{bmatrix}\begin{bmatrix} w_1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ w_2 \end{bmatrix}\begin{bmatrix} 0 & w_2 \end{bmatrix}\right) + E\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\begin{bmatrix} w_1 & w_2 \end{bmatrix}$$

$$\succeq 0$$

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

$$\begin{aligned}
\begin{bmatrix} R(\tau)w_1w_1^T & Ew_1w_2^T \\ Ew_1^Tw_2 & R(\tau)w_2w_2^T \end{bmatrix} &= (R(\tau)+E)\begin{bmatrix} w_1w_1^T & 0 \\ 0 & w_2w_2^T \end{bmatrix} - E\begin{bmatrix} w_1w_1^T & -w_1w_2^T \\ -w_1^Tw_2 & w_2w_2^T \end{bmatrix} \\
&= (R(\tau)+E)\left(\begin{bmatrix} w_1 \\ 0 \end{bmatrix}\begin{bmatrix} w_1 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ w_2 \end{bmatrix}\begin{bmatrix} 0 & w_2 \end{bmatrix}\right) - E\begin{bmatrix} w_1 \\ -w_2 \end{bmatrix}\begin{bmatrix} w_1 & -w_2 \end{bmatrix} \\
&\succeq 0
\end{aligned}$$

since $R(\tau) + E \succeq 0$. ∎

Proof of Lemma 6.4.17

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

Definition 6.4.31. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for $\gamma \in \Gamma_{U,V}$ and parity vectors $\rho, \rho' \in \mathcal{P}_U$, define the $\gamma \circ \gamma^T$-coefficient matrix $H^{-\gamma,\gamma}_{Id_V,\rho,\rho'}$ as $H^{-\gamma,\gamma}_{Id_V,\rho,\rho'}(\sigma,\sigma') = H^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma')$ if $\sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_{U,\rho'}$ and 0 otherwise.*

Proposition 6.4.32. *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 $P_\gamma \subseteq \mathcal{P}_U$ such that*

$$H^{-\gamma,\gamma}_{Id_V} = \sum_{\rho\in P_\gamma} H^{-\gamma,\gamma}_{Id_V,\rho,\rho}$$

Proof. Take any $\rho \in \mathcal{P}_U$. For $\sigma \in \mathcal{L}_{U,\rho}, \sigma' \in \mathcal{L}_U$, since $H^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma') = \frac{\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}}{|Aut(V)|}$, $H^{-\gamma,\gamma}_{Id_V}(\sigma,\sigma')$ is nonzero precisely when $\lambda_{\sigma\circ\gamma\circ\gamma^T\circ\sigma'^T}$ is nonzero. For this quantity to be nonzero, using Definition 6.4.9, we get that it is necessary, but not sufficient, that the parity

vector of σ' must also be ρ. And also observe that there exists a set P_γ of parity vectors ρ for which $H^{-\gamma,\gamma}_{Id_V,\rho,\rho}$ is nonzero and their sum is precisely $H^{-\gamma,\gamma}_{Id_V}$. ∎

Definition 6.4.33. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for all $\gamma \in \Gamma_{U,V}$ and parity vector $\rho \in \mathcal{P}_U$, define the matrix $H'_{\gamma,\rho,\rho}$ as $H'_{\gamma,\rho,\rho}(\sigma,\sigma') = H'_\gamma(\sigma,\sigma')$ if $\sigma,\sigma' \in \mathcal{L}_{U,\rho}$ and 0 otherwise.*

Proposition 6.4.34. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, for $\gamma \in \Gamma_{U,V}$, $H'_\gamma = \sum_{\rho \in P_\gamma} H'_{\gamma,\rho,\rho}$.*

We will now define vectors which are truncations of $v_{\rho,x_{U_2}}$.

Definition 6.4.35. *Let $U, V \in \mathcal{I}_{mid}$ 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 x_{U_2}. For $\rho \in \mathcal{P}_U$, define $v^{-\gamma}_{\rho,x_{U_2}}$ to be the vector indexed by left shapes $\sigma \in \mathcal{L}$ such that the σth entry is $v_{\rho,x_{U_2}}(\sigma)$ if $|V(\sigma \circ \gamma)| \leq D_V$ and 0 otherwise.*

We restate Definition 6.4.16 for convenience.

Definition 6.4.16. *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*

$$S(\gamma) = \left(\frac{k}{d}\right)^{|\gamma|_1 - \frac{|U_\gamma|_1 + |V_\gamma|_1}{2}} \Delta^{|\gamma|_2 - \frac{|U_\gamma|_2 + |V_\gamma|_2}{2}} \prod_{j \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!! \prod_{e \in E(\gamma)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}}$$

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

Proposition 6.4.36. *For any $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, and for any $\gamma \in \Gamma_{U,V}$, suppose we take $\rho \in P_\gamma$. When we compose γ with γ^T to get $\gamma \circ \gamma^T$, let $U' = (U_{\gamma\circ\gamma^T})_2, V' = (V_{\gamma\circ\gamma^T})_2$ be the type 2 vertices in $U_{\gamma\circ\gamma^T}, V_{\gamma\circ\gamma^T}$ 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 $V_\gamma = U^T_\gamma$. Let x_i*

for $i \in U' \cup W' \cup V'$ be random variables independently sampled from $\mathcal{D}$. 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,

$$H^{-\gamma,\gamma}_{Id_V,\rho,\rho} = \frac{1}{|Aut(V)|} S(\gamma)^2 \mathop{\mathbb{E}}_{x} \left[(v^{-\gamma}_{\rho,x_{U'}}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v^{-\gamma}_{\rho,x_{V'}})^T \right]$$

Proof. Fix $\sigma, \sigma' \in \mathcal{L}_{U,\rho}$ such that $|V(\sigma \circ \gamma)|, |V(\sigma' \circ \gamma)| \leq D_V$. Note that for $t \in \{1,2\}$, $|\sigma|_t - \frac{|V_\sigma|_t}{2} + |\sigma'|_t - \frac{|V_{\sigma'}|_t}{2} + 2(|\gamma|_t - \frac{|U_\gamma|_t + |V_\gamma|_t}{2}) = |\sigma \circ \gamma \circ \gamma^T \circ \sigma'^T|_t$. We can easily verify the equality using Definition 6.4.9 and Definition 6.4.25. ∎

Proposition 6.4.37. *For any $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$, and for any $\gamma \in \Gamma_{U,V}$, suppose we take $\rho \in \mathcal{P}_U$. Then,*

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

We can now prove Lemma 6.4.17.

Lemma 6.4.17. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

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

Proof. Let U', V', W' be as in Proposition 6.4.36. We have

$$\begin{aligned}
\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H^{-\gamma,\gamma}_{Id_V} &= \sum_{\rho \in P_\gamma} \frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H^{-\gamma,\gamma}_{Id_V,\rho,\rho} \\
&= \sum_{\rho \in P_\gamma} \frac{1}{|Aut(U)|} \cdot \frac{1}{R(\gamma)^2} \mathop{\mathbb{E}}_{x} \left[(v^{-\gamma}_{\rho,x_{U'}}) \left(\prod_{i \in U' \cup W' \cup V'} x_i^{deg^{\gamma \circ \gamma^T}(i)} \right) (v^{-\gamma}_{\rho,x_{V'}})^T \right]
\end{aligned}$$

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

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

This reduces to proving that

$$\frac{2}{R(\gamma)^2}\mathop{\mathbb{E}}_{x}\left[(v_{\rho,x_{U'}}^{-\gamma})\left(\prod_{i\in U'\cup W'\cup V'}x_i^{deg^{\gamma\circ\gamma^T}(i)}\right)(v_{\rho,x_{V'}}^{-\gamma})^T\right]\preceq 2\mathop{\mathbb{E}}_{y_{U_2}\sim\mathcal{D}^{U_2}}\left[(v_{\rho,y_{U_2}}^{-\gamma})(v_{\rho,y_{U_2}}^{-\gamma})^T\right]$$
$$=\mathop{\mathbb{E}}_{x}\left[(v_{\rho,x_{U'}}^{-\gamma})(v_{\rho,x_{U'}}^{-\gamma})^T+(v_{\rho,x_{V'}}^{-\gamma})(v_{\rho,x_{V'}}^{-\gamma})^T\right]$$

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

Since $H^{-\gamma,\gamma}_{Id_V,\rho,\rho}$ is symmetric, we have

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

So, it suffices to prove

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

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

$$\frac{1}{R(\gamma)^2}\left((v_{\rho,x_{U'}}^{-\gamma})\left(\prod_{i\in U'\cup W'\cup V'}x_i^{deg^{\gamma\circ\gamma^T}(i)}\right)(v_{\rho,x_{V'}}^{-\gamma})^T+(v_{\rho,x_{V'}}^{-\gamma})\left(\prod_{i\in U'\cup W'\cup V'}x_i^{deg^{\gamma\circ\gamma^T}(i)}\right)(v_{\rho,x_{U'}}^{-\gamma})^T\right)$$
$$\preceq(v_{\rho,x_{U'}}^{-\gamma})(v_{\rho,x_{U'}}^{-\gamma})^T+(v_{\rho,x_{V'}}^{-\gamma})(v_{\rho,x_{V'}}^{-\gamma})^T$$

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

$$\frac{E}{R(\gamma)^2}(w_1w_2^T+w_2w_1^T)\preceq w_1w_1^T+w_2w_2^T$$

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

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

$$\begin{aligned}\frac{E}{R(\gamma)^2}(w_1 w_2^T + w_2 w_1^T) &\preceq \frac{E}{R(\gamma)^2}(w_1 w_1^T + w_2 w_2^T) \\ &\preceq w_1 w_1^T + w_2 w_2^T\end{aligned}$$

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

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

$$\begin{aligned}\frac{E}{R(\gamma)^2}(w_1 w_2^T + w_2 w_1^T) &\preceq \frac{-E}{R(\gamma)^2}(w_1 w_1^T + w_2 w_2^T) \\ &\preceq w_1 w_1^T + w_2 w_2^T\end{aligned}$$

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

Finally, we use the fact that for all $\rho \in \mathcal{P}_U$, we have $H'_{\gamma,\rho,\rho} \succeq 0$ which can be proved the same way as the proof of Lemma 6.4.11. Therefore,

$$\begin{aligned}\frac{|Aut(V)|}{|Aut(U)|} \cdot \frac{1}{S(\gamma)^2 R(\gamma)^2} H^{-\gamma,\gamma}_{Id_V} &\preceq \sum_{\rho \in P_\gamma} H'_{\gamma,\rho,\rho} \\ &\preceq \sum_{\rho \in \mathcal{P}_U} H'_{\gamma,\rho,\rho} \\ &= H'_\gamma\end{aligned}$$

■

6.4.3 Intuition for quantitative bounds

In this section, we will give some intuition on the bounds needed for our main theorem Theorem 4.4.1, which is formally proved in Section 7.3. Informally, the theorem states that when $m \leq \frac{d}{\lambda^2}$ and $m \leq \frac{k^2}{\lambda^2}$, then $\Lambda \succeq 0$ with high probability.

We will try and understand why the inequality $\lambda^2_{\sigma\circ\tau\circ\sigma'^T} \|M_\tau\|^2 \leq \lambda_{\sigma\circ\sigma^T}\lambda_{\sigma'\circ\sigma'^T}$ holds. Assume for simplicity that $d < n$ and consider the shapes in Fig. 6.1. The assumption $d < n$ is used in this example since otherwise, if $d > n$, the decomposition differs from what's shown in the figure.

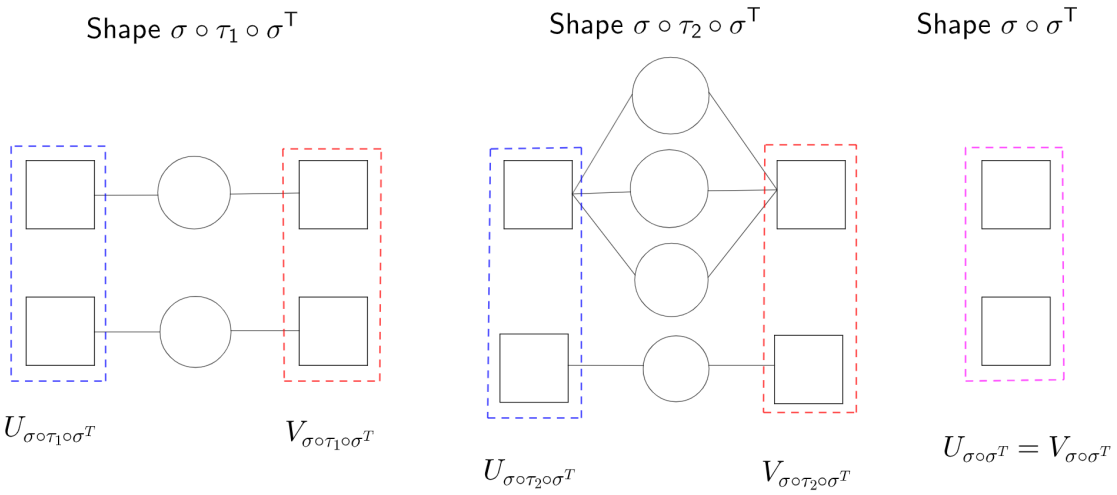

Figure 6.1: Shapes $\sigma \circ \tau_1 \circ \sigma^T, \sigma \circ \tau_2 \circ \sigma^T$ and $\sigma \circ \sigma^T$. All edges have label 1.

Firstly, the shape $\sigma \circ \sigma^T$ has a coefficient of $\lambda_{\sigma\circ\sigma^T} \approx \left(\frac{1}{\sqrt{k}}\right)^4 \left(\frac{k}{d}\right)^2$.

The first shape $\sigma \circ \tau_1 \circ \sigma^T$ has a coefficient of $\lambda_{\sigma\circ\tau_1\circ\sigma^T} \approx \left(\frac{1}{\sqrt{k}}\right)^4 \left(\frac{k}{d}\right)^4 \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^4$ and with high probability, upto lower order terms, $\|M_{\tau_1}\| \leq md$ (these norm bounds follow from [2]). So, the inequality $\lambda^2_{\sigma\circ\tau_1\circ\sigma^T} \|M_{\tau_1}\|^2 \leq \lambda_{\sigma\circ\sigma^T}\lambda_{\sigma\circ\sigma^T}$ rearranges to $m \leq \frac{d}{\lambda^2}$. But this is precisely one of the assumptions on m. Moreover, this also confirms that we need this assumption on m in order for our strategy to go through.

The second shape $\sigma \circ \tau_2 \circ \sigma^T$ has a coefficient of $\lambda_{\sigma\circ\tau_2\circ\sigma^T} \approx \left(\frac{1}{\sqrt{k}}\right)^4 \left(\frac{k}{d}\right)^4 \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^8$ and with high probability, upto lower order terms, $\|M_{\tau_2}\| \leq m^2 d$. So, the inequality

$\lambda^2_{\sigma\circ\tau_2\circ\sigma^T} \|M_{\tau_2}\|^2 \leq \lambda_{\sigma\circ\sigma^T}\lambda_{\sigma\circ\sigma^T}$ rearranges to $m^2 \leq \frac{k^2 d}{\lambda^4}$. But this is obtained simply by multiplying our assumptions on m, namely $m \leq \frac{k^2}{\lambda^2}$ and $m \leq \frac{d}{\lambda^2}$.

Moreover, consider a shape of the form $\sigma \circ \tau_3 \circ \sigma^T$ where τ_3 is similar to τ_2 except it has t (instead of 3) different circle vertices that are common neighbors to the top 2 square vertices. Analyzing our required inequality, we get for our strategy to go through, m has to satisfy $m \leq \frac{k^2}{\lambda^2} \cdot \left(\frac{d}{k^2}\right)^{\frac{2}{t+1}}$. By taking t arbitrarily large, we can see that the condition $m \leq \frac{k^2}{\lambda^2}$ is needed.

So, we get that for our analysis to go through, the assumptions $m \leq \frac{d}{\lambda^2}$ and $m \leq \frac{k^2}{\lambda^2}$ are necessary. We will prove that in fact, these are sufficient. To do this, we use a charging argument that exploits the special structure of the shapes α that appear in our decomposition of Λ and their coefficients λ_α, as we obtained in Definition 6.4.9. For details, see Section 7.3.

CHAPTER 7

QUANTITATIVE BOUNDS

In this chapter, we will prove the main SoS lower bounds Theorem 4.2.1, Theorem 4.3.1 and Theorem 4.4.1 by building on the qualitative bounds from Chapter 6. The material in this chapter is adapted from [149].

7.1 Planted slightly denser subgraph: Full verification

In this section, we will prove all the required bounds to prove Theorem 4.2.1.

Theorem 4.2.1. *Let $C_p > 0$. There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $k \leq n^{\frac{1}{2}-\varepsilon}$ and $p = \frac{1}{2} + \frac{n^{-C_p\varepsilon}}{2}$, then with high probability, the candidate moment matrix Λ given by pseudo-calibraton for degree $n^{C\varepsilon}$ Sum-of-Squares is PSD.*

In particular, we will use **??** where we choose ε in the theorem, not to be confused with the ε in Theorem 4.2.1, to be an arbitrarily small constant.

To invoke the machinery, we basically have to verify the following conditions, the qualitiative versions of which have already been shown in Section 6.2.

Lemma 7.1.1. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

Lemma 7.1.2. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

Lemma 7.1.3. *Whenever* $\|M_\alpha\| \leq B_{norm}(\alpha)$ *for all* $\alpha \in \mathcal{M}'$,

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

Corollary 7.1.4. *With constant probability,* $\Lambda \succeq 0$.

Proof. This follows by invoking **??** whose conditions follow from Lemma 6.2.7, Lemma 7.1.1, Lemma 7.1.2 and Lemma 7.1.3. ∎

7.1.1 Proof of Lemma 7.1.1

Lemma 7.1.5. *Suppose* $k \leq n^{1/2-\varepsilon}$. *For all* $U \in \mathcal{I}_{mid}$ *and* $\tau \in \mathcal{M}_U$,

$$\sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \leq \frac{1}{n^{C_p \varepsilon |E(\tau)|}}$$

Proof. This result follows by plugging in the value of $S(\tau)$. Using $k \leq n^{1/2-\varepsilon}$,

$$\begin{aligned} \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) &= \sqrt{n}^{|V(\tau)|-|U_\tau|} \left(\frac{k}{n}\right)^{|V(\tau)|-|U_\tau|} (2(\frac{1}{2} + \frac{1}{2n^{C_p\varepsilon}}) - 1)^{|E(\tau)|} \\ &\leq \frac{1}{n^{C_p \varepsilon |E(\tau)|}} \end{aligned}$$

∎

Corollary 7.1.6. *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 τ is a proper middle shape, we have $w(I_\tau) = 0$ and $w(S_\tau) = w(U_\tau)$. This implies

$$n^{\frac{w(V(\tau))+w(I_\tau)-w(S_\tau)}{2}} = \sqrt{n}^{|V(\tau)|-|U_\tau|}$$

Since τ 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, $q = n^{O(1)\cdot\varepsilon C_V}$. We can set C_V sufficiently small so that, using Lemma 7.1.5,

$$\begin{aligned}
c(\tau)B_{norm}(\tau)S(\tau) &= 100(3D_V)^{|U_\tau\setminus V_\tau|+|V_\tau\setminus U_\tau|+2|E(\tau)|}2^{|V(\tau)\setminus(U_\tau\cup V_\tau)|} \\
&\quad \cdot (6D_V \sqrt[4]{2eq})^{|V(\tau)\setminus U_\tau|+|V(\tau)\setminus V_\tau|}\sqrt{n}^{|V(\tau)|-|U_\tau|}S(\tau) \\
&\leq n^{O(1)\cdot\varepsilon C_V\cdot|E(\tau)|} \cdot \sqrt{n}^{|V(\tau)|-|U_\tau|}S(\tau) \\
&\leq n^{O(1)\cdot\varepsilon C_V\cdot|E(\tau)|} \cdot \frac{1}{n^{C_p\varepsilon|E(\tau)|}} \\
&\leq 1
\end{aligned}$$

■

We can now prove Lemma 7.1.1.

Lemma 7.1.1. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)}H_{Id_U} & B_{norm}(\tau)H_\tau \\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)}H_{Id_U} \end{bmatrix} \succeq 0$$

Proof. We have

$$\begin{aligned}
\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)}H_{Id_U} & B_{norm}(\tau)H_\tau \\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)}H_{Id_U} \end{bmatrix} &= \begin{bmatrix} \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)B_{norm}(\tau)}{|Aut(U)|}\right)H_{Id_U} & 0 \\ 0 & \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)B_{norm}(\tau)}{|Aut(U)|}\right)H_{Id_U} \end{bmatrix} \\
&\quad + B_{norm}(\tau)\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|}H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|}H_{Id_U} \end{bmatrix}
\end{aligned}$$

By Lemma 6.2.9, $\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|}H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|}H_{Id_U} \end{bmatrix} \succeq 0$, so the second term above is positive semidefinite. For the first term, by Lemma 6.2.7, $H_{Id_U} \succeq 0$ and by Corollary 7.1.6,

$\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)B_{norm}(\tau)}{|Aut(U)|} \geq 0$, which proves that the first term is also positive semidefinite. ∎

7.1.2 Proof of Lemma 7.1.2

Lemma 7.1.7. *Suppose $k \leq n^{1/2-\varepsilon}$. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and for all $\gamma \in \Gamma_{U,V}$,*

$$n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \leq \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)|+|E(\gamma)|)}}$$

for some constant B that depends only on C_p. In particular, it is independent of C_V.

Proof. Since γ is a left shape, we have $|U_\gamma| \geq |V_\gamma|$ as V_γ is the unique minimum vertex separator of γ 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 $2|V(\gamma)| - |U_\gamma| - |V_\gamma| = |U_\gamma \setminus V_\gamma| + |V_\gamma \setminus U_\gamma| + 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| \geq |V(\gamma) \setminus (U_\gamma \cap V_\gamma)|$. Therefore,

$$\begin{aligned}
n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 &= n^{|V(\gamma)\setminus U_\gamma)|} \left(\frac{k}{n}\right)^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|} (2(\frac{1}{2} + \frac{1}{2n^{C_p\varepsilon}}) - 1)^{2|E(\gamma)|} \\
&\leq n^{|V(\gamma)|-\frac{|U_\gamma|+|V_\gamma|}{2}} \left(\frac{1}{n^{1/2+\varepsilon}}\right)^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|} \left(\frac{1}{n^{2C_p\varepsilon}}\right)^{|E(\gamma)|} \\
&\leq \left(\frac{1}{n^{\varepsilon}}\right)^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|} \left(\frac{1}{n^{2C_p\varepsilon}}\right)^{|E(\gamma)|} \\
&\leq \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)|+\sum_{e\in E(\gamma)} l_e)}}
\end{aligned}$$

for a constant B that depends only on C_p. ∎

We can now prove Lemma 7.1.2.

Lemma 7.1.2. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

Proof. By Lemma 6.2.11, we have

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} = c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H'_\gamma$$

Using the same proof as in Lemma 6.2.7, we can see that $H'_\gamma \succeq 0$. 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, $q = n^{O(1)\cdot\varepsilon C_V}$. Let B be the constant from Lemma 7.1.7. We can set C_V sufficiently small so that, using Lemma 7.1.7,

$$\begin{aligned}
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} &\leq 100^2 (3D_V)^{2|U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| + 4|E(\alpha)|} 4^{|V(\gamma)\setminus(U_\gamma \cup V_\gamma)|} \\
&\quad \cdot (3D_V)^{4|V(\gamma)\setminus V_\gamma| + 2|V(\gamma)\setminus U_\gamma|} (6D_V \sqrt[4]{2eq})^{2|V(\gamma)\setminus U_\gamma| + 2|V(\gamma)\setminus V_\gamma|} \\
&\quad \cdot n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \cdot D_V^{|U_\gamma \setminus V_\gamma|} \\
&\leq n^{O(1)\cdot\varepsilon C_V \cdot (|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)} \cdot n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \\
&\leq n^{O(1)\cdot\varepsilon C_V \cdot (|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)} \cdot \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)}} \\
&\leq 1
\end{aligned}$$

■

7.1.3 Proof of Lemma 7.1.3

In this section, we will prove Lemma 7.1.3.

Lemma 7.1.3. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U\in\mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq 6 \left(\sum_{U\in\mathcal{I}_{mid}} \sum_{\gamma\in\Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

We use the strategy from [149, Section 10]. We will prove the following lemmas.

Lemma 7.1.8. *Whenever* $\|M_\alpha\| \leq B_{norm}(\alpha)$ *for all* $\alpha \in \mathcal{M}'$,

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{n^{K_1 D^2_{sos}}} Id_{sym}$$

for a constant $K_1 > 0$.

Lemma 7.1.9.

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{n^{K_2 D_{sos}}}{2^{D_V}}$$

for a constant $K_2 > 0$.

If we assume these, we can conclude the following.

Lemma 7.1.3. *Whenever* $\|M_\alpha\| \leq B_{norm}(\alpha)$ *for all* $\alpha \in \mathcal{M}'$,

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

Proof. Let $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$. By Lemma 7.1.8,

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{n^{K_1 D^2_{sos}}} Id_{sym}$$

for a constant $K_1 > 0$. By Lemma 7.1.9,

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{n^{K_2 D_{sos}}}{2^{D_V}}$$

for a constant $K_2 > 0$.

We choose C_{sos} sufficiently small so that $\frac{1}{n^{K_1 D^2_{sos}}} \geq 6\frac{n^{K_2 D_{sos}}}{2^{D_V}}$ which can be satisfied by setting $C_{sos} < K_3 C_V$ for a sufficiently small constant $K_3 > 0$. Then, since $Id_{Sym} \succeq 0$, using

Lemma 7.1.8 and Lemma 7.1.9,

$$\begin{aligned}\sum_{U\in\mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) &\succeq \frac{1}{n^{K_1 D_{sos}^2}} Id_{sym} \\ &\succeq 6\frac{n^{K_2 D_{sos}}}{2^{D_V}} Id_{sym} \\ &\succeq 6\left(\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)}\right) Id_{sym}\end{aligned}$$

■

The rest of the section is devoted to proving Lemma 7.1.8 and Lemma 7.1.9.

In the proofs of both these lemmas, we will need a bound on $B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')$ that is obtained below.

Lemma 7.1.10. *Suppose $k \le n^{1/2-\varepsilon}$. For all $U \in \mathcal{I}_{mid}$ and $\sigma, \sigma' \in \mathcal{L}_U$,*

$$B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') \le \frac{1}{n^{0.5\varepsilon|V(\alpha)|}}$$

Proof. Let $\alpha = \sigma \circ \sigma'$. Observe that $|V(\sigma)| + |V(\sigma')| = |V(\alpha)| + |U|$. By choosing C_V

sufficiently small,

$$\begin{aligned}
B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') &= (6D_V\sqrt[4]{2eq})^{|V(\sigma)\setminus U_\sigma|+|V(\sigma)\setminus V_\sigma|}n^{\frac{w(V(\sigma))-w(U)}{2}} \\
&\quad\cdot(6D_V\sqrt[4]{2eq})^{|V(\sigma')\setminus U_{\sigma'}|+|V(\sigma')\setminus V_{\sigma'}|}n^{\frac{w(V(\sigma'))-w(U)}{2}} \\
&\quad\cdot\frac{1}{|Aut(U)|}\left(\frac{k}{n}\right)^{|V(\alpha)|}(2(\frac{1}{2}+\frac{1}{2n^{C_p\varepsilon}})-1)^{|E(\alpha)|} \\
&\leq n^{O(1)\cdot\varepsilon C_V\cdot|V(\alpha)|}\sqrt{n}^{|V(\sigma)|-|U|}\sqrt{n}^{|V(\sigma')|-|U|}\left(\frac{k}{n}\right)^{|V(\alpha)|}\frac{1}{n^{C_p\varepsilon|E(\alpha)|}} \\
&\leq n^{O(1)\cdot\varepsilon C_V\cdot|V(\alpha)|}\sqrt{n}^{|V(\alpha)|-|U|}\left(\frac{1}{n^{1/2+\varepsilon}}\right)^{|V(\alpha)|}\frac{1}{n^{C_p\varepsilon|E(\alpha)|}} \\
&\leq n^{O(1)\cdot\varepsilon C_V\cdot|V(\alpha)|}\cdot\frac{1}{n^{\varepsilon|V(\alpha)|}}\cdot\frac{1}{n^{C_p\varepsilon|E(\alpha)|}} \\
&\leq \frac{1}{n^{0.5\varepsilon|V(\alpha)|}}
\end{aligned}$$

■

Proof of Lemma 7.1.8

To prove Lemma 7.1.8, we will use the strategy from [149, Section 10]. We will also use the notation from that section. We recall that for $U \in \mathcal{I}_{mid}$, $\mathcal{L}'_U \subset \mathcal{L}_U$ was the set of non-trivial shapes in $\mathcal{L}_U$.

Proposition 7.1.11. *For $V \in \mathcal{I}_{mid}$, $\lambda_V = \left(\frac{k}{n}\right)^{|V|}$.*

Proof. We have $\lambda_V = |Aut(V)|H_{Id_V}(Id_V, Id_V) = \left(\frac{k}{n}\right)^{|V|}$. ■

Corollary 7.1.12. $\lambda_V \geq \frac{1}{n^{O(1)D_{sos}}}$

Lemma 7.1.13. *For any edge $e = (V, U)$ in G, we have*

$$w(e) \leq \frac{n^{O(1)D_{sos}}}{n^{0.1\varepsilon|U|}}$$

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 7.1.10, we have

$$\begin{aligned}
2W(U,V) &= \frac{2}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_V, U_\sigma = U} \sum_{\sigma' \in \mathcal{L}_V, U_{\sigma'} \neq V} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \\
&\leq \frac{2}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_V, U_\sigma = U} \sum_{\sigma' \in \mathcal{L}_V, U_{\sigma'} \neq V} \frac{1}{n^{0.5\varepsilon |V(\sigma \circ \sigma')|}} \\
&\leq \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{2}{n^{0.5\varepsilon |V(\sigma \circ \sigma')|}} \\
&\leq \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{D_{sos}^{D_{sos}}}{n^{0.1\varepsilon |V(\sigma \circ \sigma')|} D_{sos}^{D_{sos}} n^{F\varepsilon |V(\sigma \circ \sigma')|}} \\
&\leq \frac{D_{sos}^{D_{sos}}}{n^{0.1\varepsilon |U|}} \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{1}{D_{sos}^{D_{sos}} n^{F\varepsilon |V(\sigma \circ \sigma')|}} \\
&\leq \frac{D_{sos}^{D_{sos}}}{n^{0.1\varepsilon |U|}} \\
&\leq \frac{\lambda_V n^{O(1) D_{sos}}}{n^{0.1\varepsilon |U|}}
\end{aligned}$$

where we set C_V small enough so that $0.4 \geq F$. This proves the lemma. ∎

Corollary 7.1.14. *For any $U, V \in \mathcal{I}_{mid}$ such that $w(U) > w(V)$,*

$$\sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \leq n^{O(1) D_{sos}^2}$$

Proof. The total number of vertices in G is at most $D_{sos} + 1$ since each $U \in \mathcal{I}_{mid}$ 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

$$w(e) \leq \frac{n^{O(1) D_{sos}}}{n^{0.1\varepsilon |U'|}} \leq \frac{n^{O(1) D_{sos}}}{n^{0.1\varepsilon}}$$

So, $\prod_{e\in E(P)} w(e) \leq \left(\frac{n^{O(1)D_{sos}}}{n^{0.1\varepsilon}}\right)^j$. Therefore, by setting C_{sos} small enough,

$$\begin{aligned}\sum_{P:P \text{ is a path from V to U in G}} \prod_{e\in E(P)} w(e) &\leq \sum_{j=1}^{D_{sos}} (D_{sos}+1)^j \left(\frac{n^{O(1)D_{sos}}}{n^{0.1\varepsilon}}\right)^j \\ &\leq n^{O(1)D_{sos}^2}\end{aligned}$$

■

We can now prove Lemma 7.1.8.

Lemma 7.1.8. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U\in\mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{n^{K_1 D_{sos}^2}} Id_{sym}$$

for a constant $K_1 > 0$.

Proof. For all $V \in \mathcal{I}_{mid}$, we have

$$Id_{Sym,V} \preceq 2 \sum_{U\in\mathcal{I}_{mid}:w(U)\geq w(V)} \left(\sum_{P:P \text{ is a path from V to U in } G} \prod_{e\in E(P)} w(e)\right) \frac{1}{\lambda_U} M^{fact}(H_{Id_U})$$

Summing this over all $V \in \mathcal{I}_{mid}$, we get

$$\begin{aligned}Id_{Sym} &\preceq \sum_{U\in\mathcal{I}_{mid}} \frac{2}{\lambda_U} \left(\sum_{V\in\mathcal{I}_{mid}:w(U)\geq w(V)} \sum_{P:P \text{ is a path from V to U in G}} \prod_{e\in E(P)} w(e)\right) M^{fact}(H_{Id_U}) \\ &\preceq \sum_{U\in\mathcal{I}_{mid}} \frac{2}{\lambda_U} \left(\sum_{V\in\mathcal{I}_{mid}:w(U)\geq w(V)} n^{O(1)D_{sos}^2}\right) M^{fact}(H_{Id_U})\end{aligned}$$

For any fixed $U \in \mathcal{I}_{mid}$, the number of $V \in \mathcal{I}_{mid}$ such that $w(U) \geq w(V)$ is at most $D_{sos}+1$.

Also, $\lambda_U \geq \frac{1}{d^{O(1)D_{sos}}}$ for all $U \in \mathcal{I}_{mid}$. Therefore,

$$\begin{aligned} Id_{Sym} &\preceq \sum_{U \in \mathcal{I}_{mid}} \frac{2}{\lambda_U}(D_{sos}+1)n^{O(1)D_{sos}^2} M^{fact}(H_{Id_U}) \\ &\preceq \sum_{U \in \mathcal{I}_{mid}} n^{O(1)D_{sos}^2} M^{fact}(H_{Id_U}) \end{aligned}$$

where we used the fact that for all $U \in \mathcal{I}_{mid}$, $M^{fact}(H_{Id_U}) \succeq 0$. ∎

Proof of Lemma 7.1.9

We restate the lemma for convenience.

Lemma 7.1.9.

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{n^{K_2 D_{sos}}}{2^{D_V}}$$

for a constant $K_2 > 0$.

Proof. We have

$$\begin{aligned} &\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\ &\quad = \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{1}{|Aut(U)|c(\gamma)} \sum_{\substack{\sigma,\sigma' \in \mathcal{L}_{U_\gamma} : |V(\sigma)| \leq D_V, |V(\sigma')| \leq D_V, \\ |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V}} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_{U_\gamma}}(\sigma, \sigma') \end{aligned}$$

The set of σ, σ' that could appear in the above sum must necessarily be non-trivial and hence, $\sigma, \sigma' \in \mathcal{L}'_U$. Then,

$$\begin{aligned} &\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\ &\quad = \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma,\sigma' \in \mathcal{L}'_U} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \sum_{\gamma \in \Gamma_{U,*} : |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V} \frac{1}{|Aut(U)|c(\gamma)} \end{aligned}$$

For $\sigma \in \mathcal{L}'_U$, define $m_\sigma = D_V + 1 - |V(\sigma)| \geq 1$. This is precisely set so that for all $\gamma \in \Gamma_{U,*}$, we have $|V(\sigma \circ \gamma)| > D_V$ if and only if $|V(\gamma)| \geq |U| + m_\sigma$. So, for $\sigma, \sigma' \in \mathcal{L}'_U$,

$$\begin{aligned}\sum_{\gamma \in \Gamma_{U,*}: |V(\sigma\circ\gamma)|>D_V \text{ or } |V(\sigma'\circ\gamma)|>D_V} \frac{1}{|Aut(U)|c(\gamma)} &= \sum_{\gamma \in \Gamma_{U,*}: |V(\gamma)| \geq |U| + \min(m_\sigma, m_{\sigma'})} \frac{1}{|Aut(U)|c(\gamma)} \\ &\leq \frac{1}{2^{\min(m_\sigma, m_{\sigma'})-1}}\end{aligned}$$

Also, for $\sigma, \sigma' \in \mathcal{L}'_U$, we have $|V(\sigma \circ \sigma')| + min(m_\sigma, m_{\sigma'}) - 1 \geq D_V$. Therefore,

$$\begin{aligned}\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \frac{1}{2^{\min(m_\sigma, m_{\sigma'})-1}} \\ &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)} D_{sos}}{n^{0.5\varepsilon |V(\sigma\circ\sigma')|} 2^{\min(m_\sigma, m_{\sigma'})-1}}\end{aligned}$$

where we used Lemma 7.1.10. Using $n^{0.5\varepsilon|V(\sigma\circ\sigma')|} \geq n^{0.1\varepsilon|V(\sigma\circ\sigma')|} 2^{|V(\sigma\circ\sigma')|}$,

$$\begin{aligned}\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)} D_{sos}}{n^{0.1\varepsilon |V(\sigma\circ\sigma')|} 2^{|V(\sigma\circ\sigma')|} 2^{\min(m_\sigma, m_{\sigma'})-1}} \\ &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)} D_{sos}}{n^{0.1\varepsilon |V(\sigma\circ\sigma')|} 2^{D_V}} \\ &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{n^{O(1)} D_{sos}}{D_{sos}^{D_{sos}} n^{0.1\varepsilon |V(\sigma\circ\sigma')|} 2^{D_V}}\end{aligned}$$

The final step will be to argue that $\sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{1}{D_{sos}^{D_{sos}} n^{0.1\varepsilon|V(\sigma\circ\sigma')|}} \leq 1$ which will complete the proof. But this will follow if we set C_V small enough. ∎

7.2 Tensor PCA: Full verification

In this section, we will prove all the bounds required to prove Theorem 4.3.1.

Theorem 4.3.1. *Let $k \geq 2$ be an integer. There exist constants $C, C_\Delta > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $\lambda \leq n^{\frac{k}{4}-\varepsilon}$ and $\Delta = n^{-C_\Delta \varepsilon}$ then with high probability, the candidate moment matrix Λ given by pseudo-calibration for degree $n^{C\varepsilon}$ Sum-of-Squares is PSD.*

In particular, we will use **??** where we choose ε in the theorem, not to be confused with the ε in Theorem 4.3.1, to be an arbitrarily small constant.

To invoke the machinery, we basically have to verify the following conditions, the qualitiative versions of which have already been shown in Section 6.2.

Lemma 7.2.1. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

Lemma 7.2.2. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

Lemma 7.2.3. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

Corollary 7.2.4. *With constant probability, $\Lambda \succeq 0$.*

Proof. This follows by invoking **??** whose conditions follow from Lemma 6.3.10, Lemma 7.2.1, Lemma 7.2.2 and Lemma 7.2.3. ∎

7.2.1 Proof of Lemma 7.2.1

Lemma 7.2.5. *Suppose $\lambda \leq n^{\frac{k}{4}-\varepsilon}$. For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, suppose $deg^{\tau}(i)$ is even for all $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, then*

$$\sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \leq \frac{1}{n^{0.5\varepsilon \sum_{e \in E(\tau)} l_e}}$$

Proof. Firstly, we claim that $\sum_{e \in E(\tau)} kl_e \geq 2(|V(\tau)| - |U_\tau|)$. For any vertex $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, $deg^{\tau}(i)$ is even and is not 0, hence, $deg^{\tau}(i) \geq 2$. Any vertex $i \in U_\tau \setminus V_\tau$ cannot have $deg^{\tau}(i) = 0$ otherwise $U_\tau \setminus \{i\}$ is a vertex separator of strictly smaller weight than U_τ, which is not possible, hence, $deg^{\tau}(i) \geq 1$. Therefore,

$$\begin{aligned}
\sum_{e \in E(\tau)} kl_e &= \sum_{i \in V(\tau)} deg^{\tau}(i) \\
&\geq \sum_{i \in V(\tau) \setminus U_\tau \setminus V_\tau} deg^{\tau}(i) + \sum_{i \in U_\tau \setminus V_\tau} deg^{\tau}(i) + \sum_{i \in V_\tau \setminus U_\tau} deg^{\tau}(i) \\
&\geq 2|V(\tau) \setminus U_\tau \setminus V_\tau| + |U_\tau \setminus V_\tau| + |V_\tau \setminus U_\tau| \\
&= 2(|V(\tau)| - |U_\tau|)
\end{aligned}$$

By choosing C_Δ sufficiently small, we have

$$\begin{aligned}
\sqrt{n}^{|V(\tau)|-|U_\tau|}S(\tau) &= \sqrt{n}^{|V(\tau)|-|U_\tau|}\Delta^{|V(\tau)|-|U_\tau|}\prod_{e\in E(\tau)}\left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e} \\
&\leq \sqrt{n}^{|V(\tau)|-|U_\tau|}\Delta^{|V(\tau)|-|U_\tau|}\prod_{e\in E(\tau)} n^{(-\frac{k}{4}-0.5\varepsilon)l_e} \\
&= \sqrt{n}^{|V(\tau)|-|U_\tau|-\frac{\sum_{e\in E(\tau)} kl_e}{2}}\Delta^{|V(\tau)|-|U_\tau|}\prod_{e\in E(\tau)} n^{-0.5\varepsilon l_e} \\
&= \Delta^{|V(\tau)|-|U_\tau|}\prod_{e\in E(\tau)} n^{-0.5\varepsilon l_e} \\
&\leq \frac{1}{n^{0.5\varepsilon\sum_{e\in E(\tau)} l_e}}
\end{aligned}$$

■

Corollary 7.2.6. *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 τ is a proper middle shape, we have $w(I_\tau) = 0$ and $w(S_{\tau,min}) = w(U_\tau)$. This implies

$$n^{\frac{w(V(\tau))+w(I_\tau)-w(S_{\tau,min})}{2}} = \sqrt{n}^{|V(\tau)|-|U_\tau|}$$

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 $deg^\tau(i)$ is even for all $i \in V(\tau) \setminus U_\tau \setminus V_\tau$. As was observed in the proof of Lemma 7.2.5, 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, $|E(\tau)| \leq \sum_{e\in E(\tau)} l_e$ and $q = n^{O(1)\cdot\varepsilon(C_V+C_E)}$.

We can set C_V, C_E sufficiently small so that, using Lemma 7.2.5,

$$\begin{aligned}
c(\tau)B_{norm}(\tau)S(\tau) &= 100(3D_V)^{|U_\tau \setminus V_\tau|+|V_\tau \setminus U_\tau|+k|E(\tau)|}2^{|V(\tau)\setminus(U_\tau \cup V_\tau)|} \\
&\quad \cdot 2e(6qD_V)^{|V(\tau)\setminus U_\tau|+|V(\tau)\setminus V_\tau|} \prod_{e\in E(\tau)} (400D_V^2 D_E^2 q)^{l_e} \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \\
&\leq n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot\sum_{e\in E(\tau)} l_e} \cdot \sqrt{n}^{|V(\tau)|-|U_\tau|} S(\tau) \\
&\leq n^{O(1)\cdot\varepsilon(C_V+C_E)\cdot\sum_{e\in E(\tau)} l_e} \cdot \frac{1}{n^{0.5\varepsilon\sum_{e\in E(\tau)} l_e}} \\
&\leq 1
\end{aligned}$$

■

We can now prove Lemma 7.2.1.

Lemma 7.2.1. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau)H_\tau \\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

Proof. We have

$$\begin{aligned}
\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau)H_\tau \\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} &= \begin{bmatrix} \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)B_{norm}(\tau)}{|Aut(U)|}\right) H_{Id_U} & 0 \\ 0 & \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)B_{norm}(\tau)}{|Aut(U)|}\right) H_{Id_U} \end{bmatrix} \\
&\quad + B_{norm}(\tau) \begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix}
\end{aligned}$$

By Lemma 6.3.12, $\begin{bmatrix} \frac{S(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} \succeq 0$, so the second term above is positive semidefinite. For the first term, by Lemma 6.3.10, $H_{Id_U} \succeq 0$ and by Corollary 7.2.6, $\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)B_{norm}(\tau)}{|Aut(U)|} \geq 0$, which proves that the first term is also positive semidefinite. ■

7.2.2 Proof of Lemma 7.2.2

Lemma 7.2.7. *Suppose $\lambda \le n^{\frac{k}{4}-\varepsilon}$. For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and for all $\gamma \in \Gamma_{U,V}$,*

$$n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \le \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)}}$$

for some constant B that depends only on C_Δ. 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 $deg^\gamma(i)$ is odd, then $S(\gamma) = 0$ and the inequality is true. So, assume $deg^\gamma(i)$ is even for all vertices $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$.

We first claim that $k\sum_{e\in E(\gamma)} l_e \ge 2|V(\gamma) \setminus U_\gamma|$. Since γ is a left shape, all vertices i in $V(\gamma) \setminus U_\gamma$ have $deg^\gamma(i) \ge 1$. In particular, all vertices $i \in V_\gamma \setminus U_\gamma$ have $deg^\gamma(i) \ge 1$. Moreover, if $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$, since $deg^\gamma(i)$ is even, we must have $deg^\gamma(i) \ge 2$.

Let S' be the set of vertices $i \in U_\gamma \setminus V_\gamma$ that have $deg^\gamma(i) \ge 1$. Then, note that $|S'| + |U_\gamma \cap V_\gamma| \ge |V_\gamma| \implies |S'| \ge |V_\gamma \setminus U_\gamma|$ since otherwise $S' \cup (U_\gamma \cap V_\gamma)$ will be a vertex separator of γ of weight strictly less than V_γ, which is not possible. Then,

$$\begin{aligned}
\sum_{e\in E(\gamma)} kl_e &= \sum_{i\in V(\gamma)} deg^\gamma(i) \\
&\ge \sum_{i\in V(\gamma)\setminus U_\gamma\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in U_\gamma\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in V_\gamma\setminus U_\gamma} deg^\gamma(i) \\
&\ge 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| + |S'| + |V_\gamma \setminus U_\gamma| \\
&\ge 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| \\
&= 2|V(\gamma) \setminus U_\gamma|
\end{aligned}$$

Finally, note that $2|V(\gamma)| - |U_\gamma| - |V_\gamma| = |U_\gamma \setminus V_\gamma| + |V_\gamma \setminus U_\gamma| + 2|V(\gamma) \setminus U_\gamma \setminus V_\gamma| \ge$

$|V(\gamma) \setminus (U_\gamma \cap V_\gamma)|$. By choosing C_Δ sufficiently small, we have

$$\begin{aligned} n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 &= n^{|V(\gamma)\setminus U_\gamma|} \Delta^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|} \prod_{e\in E(\gamma)} \left(\frac{\lambda^2}{(\Delta n)^k}\right)^{l_e} \\ &\leq n^{|V(\gamma)\setminus U_\gamma|} \Delta^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|} \prod_{e\in E(\gamma)} n^{-(\frac{k}{2}+\varepsilon)l_e} \\ &\leq \Delta^{2|V(\gamma)|-|U_\gamma|-|V_\gamma|} \prod_{e\in E(\gamma)} n^{-\varepsilon l_e} \\ &\leq \frac{1}{n^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)} l_e)}} \end{aligned}$$

for a constant B that depends only on C_Δ. ∎

Remark 7.2.8. *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_Δ in the planted distribution was to introduce this decay, which is needed in the current machinery.*

We can now prove Lemma 7.2.2.

Lemma 7.2.2. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

Proof. By Lemma 6.3.14, we have

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H'_\gamma$$

Using the same proof as in Lemma 6.3.10, we can see that $H'_\gamma \succeq 0$. 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 $q = n^{O(1) \cdot \varepsilon(C_V + C_E)}$. Let B be the constant from Lemma 7.2.7. We can set C_V, C_E sufficiently small so that, using Lemma 7.2.7,

$$\begin{aligned}
c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} &\leq 100^2 (3D_V)^{2|U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| + 2k|E(\alpha)|} 4^{|V(\gamma) \setminus (U_\gamma \cup V_\gamma)|} \\
&\quad \cdot (3D_V)^{4|V(\gamma) \setminus V_\gamma| + 2|V(\gamma) \setminus U_\gamma|} (6qD_V)^{2|V(\gamma) \setminus U_\gamma| + 2|V(\gamma) \setminus V_\gamma|} \prod_{e \in E(\gamma)} (400 D_V^2 D_E^2 q)^{2l_e} \\
&\quad \cdot n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 \cdot D_V^{|U_\gamma \setminus V_\gamma|} \\
&\leq n^{O(1) \cdot \varepsilon(C_V + C_E) \cdot (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)} \cdot n^{w(V(\gamma) \setminus U_\gamma)} S(\gamma)^2 \\
&\leq n^{O(1) \cdot \varepsilon(C_V + C_E) \cdot (|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)} \cdot \frac{1}{n^{B\varepsilon(|V(\gamma) \setminus (U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}} \\
&\leq 1
\end{aligned}$$

■

7.2.3 Proof of Lemma 7.2.3

In this section, we will prove Lemma 7.2.3 using the strategy sketched in [149, Section 10].

Lemma 7.2.3. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)| c(\gamma)} \right) Id_{sym}$$

In particular, we prove the following lemmas.

Lemma 7.2.9. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq \frac{\Delta^{2D_{sos}^2}}{n^{D_{sos}}} Id_{sym}$$

Lemma 7.2.10.

$$\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}\leq\frac{1}{\Delta^{2D_{sos}}2^{D_V}}$$

Assuming these, we can conclude the following.

Lemma 7.2.3. *Whenever $\|M_\alpha\|\leq B_{norm}(\alpha)$ for all $\alpha\in\mathcal{M}'$,*

$$\sum_{U\in\mathcal{I}_{mid}}M^{fact}_{Id_U}(H_{Id_U})\succeq 6\left(\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H'_\gamma,H_{Id_U})}{|Aut(U)|c(\gamma)}\right)Id_{sym}$$

Proof. We choose C_{sos} sufficiently small so that $\frac{\Delta^{2D^2_{sos}}}{n^{D_{sos}}}\geq\frac{6}{\Delta^{2D_{sos}}2^{D_V}}$ which is satisfied by setting $C_{sos}<0.5C_V$. Then, since $Id_{Sym}\succeq 0$, using Lemma 7.2.9 and Lemma 7.2.10,

$$\begin{aligned}\sum_{U\in\mathcal{I}_{mid}}M^{fact}_{Id_U}(H_{Id_U})&\succeq\frac{\Delta^{2D^2_{sos}}}{n^{D_{sos}}}Id_{sym}\\&\succeq\frac{6}{\Delta^{2D_{sos}}2^{D_V}}Id_{sym}\\&\succeq 6\left(\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H'_\gamma,H_{Id_U})}{|Aut(U)|c(\gamma)}\right)Id_{sym}\end{aligned}$$

■

The rest of the section is devoted to proving Lemma 7.2.9 and Lemma 7.2.10.

In the proofs of both these lemmas, we will need a bound on $B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')$ that is obtained below.

Lemma 7.2.11. *Suppose $\lambda=n^{\frac{k}{4}-\varepsilon}$. For all $U\in\mathcal{I}_{mid}$ and $\sigma,\sigma'\in\mathcal{L}_U$,*

$$B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')\leq\frac{1}{n^{0.5\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}n|U|}}$$

Proof. Suppose there is a vertex $i\in V(\sigma)\setminus V_\sigma$ such that $deg^\sigma(i)+deg^{U_\sigma}(i)$ is odd, then $H_{Id_U}(\sigma,\sigma')=0$ and the inequality is true. So, assume that $deg^\sigma(i)+deg^{U_\sigma}(i)$ is even for

all $i \in V(\sigma) \setminus V_\sigma$. Similarly, assume that $deg^{\sigma'}(i) + deg^{U_{\sigma'}}(i)$ 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'd be done. So, assume $\rho_\sigma = \rho_{\sigma'}$.

Let $\alpha = \sigma \circ \sigma'$. We will first prove that $\sum_{e \in E(\alpha)} kl_e + 2deg(\alpha) \geq 2|V(\alpha)| + 2|U|$. 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

$$\begin{aligned}
\sum_{e \in E(\alpha)} kl_e + 2deg(\alpha) &\geq 2|V(\alpha) \setminus (U_\alpha \cup V_\alpha)| + 2(|U| - |U_\alpha \cap V_\alpha|) + 2(|U_\alpha| + |V_\alpha|) \\
&= 2(|V(\alpha)| - |U_\alpha \cup V_\alpha|) + 2(|U| - |U_\alpha \cap V_\alpha|) + 2(|U_\alpha \cup V_\alpha| + |U_\alpha \cap V_\alpha|) \\
&= 2|V(\alpha)| + 2|U|
\end{aligned}$$

where we used the fact that $U_\alpha \cap V_\alpha \subseteq U$. Finally, by choosing C_V, C_E sufficiently small,

$$\begin{aligned}
B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma, \sigma') &= 2e(6qD_V)^{|V(\sigma)\setminus U_\sigma| + |V(\sigma)\setminus V_\sigma|} \prod_{e \in E(\sigma)} (400 D_V^2 D_E^2 q)^{l_e} n^{\frac{w(V(\sigma)) - w(U)}{2}} \\
&\quad \cdot 2e(6qD_V)^{|V(\sigma')\setminus U_{\sigma'}| + |V(\sigma')\setminus V_{\sigma'}|} \prod_{e \in E(\sigma')} (400 D_V^2 D_E^2 q)^{l_e} n^{\frac{w(V(\sigma')) - w(U)}{2}} \\
&\quad \cdot \frac{1}{|Aut(U)|} \Delta^{|V(\alpha)|} \left(\frac{1}{\sqrt{\Delta n}}\right)^{deg(\alpha)} \prod_{e \in E(\alpha)} \left(\frac{\lambda}{(\Delta n)^{\frac{k}{2}}}\right)^{l_e} \\
&\leq n^{O(1) \cdot \varepsilon(C_V + C_E) \cdot (|V(\alpha)| + \sum_{e \in E(\alpha)} l_e)} \Delta^{|V(\alpha)|} \left(\frac{1}{\sqrt{\Delta}}\right)^{deg(\alpha)} \\
&\quad \cdot \sqrt{n}^{|V(\alpha)| - |U|} \left(\frac{1}{\sqrt{n}}\right)^{deg(\alpha)} \prod_{e \in E(\alpha)} n^{(-\frac{k}{4} - 0.5\varepsilon) l_e} \\
&\leq \frac{n^{O(1) \cdot \varepsilon(C_V + C_E) \cdot (|V(\alpha)| + \sum_{e \in E(\alpha)} l_e)}}{n^{\varepsilon C_\Delta |V(\alpha)| n^{0.5\varepsilon \sum_{e \in E(\alpha)} l_e}}} \cdot \frac{1}{\Delta^{D_{sos}} n^{|U|}} \sqrt{n}^{|V(\alpha)| + |U| - deg(\alpha) - \frac{1}{2}\sum_{e \in E(\alpha)} kl_e} \\
&\leq \frac{1}{n^{0.5\varepsilon C_\Delta |V(\alpha)| \Delta^{D_{sos}} n^{|U|}}}
\end{aligned}$$

where we used the facts $\Delta \leq 1, deg(\alpha) \leq 2D_{sos}$. ■

Proof of Lemma 7.2.9

To prove Lemma 7.2.9, we will use the strategy from [149, Section 10]. We will also use the notation from that section. We recall that for $U \in \mathcal{I}_{mid}$, $\mathcal{L}'_U \subset \mathcal{L}_U$ was the set of non-trivial shapes in $\mathcal{L}_U$.

Proposition 7.2.12. *For $V \in \mathcal{I}_{mid}$, $\lambda_V = \frac{1}{n^{|V|}}$.*

Proof. We have $\lambda_V = |Aut(V)| H_{Id_V}(Id_V, Id_V) = \Delta^{|V|} \left(\frac{1}{\sqrt{\Delta n}}\right)^{2|V|} = \frac{1}{n^{|V|}}$. ∎

Lemma 7.2.13. *For any edge $e = (V, U)$ in G, we have*

$$w(e) \leq \frac{1}{n^{0.1C_\Delta \varepsilon |U|} \Delta^{2D_{sos}}}$$

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 7.2.11, we have

$$\begin{aligned}
2W(U,V) &= \frac{2}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_V, U_\sigma = U} \sum_{\sigma' \in \mathcal{L}_V, U_{\sigma'} \neq V} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \\
&\leq \frac{2}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_V, U_\sigma = U} \sum_{\sigma' \in \mathcal{L}_V, U_{\sigma'} \neq V} \frac{1}{n^{0.5\varepsilon C_\Delta |V(\sigma \circ \sigma')| \Delta^{D_{sos}} n^{|V|}}} \\
&\leq \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{2}{n^{0.5\varepsilon C_\Delta |V(\sigma \circ \sigma')| \Delta^{D_{sos}} n^{|V|}}} \\
&\leq \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{1}{n^{0.1C_\Delta \varepsilon |V(\sigma \circ \sigma')|} D_{sos}^{D_{sos}} n^{F\varepsilon |V(\sigma \circ \sigma')|} \Delta^{2D_{sos}} n^{|V|}} \\
&\leq \frac{1}{n^{0.1C_\Delta \varepsilon |U|} \Delta^{2D_{sos}} n^{|V|}} \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{1}{D_{sos}^{D_{sos}} n^{F\varepsilon |V(\sigma \circ \sigma')|}} \\
&\leq \frac{1}{n^{0.1C_\Delta \varepsilon |U|} \Delta^{2D_{sos}} n^{|V|}} \\
&= \frac{\lambda_V}{n^{0.1\varepsilon C_\Delta |U|} \Delta^{2D_{sos}}}
\end{aligned}$$

where we set C_V, C_E small enough so that $0.4C_\Delta \geq F$. This proves the lemma. ∎

Corollary 7.2.14. *For any $U, V \in \mathcal{I}_{mid}$ such that $w(U) > w(V)$,*

$$\sum_{P:P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \leq \frac{1}{2D_{sos}}$$

Proof. The total number of vertices in G is at most $D_{sos}+1$ since each $U \in \mathcal{I}_{mid}$ 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

$$w(e) \leq \frac{1}{n^{0.1C_\Delta \varepsilon |U'| \Delta^{2D_{sos}}}} \leq \frac{1}{n^{0.1C_\Delta \varepsilon \Delta^{2D_{sos}}}}$$

So, $\prod_{e \in E(P)} w(e) \leq \left(\frac{1}{n^{0.1C_\Delta \varepsilon \Delta^{2D_{sos}}}}\right)^j$. Therefore, by setting C_{sos} small enough,

$$\begin{aligned}\sum_{P:P \text{ is a path from V to U in G}} \prod_{e \in E(P)} w(e) &\leq \sum_{j=1}^{D_{sos}} (D_{sos}+1)^j \left(\frac{1}{n^{0.1C_\Delta \varepsilon \Delta^{2D_{sos}}}}\right)^j \\ &\leq \frac{1}{2D_{sos}\Delta^{2D_{sos}^2}}\end{aligned}$$

■

We can now prove Lemma 7.2.9.

Lemma 7.2.9. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{\Delta^{2D_{sos}^2}}{n^{D_{sos}}} Id_{sym}$$

Proof. For all $V \in \mathcal{I}_{mid}$, we have

$$Id_{Sym,V} \preceq 2 \sum_{U \in \mathcal{I}_{mid}: w(U) \geq w(V)} \left(\sum_{P:P \text{ is a path from V to U in } G} \prod_{e \in E(P)} w(e) \right) \frac{1}{\lambda_U} M^{fact}(H_{Id_U})$$

Summing this over all $V \in \mathcal{I}_{mid}$, we get

$$\begin{aligned} Id_{Sym} &\preceq \sum_{U \in \mathcal{I}_{mid}} \frac{2}{\lambda_U} \left(\sum_{V \in \mathcal{I}_{mid}: w(U) \geq w(V)} \sum_{P: P \text{ is a path from V to U in G}} \prod_{e \in E(P)} w(e) \right) M^{fact}(H_{Id_U}) \\ &\preceq \sum_{U \in \mathcal{I}_{mid}} \frac{2}{\lambda_U} \left(\sum_{V \in \mathcal{I}_{mid}: w(U) \geq w(V)} \frac{1}{2D_{sos}\Delta^{2D_{sos}^2}} \right) M^{fact}(H_{Id_U}) \end{aligned}$$

For any fixed $U \in \mathcal{I}_{mid}$, the number of $V \in \mathcal{I}_{mid}$ such that $w(U) \geq w(V)$ is at most D_{sos}. Therefore,

$$\begin{aligned} Id_{Sym} &\preceq \sum_{U \in \mathcal{I}_{mid}} \frac{1}{\lambda_U \Delta^{2D_{sos}^2}} M^{fact}(H_{Id_U}) \\ &= \sum_{U \in \mathcal{I}_{mid}} \frac{1}{\Delta^{2D_{sos}^2}} n^{|U|} M^{fact}(H_{Id_U}) \\ &\preceq \sum_{U \in \mathcal{I}_{mid}} \frac{1}{\Delta^{2D_{sos}^2}} n^{D_{sos}} M^{fact}(H_{Id_U}) \end{aligned}$$

where we used the fact that for all $U \in \mathcal{I}_{mid}$, we have $|U| \leq D_{sos}$ and $M^{fact}(H_{Id_U}) \succeq 0$. ■

Proof of Lemma 7.2.10

We restate the lemma for convenience.

Lemma 7.2.10.

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)| c(\gamma)} \leq \frac{1}{\Delta^{2D_{sos}} 2^{D_V}}$$

Proof. We have

$$\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}$$
$$=\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{1}{|Aut(U)|c(\gamma)}\sum_{\substack{\sigma,\sigma'\in\mathcal{L}_{U_\gamma}:|V(\sigma)|\le D_V,|V(\sigma')|\le D_V,\\ |V(\sigma\circ\gamma)|>D_V \text{ or } |V(\sigma'\circ\gamma)|>D_V}} B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_{U_\gamma}}(\sigma,\sigma')$$

The set of σ,σ' that could appear in the above sum must necessarily be non-trivial and hence, $\sigma,\sigma'\in\mathcal{L}'_U$. Then,

$$\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}$$
$$=\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')\sum_{\gamma\in\Gamma_{U,*}:|V(\sigma\circ\gamma)|>D_V \text{ or } |V(\sigma'\circ\gamma)|>D_V}\frac{1}{|Aut(U)|c(\gamma)}$$

For $\sigma\in\mathcal{L}'_U$, define $m_\sigma=D_V+1-|V(\sigma)|\ge 1$. This is precisely set so that for all $\gamma\in\Gamma_{U,*}$, we have $|V(\sigma\circ\gamma)|>D_V$ if and only if $|V(\gamma)|\ge|U|+m_\sigma$. So, for $\sigma,\sigma'\in\mathcal{L}'_U$,

$$\sum_{\gamma\in\Gamma_{U,*}:|V(\sigma\circ\gamma)|>D_V \text{ or } |V(\sigma'\circ\gamma)|>D_V}\frac{1}{|Aut(U)|c(\gamma)}=\sum_{\gamma\in\Gamma_{U,*}:|V(\gamma)|\ge|U|+\min(m_\sigma,m_{\sigma'})}\frac{1}{|Aut(U)|c(\gamma)}$$
$$\le\frac{1}{2^{\min(m_\sigma,m_{\sigma'})-1}}$$

Also, for $\sigma,\sigma'\in\mathcal{L}'_U$, we have $|V(\sigma\circ\sigma')|+min(m_\sigma,m_{\sigma'})-1\ge D_V$. Therefore,

$$\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)}\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')\frac{1}{2^{\min(m_\sigma,m_{\sigma'})-1}}$$
$$\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{\eta^{0.5\varepsilon C_\Delta|V(\sigma\circ\sigma')|_\Delta D_{sos}}|U|2^{\min(m_\sigma,m_{\sigma'})-1}}$$
$$\le\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{\eta^{0.5\varepsilon C_\Delta|V(\sigma\circ\sigma')|_\Delta D_{sos}}2^{\min(m_\sigma,m_{\sigma'})-1}}$$

where we used Lemma 7.2.11. Using $n^{0.5C_\Delta|V(\sigma\circ\sigma')|} \geq n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}2^{|V(\sigma\circ\sigma')|}$,

$$\begin{aligned}\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}}\frac{d_{Id_U}(H_{Id_U},H'_\gamma)}{|Aut(U)|c(\gamma)} &\leq \sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}}2^{|V(\sigma\circ\sigma')|}2^{\min(m_\sigma,m_{\sigma'})-1}}\\ &\leq \sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{D_{sos}}2^{D_V}}\\ &\leq \sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{D_{sos}^{D_{sos}}n^{0.1\varepsilon C_\Delta|V(\sigma\circ\sigma')|}\Delta^{2D_{sos}}2^{D_V}}\end{aligned}$$

where we set C_{sos} small enough so that $D_{sos} = n^{\varepsilon C_{sos}} \leq n^{c\varepsilon C_\Delta} = \frac{1}{\Delta}$. The final step will be to argue that $\sum_{U\in\mathcal{I}_{mid}}\sum_{\sigma,\sigma'\in\mathcal{L}'_U}\frac{1}{D_{sos}^{D_{sos}}n^{0.1C_\Delta\varepsilon|V(\sigma\circ\sigma')|}} \leq 1$ which will complete the proof. But this will follow if we set C_V, C_E small enough. ∎

7.3 Sparse PCA: Full verification

In this section, we will prove all the bounds required to prove Theorem 4.4.1.

Theorem 4.4.1. *There exists a constant $C > 0$ such that for all sufficiently small constants $\varepsilon > 0$, if $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}, m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, and there exists a constant A such that $0 < A < \frac{1}{4}$, $d^{4A} \leq k \leq d^{1-A\varepsilon}$, and $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$, then with high probability, the candidate moment matrix Λ given by pseudo-calibration for degree $d^{C\varepsilon}$ Sum-of-Squares is PSD.*

In particular, we will use **??** where we choose ε in the theorem, not to be confused with the ε in Theorem 4.4.1, to be an arbitrarily small constant.

To invoke the machinery, we basically have to verify the following conditions, the qualitiative versions of which have already been shown in Section 6.2.

Definition 7.3.1. *Define $n = \max(d, m)$.*

Remark 7.3.2. *The above definition conforms with the notation used in **??**. So, we can use the bounds as stated there.*

Lemma 7.3.3. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix} \succeq 0$$

Lemma 7.3.4. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H_{Id_V}^{-\gamma,\gamma} \preceq H'_\gamma$$

Lemma 7.3.5. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M_{Id_U}^{fact}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

Corollary 7.3.6. *With high probability, $\Lambda \succeq 0$.*

Proof. This follows by invoking **??** whose conditions follow from Lemma 6.4.11, Lemma 7.3.3, Lemma 7.3.4, and Lemma 7.3.5. ∎

7.3.1 Proof of Lemma 7.3.3

Lemma 7.3.7. *Suppose $0 < A < \frac{1}{4}$ is a constant such that $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}$ and $\frac{1}{\sqrt{k}} \leq d^{-2A}$. For all m such that $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}, m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, for all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, suppose $deg^\tau(i)$ is even for all $i \in V(\tau) \setminus U_\tau \setminus V_\tau$, then*

$$\sqrt{d}^{|\tau|_1 - |U_\tau|_1} \sqrt{m}^{|\tau|_2 - |U_\tau|_2} S(\tau) \leq \prod_{j \in V_2(\tau) \setminus U_\tau \setminus V_\tau} (deg^\tau(j) - 1)!! \cdot \frac{1}{d^{A\varepsilon \sum_{e \in E(\tau)} l_e}}$$

Proof. Let $r_1 = |\tau|_1 - |U_\tau|_1, r_2 = |\tau|_2 - |U_\tau|_2$. Since $\Delta \leq 1$, it suffices to prove

$$E := \sqrt{d}^{r_1} \sqrt{m}^{r_2} \left(\frac{k}{d}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{\sum_{e \in E(\tau)} l_e} \leq \frac{1}{d^{A\varepsilon \sum_{e \in E(\tau)} l_e}}$$

We will need the following claim.

Claim 7.3.8. $\sum_{e \in E(\tau)} l_e \geq 2 \max(r_1, r_2)$.

Proof. We will first prove $\sum_{e \in E(\tau)} l_e \geq 2r_1$. For any vertex $i \in V_1(\tau) \setminus U_\tau \setminus V_\tau$, $deg^\tau(i)$ is even and is not 0, hence, $deg^\tau(i) \geq 2$. Any vertex $i \in U_\tau \setminus V_\tau$ cannot have $deg^\tau(i) = 0$ otherwise $U_\tau \setminus \{i\}$ is a vertex separator of strictly smaller weight than U_τ, which is not possible, hence, $deg^\tau(i) \geq 1$. Similarly, for $i \in V_\tau \setminus U_\tau$, $deg^\tau(i) \geq 1$. Also, since H_τ is bipartite, we have $\sum_{i \in V_1(\tau)} deg^\tau(i) = \sum_{j \in V_2(\tau)} deg^\tau(j) = \sum_{e \in E(\tau)} l_e$. Consider

$$\begin{aligned}
\sum_{e \in E(\tau)} l_e &= \sum_{i \in V_1(\tau)} deg^\tau(i) \\
&\geq \sum_{i \in V_1(\tau) \setminus U_\tau \setminus V_\tau} deg^\tau(i) + \sum_{i \in (U_\tau)_1 \setminus V_\tau} deg^\tau(i) + \sum_{i \in (V_\tau)_1 \setminus U_\tau} deg^\tau(i) \\
&\geq 2|V_1(\tau) \setminus U_\tau \setminus V_\tau| + |(U_\tau)_1 \setminus V_\tau| + |(V_\tau)_1 \setminus U_\tau| \\
&= 2r_1
\end{aligned}$$

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 $m \leq \frac{d}{\lambda^2}, m \leq \frac{k^2}{\lambda^2}$ and also, $E \leq \sqrt{d}^{r_1} \sqrt{m}^{r_2} \left(\frac{k}{d}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2 \max(r_1, r_2)}$ where we used the fact that $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon} \leq 1$.

Claim 7.3.9. *For integers $r_1, r_2 \geq 0$, if $m \leq \frac{d}{\lambda^2}$ and $m \leq \frac{k^2}{\lambda^2}$, then,*

$$\sqrt{d}^{r_1} \sqrt{m}^{r_2} \left(\frac{k}{d}\right)^{r_1} \left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2 \max(r_1, r_2)} \leq 1$$

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

$$\begin{aligned}
\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1} &\leq \sqrt{d}^{r_1}\left(\frac{\sqrt{d}}{\lambda}\right)^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1} \\
&= \left(\frac{\lambda}{\sqrt{d}}\right)^{r_1-r_2} \\
&\leq \left(\frac{1}{\sqrt{m}}\right)^{r_1-r_2} \\
&\leq 1
\end{aligned}$$

And if $r_1 < r_2$, we have

$$\begin{aligned}
\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} &= \sqrt{d}^{r_1}\sqrt{m}^{r_2-r_1}\sqrt{m}^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} \\
&\leq \sqrt{d}^{r_1}\left(\frac{k}{\lambda}\right)^{r_2-r_1}\left(\frac{\sqrt{d}}{\lambda}\right)^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2} \\
&= 1
\end{aligned}$$

■

For the desired bounds, we mimic this argument while carefully keeping track of factors of d^{ε}.

Claim 7.3.10. *For integers $r_1, r_2 \geq 0$ and an integer $r \geq 2\max(r_1, r_2)$, if $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}$ and $m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, then,*

$$\sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r} \leq \left(\frac{1}{d^{A\varepsilon}}\right)^{r}$$

Proof. If $r_1 \geq r_2$,

$$\begin{aligned}
E &= \sqrt{d}^{r_1}\sqrt{m}^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_1} \\
&\leq \sqrt{d}^{r_1}\left(\frac{\sqrt{d}^{1-\varepsilon}}{\lambda}\right)^{r_2}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_1} \\
&= \left(\frac{\lambda}{\sqrt{d}^{1-\varepsilon}}\right)^{r_1-r_2}\left(\frac{1}{\sqrt{d}}\right)^{\varepsilon r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_1} \\
&\leq \left(\frac{1}{\sqrt{m}}\right)^{r_1-r_2}\left(\frac{1}{\sqrt{d}}\right)^{\varepsilon r_1}\left(\frac{1}{d^{A\varepsilon}}\right)^{r-2r_1} \\
&\leq \left(\frac{1}{d^{2A}}\right)^{\varepsilon r_1}\left(\frac{1}{d^{A\varepsilon}}\right)^{r-2r_1} \\
&= \left(\frac{1}{d^{A\varepsilon}}\right)^{r}
\end{aligned}$$

And if $r_1 < r_2$,

$$\begin{aligned}
E &= \sqrt{d}^{r_1}\sqrt{m}^{r_2-r_1}\sqrt{m}^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&\leq \sqrt{d}^{r_1}\left(\frac{\sqrt{k}^{2-\varepsilon}}{\lambda}\right)^{r_2-r_1}\left(\frac{\sqrt{d}^{1-\varepsilon}}{\lambda}\right)^{r_1}\left(\frac{k}{d}\right)^{r_1}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{2r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&= \left(\frac{\sqrt{k}}{\sqrt{d}}\right)^{\varepsilon r_1}\left(\frac{1}{\sqrt{k}}\right)^{\varepsilon r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&\leq \left(\frac{1}{\sqrt{k}}\right)^{\varepsilon r_2}\left(\frac{\sqrt{\lambda}}{\sqrt{k}}\right)^{r-2r_2} \\
&\leq \left(\frac{1}{d^{2A}}\right)^{\varepsilon r_2}\left(\frac{1}{d^{A\varepsilon}}\right)^{r-2r_2} \\
&\leq \left(\frac{1}{d^{A\varepsilon}}\right)^{\sum_{e\in E(\tau)} l_e}
\end{aligned}$$

■

The result follows by setting $r = \sum_{e\in E(\tau)} l_e$ in the above claim. ■

Corollary 7.3.11. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$, we have*

$$c(\tau)B_{norm}(\tau)S(\tau)R(\tau) \leq 1$$

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 τ is a proper middle shape, we have $w(I_\tau) = 0$ and $w(S_{\tau,min}) = w(U_\tau)$. This implies

$$n^{\frac{w(V(\tau))+w(I_\tau)-w(S_{\tau,min})}{2}} = \sqrt{d}^{|\tau|_1-|U_\tau|_1}\sqrt{m}^{|\tau|_2-|U_\tau|_2}$$

As was observed in the proof of Lemma 7.3.7, 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, $q = d^{O(1)\cdot\varepsilon(C_V+C_E)}$. We can set C_V, C_E sufficiently small so that

$$\begin{aligned}
c(\tau)B_{norm}(\tau)S(\tau)R(\tau) &= 100(6D_V)^{|U_\tau\setminus V_\tau|+|V_\tau\setminus U_\tau|+2|E(\tau)|}4^{|V(\tau)\setminus(U_\tau\cup V_\tau)|} \\
&\quad \cdot 2e(6qD_V)^{|V(\tau)\setminus U_\tau|+|V(\tau)\setminus V_\tau|}\prod_{e\in E(\tau)}(400D_V^2D_E^2q)^{l_e} \\
&\quad \cdot \sqrt{d}^{|\tau|_1-|U_\tau|_1}\sqrt{m}^{|\tau|_2-|U_\tau|_2}S(\tau)(C_{disc}\sqrt{D_E})^{\sum_{j\in(U_\tau)_2\cup(V_\tau)_2} deg^\tau(j)} \\
&\leq d^{O(1)\cdot(C_V+C_E)\cdot\varepsilon\sum_{e\in E(\tau)} l_e} \cdot \prod_{j\in V_2(\tau)\setminus V_2(U_\tau)\setminus V_2(V_\tau)} (deg^\tau(j)-1)!! \cdot \frac{1}{d^{A\varepsilon\sum_{e\in E(\tau)} l_e}} \\
&\leq d^{O(1)\cdot(C_V+C_E)\cdot\varepsilon\sum_{e\in E(\tau)} l_e} \cdot (D_VD_E)^{\sum_{e\in E(\tau)} l_e} \cdot \frac{1}{d^{A\varepsilon\sum_{e\in E(\tau)} l_e}} \\
&\leq 1
\end{aligned}$$

■

We can now prove Lemma 7.3.3.

Lemma 7.3.3. *For all $U \in \mathcal{I}_{mid}$ and $\tau \in \mathcal{M}_U$,*

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)}H_{Id_U} & B_{norm}(\tau)H_\tau \\ B_{norm}(\tau)H_\tau^T & \frac{1}{|Aut(U)|c(\tau)}H_{Id_U} \end{bmatrix} \succeq 0$$

Proof. We have

$$\begin{bmatrix} \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} & B_{norm}(\tau) H_\tau \\ B_{norm}(\tau) H_\tau^T & \frac{1}{|Aut(U)|c(\tau)} H_{Id_U} \end{bmatrix}$$
$$= \begin{bmatrix} \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)R(\tau)B_{norm}(\tau)}{|Aut(U)|} \right) H_{Id_U} & 0 \\ 0 & \left(\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)R(\tau)B_{norm}(\tau)}{|Aut(U)|} \right) H_{Id_U} \end{bmatrix}$$
$$+ B_{norm}(\tau) \begin{bmatrix} \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix}$$

By Lemma 6.4.15, $\begin{bmatrix} \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} & H_\tau \\ H_\tau^T & \frac{S(\tau)R(\tau)}{|Aut(U)|} H_{Id_U} \end{bmatrix} \succeq 0$, so the second term above is positive semidefinite. For the first term, by Lemma 6.4.11, $H_{Id_U} \succeq 0$ and by Corollary 7.3.11, $\frac{1}{|Aut(U)|c(\tau)} - \frac{S(\tau)R(\tau)B_{norm}(\tau)}{|Aut(U)|} \geq 0$, which proves that the first term is also positive semidefinite. ∎

7.3.2 Proof of Lemma 7.3.4

Lemma 7.3.12. *Suppose $0 < A < \frac{1}{4}$ is a constant such that $\frac{\sqrt{\lambda}}{\sqrt{k}} \leq d^{-A\varepsilon}, \frac{1}{\sqrt{k}} \leq d^{-2A}$ and $\frac{k}{d} \leq d^{-A\varepsilon}$. For all m such that $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}, m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$, for all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and for all $\gamma \in \Gamma_{U,V}$,*

$$n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \leq \left(\prod_{j \in V_2(\gamma)\setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!! \right)^2 \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e \in E(\gamma)} l_e)}}$$

for some constant $B > 0$ that depends only on C_Δ. 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 $deg^\gamma(i)$ is odd, then $S(\gamma) = 0$ and the inequality is true. So, assume $deg^\gamma(i)$ is even for all vertices $i \in V(\gamma) \setminus U_\gamma \setminus V_\gamma$. We

have $n^{w(V(\gamma)\setminus U_\gamma)} = d^{|\gamma|_1-|U_\gamma|_1}m^{|\gamma|_2-|U_\gamma|_2}$. Plugging in $S(\gamma)$, we get that we have to prove

$$E := d^{|\gamma|_1-|U_\gamma|_1}m^{|\gamma|_2-|U_\gamma|_2}\left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1}\Delta^{2|\gamma|_2-|U_\gamma|_2-|V_\gamma|_2}\prod_{e\in E(\gamma)}\frac{\lambda^{l_e}}{k^{l_e}} \leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)}l_e)}}$$

Let S' be the set of vertices $i \in U_\gamma \setminus V_\gamma$ that have $deg^\gamma(i) \geq 1$. Let e, f be the number of type 1 vertices and the number of type 2 vertices in S' respectively. Observe that $S' \cup (U_\gamma \cap V_\gamma)$ is a vertex separator of γ.

Let $g = |V_\gamma \setminus U_\gamma|_1$ (resp. $h = |V_\gamma \setminus U_\gamma|_2$) 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 $S' \cup (U_\gamma \cap V_\gamma)$ has weight $\sqrt{d}^{e+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{f+|U_\gamma\cap V_\gamma|_2}$. On the other hand, V_γ has weight $\sqrt{d}^{g+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{h+|U_\gamma\cap V_\gamma|_2}$. Since γ is a left shape, V_γ is the unique minimum vertex separator and hence, $\sqrt{d}^{e+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{f+|U_\gamma\cap V_\gamma|_2} \geq \sqrt{d}^{g+|U_\gamma\cap V_\gamma|_1}\sqrt{m}^{h+|U_\gamma\cap V_\gamma|_2}$ which implies $d^e m^f \geq d^g m^h$.

Let $p = |V(\gamma) \setminus (U_\gamma \cup V_\gamma)|_1$ (resp. $q = |V(\gamma) \setminus (U_\gamma \cup V_\gamma)|_2$) be the number of type 1 vertices (resp. type 2 vertices) in $V(\gamma) \setminus (U_\gamma \cup V_\gamma)$.

To illustrate the main idea, we will first prove the weaker inequality $E \leq 1$. Since $\Delta \leq 1$, it suffices to prove

$$d^{|\gamma|_1-|U_\gamma|_1}m^{|\gamma|_2-|U_\gamma|_2}\left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1}\prod_{e\in E(\gamma)}\frac{\lambda^{l_e}}{k^{l_e}} \leq 1$$

We have

$$d^{|\gamma|_1-|U_\gamma|_1}m^{|\gamma|_2-|U_\gamma|_2} = d^{p+g}m^{q+h} \leq n^{p+\frac{e+g}{2}}m^{q+\frac{f+h}{2}}$$

since $d^e m^f \geq d^g m^h$. Also, $2|\gamma|_1 - |U_\gamma|_1 - |V_\gamma|_1 = 2p + e + g$. So, it suffices to prove

$$n^{p+\frac{e+g}{2}}m^{q+\frac{f+h}{2}}\left(\frac{k}{d}\right)^{2p+e+g}\prod_{e\in E(\gamma)}\left(\frac{\lambda}{k}\right)^{l_e} \leq 1$$

We will need the following claim.

Claim 7.3.13. $\sum_{e\in E(\gamma)} l_e \geq \max(2p+e+g, 2q+f+h)$

Proof. Since H_γ 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,

$$\begin{aligned}\sum_{e\in E(\gamma)} l_e &= \sum_{i\in V_1(\gamma)} deg^\gamma(i) \\ &\geq \sum_{i\in V_1(\gamma)\setminus U_\gamma\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in (U_\gamma)_1\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in (V_\gamma)_1\setminus U_\gamma} deg^\gamma(i) \\ &\geq 2p+e+g\end{aligned}$$

Similarly,

$$\begin{aligned}\sum_{e\in E(\gamma)} l_e &= \sum_{i\in V_2(\gamma)} deg^\gamma(i) \\ &\geq \sum_{i\in V_2(\gamma)\setminus U_\gamma\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in (U_\gamma)_2\setminus V_\gamma} deg^\gamma(i) + \sum_{i\in (V_\gamma)_2\setminus U_\gamma} deg^\gamma(i) \\ &\geq 2q+f+h\end{aligned}$$

Therefore, $\sum_{e\in E(\gamma)} l_e \geq \max(2p+e+g, 2q+f+h)$. ∎

Now, let $r_1 = p + \frac{e+g}{2}, r_2 = q + \frac{f+h}{2}$. Then, $\sum_{e\in E(\gamma)} l_e \geq 2\max(r_1, r_2)$ and we wish to prove

$$d^{r_1} m^{r_2} \left(\frac{k}{d}\right)^{2r_1} \left(\frac{\lambda}{k}\right)^{2\max(r_1,r_2)} \leq 1$$

This expression simply follows by squaring Claim 7.3.9.

Now, to prove that $E \leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)} l_e)}}$, we mimic this argument while carefully keeping track of factors of d^ε. Again, using $d^e m^f \geq d^g m^h$, it suffices to prove that

$$d^{p+\frac{e+g}{2}} m^{q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1} \Delta^{2|\gamma|_2-|U_\gamma|_2-|V_\gamma|_2} \prod_{e\in E(\gamma)} \frac{\lambda^{l_e}}{k^{l_e}} \leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|+\sum_{e\in E(\gamma)} l_e)}}$$

The idea is that the $d^{B\varepsilon}$ decay for the edges are obtained from the stronger assumption on m, namely $m \leq \frac{d^{1-\varepsilon}}{\lambda^2}, m \leq \frac{k^{2-\varepsilon}}{\lambda^2}$. And the $d^{B\varepsilon}$ 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 $\frac{k}{d}$, the latter especially useful for the degree 0 vertices. Finally, the $d^{B\varepsilon}$ decay for the type 2 vertices of $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$ are obtained from the factors of Δ.

Indeed, note that for a constant B that depends on C_Δ, $\Delta^{2|\gamma|_2-|U_\gamma|_2-|V_\gamma|_2} \leq d^{-B\varepsilon|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|_2}$. So, we would be done if we prove

$$d^{p+\frac{e+g}{2}} m^{q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1} \left(\frac{\lambda}{k}\right)^{\sum_{e\in E(\gamma)} l_e} \leq \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma\cap V_\gamma)|_1+\sum_{e\in E(\gamma)} l_e)}}$$

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, $\left(\frac{k}{d}\right)^{2|\gamma|_1-|U_\gamma|_1-|V_\gamma|_1} = \left(\frac{k}{d}\right)^{2p+e+g+c_0}$. For these degree 0 vertices, we have that the factors of $\frac{k}{d} \leq d^{-A\varepsilon}$ offer a decay of $\frac{1}{d^{B\varepsilon}}$. Therefore, it suffices to prove

$$d^{p+\frac{e+g}{2}} m^{q+\frac{f+h}{2}} \left(\frac{k}{d}\right)^{2p+e+g} \left(\frac{\lambda}{k}\right)^{\sum_{e\in E(\gamma)} l_e} \leq \frac{1}{d^{B\varepsilon(p+q+e+f+g+h)+\sum_{e\in E(\gamma)} l_e)}}$$

for a constant $B > 0$. Observe that $p+q+e+f+g+h \leq 2(\sum_{e\in E(\gamma)} l_e)$. Therefore, using the notation $r_1 = p + \frac{e+g}{2}, r_2 = q + \frac{f+h}{2}$, it suffices to prove

$$d^{r_1} m^{r_2} \left(\frac{k}{d}\right)^{2r_1} \left(\frac{\lambda}{k}\right)^{\sum_{e\in E(\gamma)} l_e} \leq \frac{1}{d^{B\varepsilon \sum_{e\in E(\gamma)} l_e}}$$

for a constant $B > 0$. But this follows by squaring Claim 7.3.10 where we set $r = \sum_{e \in E(\gamma)} l_e$.

■

Remark 7.3.14. *In the above bounds, note that there is a decay of $d^{B\varepsilon}$ for each vertex in $V(\gamma) \setminus (U_\gamma \cap V_\gamma)$. One of the main technical reasons for introducing the slack parameter C_Δ in the planted distribution was to introduce this decay, which is needed in the current machinery.*

We can now prove Lemma 7.3.4.

Lemma 7.3.4. *For all $U, V \in \mathcal{I}_{mid}$ where $w(U) > w(V)$ and all $\gamma \in \Gamma_{U,V}$,*

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H^{-\gamma,\gamma}_{Id_V} \preceq H'_\gamma$$

Proof. By Lemma 6.4.17, we have

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 H^{-\gamma,\gamma}_{Id_V} \preceq c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} H'_\gamma$$

Using the same proof as in Lemma 6.4.11, we can see that $H'_\gamma \succeq 0$. Therefore, it suffices to prove that

$$c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \leq 1$$

Since $U, V \in \mathcal{I}_{mid}$, $Aut(U) = |U|_1!|U|_2!, Aut(V) = |V|_1!|V|_2!$. Therefore, $\frac{|Aut(U)|}{|Aut(V)|} = \frac{|U|_1!|U|_2!}{|V|_1!|V|_2!} \leq D_V^{|U_\gamma \setminus V_\gamma|}$. Also, $|E(\gamma)| \leq \sum_{e \in E(\gamma)} l_e$ and $q = d^{O(1)\cdot\varepsilon(C_V + C_E)}$. Note that

$$R(\gamma)^2 = (C_{disc}\sqrt{D_E})^{2\sum_{j \in (U_\gamma)_2 \cup (V_\gamma)_2} deg^\gamma(j)} \leq d^{O(1)\cdot\varepsilon C_E \cdot \sum_{e \in E(\gamma)} l_e}$$

and

$$\left(\prod_{j \in V_2(\gamma) \setminus U_\gamma \setminus V_\gamma} (deg^\gamma(j) - 1)!!\right)^2 \leq (D_V D_E)^{2\sum_{e \in E(\tau)} l_e} \leq d^{O(1)\cdot\varepsilon(C_V + C_E)\cdot\sum_{e \in E(\gamma)} l_e}$$

Let B be the constant from Lemma 7.3.12. We can set C_V, C_E sufficiently small so that, using Lemma 7.3.12,

$$\begin{aligned}
&c(\gamma)^2 N(\gamma)^2 B(\gamma)^2 S(\gamma)^2 R(\gamma)^2 \frac{|Aut(U)|}{|Aut(V)|} \\
&\qquad \leq 100^2 (6D_V)^{2|U_\gamma \setminus V_\gamma| + 2|V_\gamma \setminus U_\gamma| + |E(\alpha)|} 16^{|V(\gamma)\setminus(U_\gamma \cup V_\gamma)|} \\
&\qquad\quad \cdot (3D_V)^{4|V(\gamma)\setminus V_\gamma| + 2|V(\gamma)\setminus U_\gamma|} (6qD_V)^{2|V(\gamma)\setminus U_\gamma| + 2|V(\gamma)\setminus V_\gamma|} \prod_{e \in E(\gamma)} (400 D_V^2 D_E^2 q)^{2l_e} \\
&\qquad\quad \cdot n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 d^{O(1)\cdot \varepsilon C_E \cdot \sum_{e\in E(\gamma)} l_e} \cdot D_V^{|U_\gamma \setminus V_\gamma|} \\
&\qquad \leq d^{O(1)\cdot\varepsilon(C_V + C_E)\cdot(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)} \cdot n^{w(V(\gamma)\setminus U_\gamma)} S(\gamma)^2 \\
&\qquad \leq d^{O(1)\cdot\varepsilon(C_V + C_E)\cdot(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)} \cdot \frac{1}{d^{B\varepsilon(|V(\gamma)\setminus(U_\gamma \cap V_\gamma)| + \sum_{e\in E(\gamma)} l_e)}} \\
&\qquad \leq 1
\end{aligned}$$

■

7.3.3 Proof of Lemma 7.3.5

In this section, we will prove Lemma 7.3.5 using the strategy sketched in [149, Section 10].

Lemma 7.3.5. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)| c(\gamma)} \right) Id_{sym}$$

In particular, we prove the following lemmas.

Lemma 7.3.15. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{d^{K_1 D_{sos}^2}} Id_{sym}$$

for a constant $K_1 > 0$ that can depend on C_Δ.

Lemma 7.3.16.

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{d^{K_2 D_{sos}}}{2^{D_V}}$$

for a constant $K_2 > 0$ that can depend on C_Δ.

If we assume the above lemmas, we can prove Lemma 7.3.5.

Lemma 7.3.5. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq 6 \left(\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)} \right) Id_{sym}$$

Proof. Let $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$. By Lemma 7.3.15,

$$\sum_{U \in \mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{d^{K_1 D^2_{sos}}} Id_{sym}$$

for a constant $K_1 > 0$. By Lemma 7.3.16,

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{d^{K_2 D_{sos}}}{2^{D_V}}$$

for a constant $K_2 > 0$.

We choose C_{sos} sufficiently small so that $\frac{1}{d^{K_1 D^2_{sos}}} \geq 6 \frac{d^{K_2 D_{sos}}}{2^{D_V}}$ which can be satisfied by setting $C_{sos} < K_3 C_V$ for a sufficiently small constant $K_3 > 0$. Then, since $Id_{Sym} \succeq 0$, using

Lemma 7.3.15 and Lemma 7.3.16,

$$\begin{aligned}\sum_{U\in\mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) &\succeq \frac{1}{d^{K_1 D_{sos}^2}} Id_{sym}\\ &\succeq 6\frac{d^{K_2 D_{sos}}}{2^{D_V}} Id_{sym}\\ &\succeq 6\left(\sum_{U\in\mathcal{I}_{mid}}\sum_{\gamma\in\Gamma_{U,*}} \frac{d_{Id_U}(H'_\gamma, H_{Id_U})}{|Aut(U)|c(\gamma)}\right) Id_{sym}\end{aligned}$$

■

In the rest of the section, we will prove Lemma 7.3.15 and Lemma 7.3.16.

To begin with, we will need a bound on $B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma')$.

Lemma 7.3.17. *Suppose $0 < A < \frac{1}{4}$ is a constant such that $\frac{\sqrt{\lambda}}{\sqrt{k}} \le d^{-A\varepsilon}$ and $\frac{1}{\sqrt{k}} \le d^{-2A}$. Suppose m is such that $m \le \frac{d^{1-\varepsilon}}{\lambda^2}, m \le \frac{k^{2-\varepsilon}}{\lambda^2}$. For all $U \in \mathcal{I}_{mid}$ and $\sigma, \sigma' \in \mathcal{L}_U$,*

$$B_{norm}(\sigma)B_{norm}(\sigma')H_{Id_U}(\sigma,\sigma') \le \frac{d^{O(1)D_{sos}}}{d^{0.5A\varepsilon|V(\sigma\circ\sigma')|}}$$

Proof. Suppose there is a vertex $i \in V(\sigma)\setminus V_\sigma$ such that $deg^\sigma(i) + deg^{U_\sigma}(i)$ is odd, then $H_{Id_U}(\sigma,\sigma') = 0$ and the inequality is true. So, assume that $deg^\sigma(i) + deg^{U_\sigma}(i)$ is even for all $i \in V(\sigma)\setminus V_\sigma$. Similarly, assume that $deg^{\sigma'}(i) + deg^{U_{\sigma'}}(i)$ 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,

$$\begin{aligned}n^{\frac{w(V(\sigma))-w(U)}{2}} &= \sqrt{d}^{|V(\sigma)|_1-|U|_1}\sqrt{m}^{|V(\sigma)|_2-|U|_2}\\ &= \sqrt{d}^{|U_\sigma|_1}\sqrt{m}^{|U_\sigma|_2}\sqrt{d}^{e}\sqrt{m}^{f}\\ &\le d^{O(1)D_{sos}}\sqrt{d}^{e}\sqrt{m}^{f}\end{aligned}$$

where we used the fact that $|U_\sigma| \le D_{sos}$.

Let there be g (resp. h) vertices of type 1 (resp. type 2) in $V(\sigma') \setminus U_{\sigma'} \setminus V_{\sigma'}$. Then, similarly, $n^{\frac{w(V(\sigma'))-w(U)}{2}} \leq d^{O(1)D_{sos}} \sqrt{d}^g \sqrt{m}^h$.

Let $\alpha = \sigma \circ \sigma'$. Since all vertices in $V(\alpha) \setminus U_\alpha \setminus V_\alpha$ have degree at least 2, we have $\sum_{e \in E(\alpha)} l_e \geq \sum_{i \in V_1(\alpha) \setminus U_\alpha \setminus V_\alpha} deg^\alpha(i) \geq 2(e+g)$. 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 7.3.10, we have

$$\sqrt{d}^{e+g} \sqrt{m}^{f+h} \left(\frac{k}{d}\right)^{e+g} \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}} \leq \frac{1}{d^{A\varepsilon \sum_{e \in E(\alpha)} l_e}}$$

Also, $\left(\frac{k}{d}\right)^{|\alpha|_1} \leq \left(\frac{k}{d}\right)^{e+g}$ and $\prod_{j \in V_2(\alpha)} (deg^\alpha(j)-1)!! \leq d^{\varepsilon C_V \sum_{e \in E(\alpha)} l_e}$. Therefore,

$$\begin{aligned}
&n^{\frac{w(V(\sigma))-w(U)}{2}} n^{\frac{w(V(\sigma'))-w(U)}{2}} H_{Id_U}(\sigma, \sigma') \\
&\quad \leq d^{O(1)D_{sos}} \sqrt{d}^e \sqrt{m}^f d^{O(1)D_{sos}} \sqrt{d}^g \sqrt{m}^h \\
&\qquad \cdot \frac{1}{|Aut(U)|} \left(\frac{1}{\sqrt{k}}\right)^{deg(\alpha)} \left(\frac{k}{d}\right)^{|\alpha|_1} \Delta^{|\alpha|_2} \prod_{j \in V_2(\alpha)} (deg^\alpha(j)-1)!! \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}} \\
&\quad \leq d^{O(1)D_{sos}} d^{\varepsilon C_V \sum_{e \in E(\alpha)} l_e} \sqrt{d}^{e+g} \sqrt{m}^{f+h} \left(\frac{k}{d}\right)^{e+g} \prod_{e \in E(\alpha)} \frac{\sqrt{\lambda}^{l_e}}{\sqrt{k}^{l_e}} \\
&\quad \leq d^{O(1)D_{sos}} d^{\varepsilon C_V \sum_{e \in E(\alpha)} l_e} \frac{1}{d^{A\varepsilon \sum_{e \in E(\alpha)} l_e}}
\end{aligned}$$

Now, observe that since all vertices in $V(\alpha) \setminus U_\alpha \setminus V_\alpha$ have degree at least 1, $|V(\alpha)| \leq 2D_{sos} + 2\sum_{e \in E(\alpha)} l_e$. So, by setting C_V, C_E sufficiently small,

$$\begin{aligned}
B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') &= 2e(6qD_V)^{|V(\sigma) \setminus U_\sigma| + |V(\sigma) \setminus V_\sigma|} \prod_{e \in E(\sigma)} (400 D_V^2 D_E^2 q)^{l_e} n^{\frac{w(V(\sigma))-w(U)}{2}} \\
&\quad \cdot 2e(6qD_V)^{|V(\sigma') \setminus U_{\sigma'}| + |V(\sigma') \setminus V_{\sigma'}|} \prod_{e \in E(\sigma')} (400 D_V^2 D_E^2 q)^{l_e} n^{\frac{w(V(\sigma'))-w(U)}{2}} \\
&\quad \cdot H_{Id_U}(\sigma, \sigma') \\
&\leq d^{O(1) \cdot \varepsilon(C_V + C_E) \cdot (|V(\alpha)| + \sum_{e \in E(\alpha)} l_e)} d^{O(1)D_{sos}} d^{\varepsilon C_V \sum_{e \in E(\alpha)} l_e} \frac{1}{d^{A\varepsilon \sum_{e \in E(\alpha)} l_e}} \\
&\leq \frac{d^{O(1)D_{sos}}}{d^{0.5A\varepsilon |V(\alpha)|}}
\end{aligned}$$

■

Proof of Lemma 7.3.15

To prove Lemma 7.3.15, we will use the strategy from [149, Section 10]. We will also use the notation from that section. We recall that for $U \in \mathcal{I}_{mid}$, $\mathcal{L}'_U \subset \mathcal{L}_U$ was the set of non-trivial shapes in $\mathcal{L}_U$.

Proposition 7.3.18. *For $V \in \mathcal{I}_{mid}$,*

$$\lambda_V = \frac{\Delta^{|V|_2}}{d^{|V|_1} k^{|V|_2}}$$

Proof. We have $\lambda_V = |Aut(V)| H_{Id_V}(Id_V, Id_V) = \left(\frac{1}{\sqrt{k}}\right)^{2|V|} \left(\frac{k}{d}\right)^{|V|_1} \Delta^{|V|_2} = \frac{\Delta^{|V|_2}}{d^{|V|_1} k^{|V|_2}}$. ■

Corollary 7.3.19. $\lambda_V \geq \frac{1}{d^{O(1)D_{sos}}}$

Lemma 7.3.20. *For any edge $e = (V, U)$ in G, we have*

$$w(e) \leq \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma \circ \sigma')|}}$$

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 7.3.17, we have

$$\begin{aligned}
2W(U,V) &= \frac{2}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_V, U_\sigma = U} \sum_{\sigma' \in \mathcal{L}_V, U_{\sigma'} \neq V} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \\
&\leq \frac{2}{|Aut(U)|} \sum_{\sigma \in \mathcal{L}_V, U_\sigma = U} \sum_{\sigma' \in \mathcal{L}_V, U_{\sigma'} \neq V} \frac{d^{O(1)D_{sos}}}{d^{0.5A\varepsilon|V(\sigma \circ \sigma')|}} \\
&\leq \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{2d^{O(1)D_{sos}}}{d^{0.5A\varepsilon|V(\sigma \circ \sigma')|}} \\
&\leq \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma \circ \sigma')|} D_{sos}^{D_{sos}} d^{F\varepsilon|V(\sigma \circ \sigma')|}} \\
&\leq \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma \circ \sigma')|}} \sum_{\sigma, \sigma' \in \mathcal{L}'_V} \frac{1}{D_{sos}^{D_{sos}} d^{F\varepsilon|V(\sigma \circ \sigma')|}} \\
&\leq \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma \circ \sigma')|}} \\
&\leq \frac{\lambda_V d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma \circ \sigma')|}}
\end{aligned}$$

where we set C_V, C_E small enough. Rearranging proves the lemma. ∎

Corollary 7.3.21. *For any $U, V \in \mathcal{I}_{mid}$ such that $w(U) > w(V)$,*

$$\sum_{P: P \text{ is a path from } V \text{ to } U \text{ in } G} \prod_{e \in E(P)} w(e) \leq d^{O(1)D_{sos}^2}$$

Proof. The total number of vertices in G is at most $(D_{sos}+1)^2$ since each $U \in \mathcal{I}_{mid}$ has at most 2 index shape pieces corresponding to each type and each index shape piece 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 $(D_{sos}+1)^{2j}$. 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

$$w(e) \leq \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|U'|}} \leq \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon}}$$

So, $\prod_{e\in E(P)} w(e) \leq \left(\frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon}}\right)^j$. Therefore, by setting C_{sos} small enough,

$$\sum_{P:P \text{ is a path from V to U in G}} \prod_{e\in E(P)} w(e) \leq \sum_{j=1}^{(D_{sos}+1)^2} (D_{sos}+1)^{2j} \left(\frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon}}\right)^j$$
$$\leq d^{O(1)D_{sos}^2}$$

■

We can now prove Lemma 7.3.15.

Lemma 7.3.15. *Whenever $\|M_\alpha\| \leq B_{norm}(\alpha)$ for all $\alpha \in \mathcal{M}'$,*

$$\sum_{U\in\mathcal{I}_{mid}} M^{fact}_{Id_U}(H_{Id_U}) \succeq \frac{1}{d^{K_1 D_{sos}^2}} Id_{sym}$$

for a constant $K_1 > 0$ that can depend on C_Δ.

Proof. For all $V \in \mathcal{I}_{mid}$, we have

$$Id_{Sym,V} \preceq 2 \sum_{U\in\mathcal{I}_{mid}:w(U)\geq w(V)} \left(\sum_{P:P \text{ is a path from V to U in } G} \prod_{e\in E(P)} w(e) \right) \frac{1}{\lambda_U} M^{fact}(H_{Id_U})$$

Summing this over all $V \in \mathcal{I}_{mid}$, we get

$$Id_{Sym} \preceq \sum_{U\in\mathcal{I}_{mid}} \frac{2}{\lambda_U} \left(\sum_{V\in\mathcal{I}_{mid}:w(U)\geq w(V)} \sum_{P:P \text{ is a path from V to U in G}} \prod_{e\in E(P)} w(e) \right) M^{fact}(H_{Id_U})$$
$$\preceq \sum_{U\in\mathcal{I}_{mid}} \frac{2}{\lambda_U} \left(\sum_{V\in\mathcal{I}_{mid}:w(U)\geq w(V)} d^{O(1)D_{sos}^2} \right) M^{fact}(H_{Id_U})$$

For any fixed $U \in \mathcal{I}_{mid}$, the number of $V \in \mathcal{I}_{mid}$ such that $w(U) \geq w(V)$ is at most

$(D_{sos}+1)^2$. Also, $\lambda_U \geq \frac{1}{d^{O(1)D_{sos}}}$ for all $U \in \mathcal{I}_{mid}$. Therefore,

$$\begin{aligned} Id_{Sym} &\preceq \sum_{U \in \mathcal{I}_{mid}} \frac{2}{\lambda_U}(D_{sos}+1)^2 d^{O(1)D_{sos}^2} M^{fact}(H_{Id_U}) \\ &\preceq \sum_{U \in \mathcal{I}_{mid}} d^{O(1)D_{sos}^2} M^{fact}(H_{Id_U}) \end{aligned}$$

where we used the fact that for all $U \in \mathcal{I}_{mid}$, $M^{fact}(H_{Id_U}) \succeq 0$. ∎

Proof of Lemma 7.3.16

We restate the lemma for convenience.

Lemma 7.3.16.

$$\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \leq \frac{d^{K_2 D_{sos}}}{2^{D_V}}$$

for a constant $K_2 > 0$ that can depend on C_Δ.

Proof. We have

$$\begin{aligned} &\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\ &\quad = \sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{1}{|Aut(U)|c(\gamma)} \sum_{\substack{\sigma,\sigma' \in \mathcal{L}_{U_\gamma} : |V(\sigma)| \leq D_V, |V(\sigma')| \leq D_V, \\ |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V}} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_{U_\gamma}}(\sigma, \sigma') \end{aligned}$$

The set of σ, σ' that could appear in the above sum must necessarily be non-trivial and hence, $\sigma, \sigma' \in \mathcal{L}'_U$. Then,

$$\begin{aligned} &\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} \\ &\quad = \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma,\sigma' \in \mathcal{L}'_U} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \sum_{\gamma \in \Gamma_{U,*} : |V(\sigma \circ \gamma)| > D_V \text{ or } |V(\sigma' \circ \gamma)| > D_V} \frac{1}{|Aut(U)|c(\gamma)} \end{aligned}$$

For $\sigma \in \mathcal{L}'_U$, define $m_\sigma = D_V + 1 - |V(\sigma)| \geq 1$. This is precisely set so that for all $\gamma \in \Gamma_{U,*}$, we have $|V(\sigma \circ \gamma)| > D_V$ if and only if $|V(\gamma)| \geq |U| + m_\sigma$. So, for $\sigma, \sigma' \in \mathcal{L}'_U$,

$$\begin{aligned}\sum_{\gamma \in \Gamma_{U,*}: |V(\sigma\circ\gamma)|>D_V \text{ or } |V(\sigma'\circ\gamma)|>D_V} \frac{1}{|Aut(U)|c(\gamma)} &= \sum_{\gamma \in \Gamma_{U,*}: |V(\gamma)| \geq |U| + \min(m_\sigma, m_{\sigma'})} \frac{1}{|Aut(U)|c(\gamma)} \\ &\leq \frac{1}{2^{\min(m_\sigma, m_{\sigma'})-1}}\end{aligned}$$

Also, for $\sigma, \sigma' \in \mathcal{L}'_U$, we have $|V(\sigma \circ \sigma')| + min(m_\sigma, m_{\sigma'}) - 1 \geq D_V$. Therefore,

$$\begin{aligned}\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} B_{norm}(\sigma) B_{norm}(\sigma') H_{Id_U}(\sigma, \sigma') \frac{1}{2^{\min(m_\sigma, m_{\sigma'})-1}} \\ &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{d^{O(1)D_{sos}}}{d^{0.5A\varepsilon|V(\sigma\circ\sigma')|} 2^{\min(m_\sigma, m_{\sigma'})-1}}\end{aligned}$$

where we used Lemma 7.3.17. Using $d^{0.5A\varepsilon|V(\sigma\circ\sigma')|} \geq d^{0.1A\varepsilon|V(\sigma\circ\sigma')|} 2^{|V(\sigma\circ\sigma')|}$,

$$\begin{aligned}\sum_{U \in \mathcal{I}_{mid}} \sum_{\gamma \in \Gamma_{U,*}} \frac{d_{Id_U}(H_{Id_U}, H'_\gamma)}{|Aut(U)|c(\gamma)} &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma\circ\sigma')|} 2^{|V(\sigma\circ\sigma')|} 2^{\min(m_\sigma, m_{\sigma'})-1}} \\ &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{d^{O(1)D_{sos}}}{d^{0.1A\varepsilon|V(\sigma\circ\sigma')|} 2^{D_V}} \\ &\leq \sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{d^{O(1)D_{sos}}}{D_{sos}^{D_{sos}} d^{0.1A\varepsilon|V(\sigma\circ\sigma')|} 2^{D_V}}\end{aligned}$$

The final step will be to argue that $\sum_{U \in \mathcal{I}_{mid}} \sum_{\sigma, \sigma' \in \mathcal{L}'_U} \frac{1}{D_{sos}^{D_{sos}} d^{0.1A\varepsilon|V(\sigma\circ\sigma')|}} \leq 1$ which will complete the proof. But this will follow if we set C_V, C_E small enough. ∎

CHAPTER 8

FOLLOWUP AND FUTURE WORK

In this chapter, we go over some followup works that are not covered in this dissertation and also suggest directions for future work.

8.1 Nonlinear concentration for non-product distributions

Our techniques in Chapter 2 apply to a collection of random variables that are sampled independently of each other. A natural question is to ask if we can generalize to the case when they are not independent. For example, this is useful when instead of analyzing Erdős-Rényi random graphs, we wish to analyze uniform d-regular graphs. Such a generalization seems extremely likely because our proof techniques essentially requires a Markov Chain that mixes rapidly to the given distribution, and then we can recursively apply the Poincaré inequality. We leave this for future work.

8.2 Sum-of-Squares lower bounds

In this dissertation, we saw several SoS lower bounds and while they build on fundamental conceptual blocks such as the nonlinear concentration results we show, and simple heuristics like pseudocalibration, an important technical barrier in the current proofs is that the proofs are highly technical and have many moving parts. It's an important research question to understand if the proofs can be simplified. Apart from enabling a better understanding of the SoS hierarchy, this will also help us understand the computational barriers of several fundamental problems in computer science. Examples of such problems follow.

8.2.1 Sparse independent set

In a followup work [95], we prove SoS lower bounds for the important problem of maximum independent set on sparse Erdős-Rényi random graphs.

An important aspect of the SoS lower bounds until this work (including the ones in this dissertation) is that they apply for the so-called *dense setting*, which corresponds to cases when the input distribution can be specified by a collection of independent Rademacher or Gaussian variables. In the case of planted clique, this corresponds to the case when the input is a random graph distributed according to $G_{n,\frac{1}{2}}$ i.e. specified by a collection of $\binom{n}{2}$ independent Rademacher variables. Similarly, the tensor PCA and sparse PCA lower bounds we show in this dissertation apply when the input tensor has independent Rademacher or Gaussian entries. The techniques used to establish these lower bounds have proved difficult to extend to the case when the input distribution naturally corresponds to a *sparse graph* (or more generally, when it is specified by a collection of independent sub-gaussian variables, with Orlicz norm $\omega(1)$ instead of $O(1)$).

In [95], we extend lower bound technology for SoS to the *sparse setting*, where the input is a graph with average degree $d \leq n/2$. We use as a case study the fundamental combinatorial optimization problem of independent set. For the dense case $d = n/2$, finding an independent set is equivalent to finding a clique and the paper [10] shows an average-case lower bound against the Sum-of-Squares algorithm. We extend the techniques used in this dissertation, namely pseudocalibration, graph matrices, and the approximate decomposition into positive semidefinite matrices, in order to show the first average-case lower bound for the sparse setting. We hope that the techniques developed here offer a gateway for the analysis of SoS on other sparse problems.

Sample $G \sim G_{n,\frac{d}{n}}$ as an Erdős-Rényi random graph with average degree d, where we think of $d \ll n$. Specializing to the problem of independent set, a maximum independent set in G has size:

Fact 8.2.1 ([44, 46, 52])**.** *W.h.p. the max independent set in G has size $(1+o_d(1)) \cdot \frac{2\ln d}{d} \cdot n$.*

The value of the degree-2 SoS relaxation for independent set equals the Lovász ϑ function, which is an upper bound on the independence number $\alpha(G)$, by virtue of being a relaxation. For random graphs $G \sim G_{n,d/n}$ this value is larger by a factor of about $\sqrt{d}$ than the true value of $\alpha(G)$ with high probability.

Fact 8.2.2 ([43])**.** *W.h.p. $\vartheta(G) = \Theta(\frac{n}{\sqrt{d}})$.*

In [95], we prove that the value of higher-degree SoS is also on the order of $n/\sqrt{d}$, rather than n/d, and thereby demonstrate that the information-computation gap against basic SDP/spectral algorithms persists against higher-degree SoS.

In considering relaxations for independent set of a graph $G = (V, E)$, with variables x_v being the 0/1 indicators of the independent set, the SoS relaxation searches for pseudoexpectation operators satisfying the polynomial constraints

$$\forall v \in V. \quad x_v^2 = x_v \qquad \text{and} \qquad \forall (u,v) \in E. \quad x_u x_v = 0\,.$$

The objective value of the convex relaxation is given by the quantity $\widetilde{\mathbb{E}}[\sum_{v\in V} x_v] = \sum_{v\in V} \widetilde{\mathbb{E}}[x_v]$. For the results below, we say that an event occurs with high probability (w.h.p.) when it occurs with probability at least $1 - O(1/n^c)$ for some $c > 0$. The following theorem states our main result.

Theorem 8.2.3. *There is an absolute constant $c_0 \in \mathbb{N}$ such that for sufficiently large $n \in \mathbb{N}$ and $d \in [(\log n)^2, n^{0.5}]$, and parameters k, D_{SoS} satisfying*

$$k \;\leq\; \frac{n}{D_{SoS}^{c_0} \cdot \log n \cdot d^{1/2}}\,,$$

it holds w.h.p. for $G = (V, E) \;\sim\; G_{n,\,d/n}$ that there exists a degree-D_{SoS} pseudoexpectation

satisfying

$$\forall v \in V. \quad x_v^2 = x_v \qquad \text{and} \qquad \forall (u,v) \in E. \quad x_u x_v = 0\,,$$

and objective value $\widetilde{\mathbb{E}}[\sum_{v\in V} x_v] \;\geq\; (1-o(1))k$.

Remark 8.2.4. *This is a non-trivial lower bound whenever* $D_{SoS} \;\leq\; \left(\frac{d^{1/2}}{\log n}\right)^{1/c_0}$.

Remark 8.2.5. *It suffices to set* $c_0 = 20$ *for our current proof. We did not optimize the tradeoff in* D_{SoS} *with* k*, but we did optimize the log factor (with the hope of eventually removing it).*

Remark 8.2.6. *Using the same technique, we can prove an* $n^{\Omega(\varepsilon)}$ *SoS-degree lower bound for all* $d \in [\sqrt{n}, n^{1-\varepsilon}]$.

For $n^\varepsilon \leq d \leq n^{0.5}$, the theorem gives a polynomial n^δ SoS-degree lower bound. For smaller d, the bound is still strong against low-degree SoS, but it becomes trivial as D_{SoS} approaches $(d^{1/2}/\log n)^{1/c_0}$ or d approaches $(\log n)^2$ since k matches the size of the maximum independent set in G, hence there is an actual distribution over independent sets of this size (the expectation operator for which is trivially is also a pseudoexpectation operator).

The above bound says nothing about the "almost dense" regime $d \in [n^{1-\varepsilon}, n/2]$. To handle this regime, we observe that our techniques, along with the ideas from the $\Omega(\log n)$-degree SoS bound from [10] for the dense case, prove a lower bound for any degree $d \geq n^\varepsilon$.

Theorem 8.2.7. *For any* $\varepsilon_1, \varepsilon_2 > 0$ *there is* $\delta > 0$*, such that for* $d \in [n^{\varepsilon_1}, n/2]$ *and* $k \leq \frac{n}{d^{1/2+\varepsilon_2}}$*, it holds w.h.p. for* $G = (V,E) \;\sim\; G_{n,\, d/n}$ *that there exists a degree-*$(\delta \log d)$ *pseudoexpectation satisfying*

$$\forall v \in V. \quad x_v^2 = x_v \qquad \text{and} \qquad \forall (u,v) \in E. \quad x_u x_v = 0\,,$$

and objective value $\widetilde{\mathbb{E}}[\sum_{v\in V} x_v] \;\geq\; (1-o(1))k$.

In particular, these theorems rule out polynomial-time certification (i.e. constant degree SoS) for any $d \geq \text{polylog}(n)$.

One of the important conceptual innovations in this work was to improve our understanding of sparse graph matrices by obtaining better norm bounds, which was done in Chapter 2. We also employ several other techniques such as pseudocalibration with connected truncation, conditioning and generalized intersection tradeoff lemmas.

8.2.2 Planted Affine Planes and Maximum Cut

We conjecture that for the Planted Affine Planes problem, defined in Chapter 5, the problem remains difficult for SoS even with the number of vectors increased to $m = n^{2-\varepsilon}$.

Conjecture 8.2.8. *Theorem 4.1.4 holds with the bound on the number of sampled vectors m loosened to $m \leq n^{2-\varepsilon}$.*

The reason for the upper bound comes from Remark 5.4.9. As we saw in Chapter 3, analyzing $\widetilde{\mathbb{E}}[1]$ is an established way to hypothesize about the power of SoS. We will revisit this point in the next section.

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

Conjecture 8.2.9. *Theorem 4.1.5 holds with the bound on the dimension p of a random subspace loosened to $p \geq n^{1/2+\varepsilon}$.*

We remark that recent work [186] has exhibited a polynomial time for the search variant of Planted Affine Planes for $m \geq n+1$, as opposed to prior known algorithms that required $m \gg n^2$, some of which were SoS based. Their algorithm is lattice-based, uses the special algebraic structure present in the problem, and is not captured by SoS. It's not clear if their search algorithm can be used for certification.

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

Conjecture 8.2.10. *Let $d \geq 3$, and let G be a random d-regular graph on n vertices. For some $\delta > 0$, w.h.p. there is a degree-n^δ pseudoexpectation operator $\widetilde{\mathbb{E}}$ on boolean variables x_i with maximum cut value at least*

$$\frac{1}{2} + \frac{\sqrt{d-1}}{d}(1 - \mathrm{o}_{d,n}(1))$$

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

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

Evidence for this conjecture comes from the fact that the only property required of the random inputs $d_1, \ldots, d_m$ was that norm bounds hold for the graph matrix with Hermite polynomial entries. When the variables $\{d_{u,i}\}$ are i.i.d from some other distribution, if we use graph matrices for the orthonormal polynomials under the distribution and assuming suitable bounds on the moments of the distribution, the same norm bounds hold [2]. When $d_u \in_{\mathrm{R}} S^n$ or another distribution for which the coordinates are not i.i.d, it seems likely that

if we use e.g. the spherical harmonics then similar norm bounds hold, but this is not proven.

8.2.3 Unique Games

The famous Unique Games conjecture (UGC) [100] postulates that a graph theory problem known as the Unique Games problem is NP-hard. This conjecture gained tremendous traction in the community because of it's numerous consequences (e.g. [100, 102, 152])) and connections to various other fields such as metric geometry [104] and discrete Fourier analysis [103]. An exciting array of recent works [53, 16, 175] has shown that a problem closely related to unique games, known as 2-to-2 games, is NP-hard. This is an important step towards proving the UGC and offers evidence that the UGC is true.

On the algorithmic side, there have been various attempts (see for e.g. [179, 37, 5]) to disprove the UGC. In particular, Barak et al. [12] showed that degree 8 SoS can efficiently solve integrality gap instances of the Unique Games problem that were proposed for linear programs and SDPs considered earlier. This work caused significant interest in the community, since it suggests that SoS might be a way to refute the UGC.

Therefore, it's tremendously important to understand the performance of SoS on the unique games problem. A good first step would be to understand the performance of SoS for the problem of maximum cut, which is also a Unique Games problem. In fact, we can be even more concrete and ask for the performance of SoS for the problem of maximum cut on random graphs, more precisely Conjecture 8.2.10. Lower bounds were shown for degree 2 and degree 4 in [133, 129] and generalizing their analyses for higher degree SoS is a nontrivial but important open problem.

8.3 Low degree likelihood ratio hypothesis

As explained in Chapter 3, the low-degree likelihood ratio hypothesis analytically predicts the computational barriers for hypothesis testing in bounded time, for *sufficiently nice* dis-

tributions. See [91, 116, 79] and references therein for more details. A full proof of this hypothesis is beyond current techniques, since it's likely harder than proving say $P \neq NP$. But confirming the hypothesis in restricted proof systems is a fascinating and important field for future research. In particular, building on the notation from Chapter 3, we would like to prove that for sufficiently nice distributions ν, μ, after pseudo-calibrating, if $\widetilde{\mathbb{E}}[1] = 1 + o(1)$, then there exists an SoS lower bound. Indeed, in this work, we confirm this for several fundamental problems. But proving this in general will go a long way towards understanding the power of bounded-time algorithms.

8.4 Technical improvements

Having covered the general directions for future research, we now specify a few directions for improving some technical aspects of our results.

8.4.1 Improving parameter dependences

In many of our lower bounds, we require polynomial decay in the Fourier coefficients. For example, we require a decay of n^{ε} for each new Fourier character, where n is the input size. This is done to handle various other factors that appear in norm bounds when doing the charging arguments. In the proofs, we term these as vertex or edge decay, corresponding to how they are encoded in the graph matrix arguments we use. By doing this, we obtain a slightly weaker lower bound. For example, instead of getting a $n^{1/4}$ lower bound (upto polylogarithmic factors) for Tensor PCA, we obtain a $n^{1/4-\varepsilon}$ lower bound for any $\varepsilon > 0$. In general, while they facilitate the proof, it's not clear that this sort of decay is necessary and it's open to find a tighter analysis so as to close the gap from known upper bounds upto a polylogarithmic factor.

Related to the above discussion, another open problem is to push the degree of SoS higher in our lower bounds. For example, for the Sherrington-Kirkpatrick lower bound, it's open

to push the SoS degree from n^{ε} to $\Omega(n)$. Our current techniques do not handle this but we expect the lower bound to nevertheless hold.

8.4.2 Satisfying constraints exactly

In some of our lower bounds, 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, the moment matrix becomes much harder to analyze.

We do resolve it for the Sherrington-Kirkpatrick lower bound by using a rounding technique. This same issue also appeared in the SoS lower bounds for planted clique [10], which was fixed in a recent paper by Pang [140]. We leave it to future work to resolve this in general.

REFERENCES

[1] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. abs/1604.03423, 2020.

[2] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. abs/1604.03423, 2020.

[3] Genevera I Allen and Mirjana Maletić-Savatić. Sparse non-negative generalized pca with applications to metabolomics. *Bioinformatics*, 27(21):3029–3035, 2011.

[4] 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.

[5] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *Journal of the ACM (JACM)*, 62(5):1–25, 2015.

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

[7] 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.

[8] Ainesh Bakshi and Adarsh Prasad. Robust linear regression: Optimal rates in polynomial time. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 102–115, 2021.

[9] Afonso S. Bandeira, Dmitriy Kunisky, and Alexander S. Wein. Computational hardness of certifying bounds on constrained pca problems, 2019.

[10] B. Barak, S. B. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science*, pages 428–437, 2016.

[11] 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.

[12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326, 2012.

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

[14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 31–40, 2014.

[15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 143–151, 2015.

[16] Boaz Barak, Pravesh K Kothari, and David Steurer. Small-set expansion in shortcode graph and the 2-to-2 conjecture. *arXiv preprint arXiv:1804.08662*, 2018.

[17] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. In *ICM*, 2014.

[18] Avraham Ben-Aroya, Oded Regev, and Ronald De Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and ldcs. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, pages 477–486. IEEE, 2008.

[19] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 2141–2161. Institut Henri Poincaré, 2020.

[20] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, pages 1046–1066, 2013.

[21] Quentin Berthet, Philippe Rigollet, et al. Optimal detection of sparse principal components in high dimension. *The Annals of Statistics*, 41(4):1780–1815, 2013.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] K-H Borgwardt. The average number of pivot steps required by the simplex-method is polynomial. *Zeitschrift für Operations Research*, 26(1):157–177, 1982.

[27] Karl Heinx Borgwardt. Probabilistic analysis of the simplex method. In *DGOR/NSOR*, pages 564–575. Springer, 1988.

[28] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart. Moment inequalities for functions of independent random variables. *The Annals of Probability*, 33(2):514 – 560, 2005.

[29] Fernando GSL Brandao and Aram W Harrow. Quantum de finetti theorems under local measurements with applications. *Communications in Mathematical Physics*, 353(2):469–506, 2017.

[30] Mark Braverman, Young Kun Ko, Aviad Rubinstein, and Omri Weinstein. Eth hardness for densest-k-subgraph with perfect completeness. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1326–1341. SIAM, 2017.

[31] Matthew Brennan and Guy Bresler. Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. *arXiv preprint arXiv:1902.07380*, 2019.

[32] 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.

[33] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.

[34] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory*, pages 417–468. PMLR, 2019.

[35] Jonah Brown-Cohen and Prasad Raghavendra. Extended formulation lower bounds for refuting random csps. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 305–324. SIAM, 2020.

[36] 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.

[37] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, 2006.

[38] Sourav Chatterjee. *Concentration inequalities with exchangeable pairs*. Stanford University, 2005.

[39] Sourav Chatterjee. Stein's method for concentration inequalities. *arXiv preprint math/0604352*, 2006.

[40] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *arXiv preprint arXiv:1402.1267*, 2014.

[41] Eden Chlamtáč and Pasin Manurangsi. Sherali-adams integrality gaps matching the log-density threshold. *arXiv preprint arXiv:1804.07842*, 2018.

[42] Hyonho Chun and Sündüz Keleş. Expression quantitative trait loci mapping with multivariate sparse partial least squares regression. *Genetics*, 182(1):79–90, 2009.

[43] Amin Coja-Oghlan. The Lovász number of random graphs. *Combinatorics, Probability and Computing*, 14(4):439–465, 2005.

[44] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015.

[45] Andrea Crisanti and Tommaso Rizzo. Analysis of the ∞-replica symmetry breaking solution of the sherrington-kirkpatrick model. *Physical Review E*, 65(4):046137, 2002.

[46] Varsha Dani and Cristopher Moore. Independent sets in random graphs from the weighted second moment method. In *Approximation, randomization, and combinatorial optimization*, volume 6845 of *Lecture Notes in Comput. Sci.*, pages 472–482. Springer, Heidelberg, 2011.

[47] George Dantzig. *Linear programming and extensions*. Princeton university press, 2016.

[48] Amir Dembo, Andrea Montanari, and Subhabrata Sen. Extremal cuts of sparse random graphs. *The Annals of Probability*, 45(2):1190–1217, 2017.

[49] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory*, pages 523–562. PMLR, 2015.

[50] Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. *The Journal of Machine Learning Research*, 17(1):4913–4953, 2016.

[51] 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.

[52] Jian Ding, Allan Sly, and Nike Sun. Maximum independent sets on random regular graphs. *Acta Math.*, 217(2):263–340, 2016.

[53] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 376–389, 2018.

[54] 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.

[55] Gábor Elek and Gábor Lippner. Borel oracles. an analytical approach to constant-time algorithms. *Proceedings of the American Mathematical Society*, 138(8):2939–2947, 2010.

[56] Zhou Fan and Andrea Montanari. How well do local algorithms solve semidefinite programs? In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614, 2017.

[57] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000.

[58] Uriel Feige, David Peleg, and Guy Kortsarz. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001.

[59] Uriel Feige, Michael Seltser, et al. *On the densest k-subgraph problem*. Citeseer, 1997.

[60] Vitaly Feldman. *Statistical Query Learning*, pages 2090–2095. Springer New York, New York, NY, 2016.

[61] 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.

[62] Vitaly Feldman, Cristobal Guzman, and Santosh Vempala. Statistical query algorithms for mean vector estimation and stochastic convex optimization. *Mathematics of Operations Research*, 2021.

[63] 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.

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

[65] PJ Forrester, NC Snaith, and JJM Verbaarschot. Developments in random matrix theory. *Journal of Physics A: Mathematical and General*, 36(12):R1, 2003.

[66] 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.

[67] Zoltán Füredi and János Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[68] 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.

[69] Rong Ge and Tengyu Ma. Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms. *arXiv preprint arXiv:1504.05287*, 2015.

[70] 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)*, pages 954–965. IEEE, 2020.

[71] M. X. Goemans and D. P. Williamson. .878-approximation algorithms for MAX CUT and MAX 2SAT. In *Proceedings of the 26th ACM Symposium on Theory of Computing*, pages 422–431, 1994.

[72] 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*.

[73] Dima Grigoriev. Complexity of positivstellensatz proofs for the knapsack. *computational complexity*, 10(2):139–154, 2001.

[74] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

[75] Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

[76] 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.

[77] Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928. PMLR, 2015.

[78] Frank Hansen and Gert K Pedersen. Jensen's operator inequality. *Bulletin of the London Mathematical Society*, 35(4):553–564, 2003.

[79] Justin Holmgren and Alexander S Wein. Counterexamples to the low-degree conjecture. *arXiv preprint arXiv:2004.08454*, 2020.

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

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

[82] Samuel B Hopkins. Mean estimation with sub-gaussian rates in polynomial time. *The Annals of Statistics*, 48(2):1193–1213, 2020.

[83] Samuel B Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. *ACM Transactions on Algorithms (TALG)*, 14(3):1–31, 2018.

[84] Samuel B Hopkins, Pravesh K Kothari, and Aaron Potechin. Sos and planted clique: Tight analysis of mpw moments at all degrees and an optimal lower bound at degree four. *arXiv preprint arXiv:1507.05230*, 2015.

[85] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, pages 720–731. IEEE, 2017.

[86] Samuel B Hopkins, Tselil Schramm, and Jonathan Shi. A robust spectral algorithm for overcomplete tensor decomposition. In *Conference on Learning Theory*, pages 1683–1722. PMLR, 2019.

[87] Samuel B Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 178–191, 2016.

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

[89] 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.

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

[91] Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method. 2018.

[92] Carlos Hoppen and Nicholas Wormald. Local algorithms, regular graphs of large girth, and random regular graphs. *Combinatorica*, 38(3):619–664, 2018.

[93] De Huang and Joel A. Tropp. From Poincaré inequalities to nonlinear matrix concentration. *Bernoulli*, 27(3):1724 – 1744, 2021.

[94] Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *arXiv preprint arXiv:0901.4392*, 2009.

[95] Chris Jones, Aaron Potechin, Goutham Rajendran, Madhur Tulsiani, and Jeff Xu. Sum-of-squares lower bounds for sparse independent set. *IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, 2021.

[96] Sushrut Karmalkar, Adam R Klivans, and Pravesh K Kothari. List-decodable linear regression. *arXiv preprint arXiv:1905.05679*, 2019.

[97] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

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

[99] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.

[100] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[101] Subhash Khot. Ruling out ptas for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM Journal on Computing*, 36(4):1025–1071, 2006.

[102] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? In *Proceedings of the 45th IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004.

[103] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within $2-\varepsilon$. In *Proceedings of the 18th IEEE Conference on Computational Complexity*, 2003.

[104] Subhash Khot and Nisheeth Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into ℓ_1. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, pages 53–63, 2005.

[105] Bohdan Kivva, Goutham Rajendran, Pradeep Ravikumar, and Bryon Aragam. Learning latent causal graphs via mixture oracles. 2021.

[106] Victor Klee and George J Minty. How good is the simplex algorithm. *Inequalities*, 3(3):159–175, 1972.

[107] Pravesh Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, 2017.

[108] Pravesh K Kothari and Peter Manohar. A stress-free sum-of-squares lower bound for coloring. *arXiv preprint arXiv:2105.07517*, 2021.

[109] Pravesh K Kothari and Ruta Mehta. Sum-of-squares meets nash: lower bounds for finding any equilibrium. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1241–1248, 2018.

[110] Pravesh K Kothari and David Steurer. Outlier-robust moment-estimation via sum-of-squares. *arXiv preprint arXiv:1711.11581*, 2017.

[111] Robert Krauthgamer, Boaz Nadler, Dan Vilenchik, et al. Do semidefinite relaxations solve sparse pca up to the information limit? *The Annals of Statistics*, 43(3):1300–1322, 2015.

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

[113] Dmitriy Kunisky. Positivity-preserving extensions of sum-of-squares pseudomoments over the hypercube. *arXiv preprint arXiv:2009.07269*, 2020.

[114] Dmitriy Kunisky. *Spectral Barriers in Certification Problems*. PhD thesis, New York University, 2021.

[115] Dmitriy Kunisky and Afonso S. Bandeira. A tight degree 4 sum-of-squares lower bound for the sherrington-kirkpatrick hamiltonian. abs/1907.11686, 2019.

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

[117] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001.

[118] Massimo Lauria. Sum of squares and integer programming relaxations. Course at KTH Royal Institute of Technology.

[119] James R Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 567–576, 2015.

[120] L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. on Optimization*, 1(12):166–190, 1991.

[121] Tengyu Ma and Avi Wigderson. Sum-of-squares lower bounds for sparse pca. In *Advances in Neural Information Processing Systems*, pages 1612–1620, 2015.

[122] Zongming Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772–801, 2013.

[123] Angshul Majumdar. Image compression by sparse pca coding in curvelet domain. *Signal, image and video processing*, 3(1):27–34, 2009.

[124] Pasin Manurangsi. Almost-polynomial ratio eth-hardness of approximating densest k-subgraph. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 954–961, 2017.

[125] David W Matula. *The largest clique size in a random graph*. Department of Computer Science, Southern Methodist University Dallas, Texas . . . , 1976.

[126] Dhruv Medarametla and Aaron Potechin. Bounds on the norms of uniform low degree graph matrices. *RANDOM*, 2016.

[127] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 87–96, 2015.

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

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

[130] Ankur Moitra. Sum of squares in theoretical computer science. *Sum of Squares: Theory and Applications*, 77:83, 2020.

[131] Ankur Moitra and Alexander S Wein. Spectral methods from tensor networks. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 926–937, 2019.

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

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

[134] Nikhil Naikal, Allen Y Yang, and S Shankar Sastry. Informative feature selection for object recognition via sparse pca. In *2011 International Conference on Computer Vision*, pages 818–825. IEEE, 2011.

[135] Yurii Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, pages 405–440. Springer, 2000.

[136] R. O'Donnell. Some topics in analysis of Boolean functions. In *STOC*, pages 569–578, 2008.

[137] Ryan O'Donnell. Sos is not obviously automatizable, even approximately. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

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

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

[140] Shuo Pang. SOS lower bound for exact planted clique. In *36th Computational Complexity Conference*, volume 200 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. 26, 63. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2021.

[141] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, 1982.

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

[143] Giorgio Parisi. A sequence of approximated solutions to the sk model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

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

[145] Pablo A Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical programming*, 96(2):293–320, 2003.

[146] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[147] Daniel Paulin, Lester Mackey, and Joel A. Tropp. Efron–stein inequalities for random matrices. *Ann. Probab.*, 44(5):3431–3473, 09 2016.

[148] Dénes Petz. A survey of certain trace inequalities. *Banach Center Publications*, 30(1):287–298, 1994.

[149] Aaron Potechin and Goutham Rajendran. Machinery for proving sum-of-squares lower bounds on certification problems. *arXiv preprint arXiv:2011.04253*, 2020.

[150] Aaron Potechin and David Steurer. Exact tensor completion with sum-of-squares. *arXiv preprint arXiv:1702.06237*, 2017.

[151] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

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

[153] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random csps below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 121–131, 2017.

[154] Prasad Raghavendra and Tselil Schramm. Tight lower bounds for planted clique in the degree-4 sos program. *arXiv preprint arXiv:1507.05136*, 2015.

[155] Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. In *Proceedings of the 44th International Colloquium on Automata, Languages and Programming*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[156] Goutham Rajendran. Combinatorial optimization via the sum of squares hierarchy. 2018.

[157] Goutham Rajendran, Bohdan Kivva, Ming Gao, and Bryon Aragam. Structure learning in polynomial time: Greedy algorithms, bregman information, and exponential families. 2021.

[158] Goutham Rajendran and Madhur Tulsiani. Nonlinear concentration via matrix efronstein. *Manuscript*, 2021.

[159] Bruce Reznick. Some concrete aspects of hilbert's 17th problem. *Contemporary mathematics*, 253:251–272, 2000.

[160] Emile Richard and Andrea Montanari. A statistical model for tensor pca. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

[161] Benjamin Rossman. *Average-case complexity of detecting cliques*. PhD thesis, Massachusetts Institute of Technology, 2010.

[162] Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM Journal on Computing*, 43(1):256–279, 2014.

[163] Gian-Carlo Rota and Timothy C. Wallstrom. Stochastic integrals: a combinatorial approach. *Ann. Probab.*, 25(3):1257–1283, 1997.

[164] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008.

[165] Tselil Schramm and David Steurer. Fast and robust tensor decomposition with applications to dictionary learning. In *Conference on Learning Theory*, pages 1760–1793. PMLR, 2017.

[166] Warren Schudy and Maxim Sviridenko. Bernstein-like concentration and moment inequalities for polynomials of independent random variables: multilinear case. *arXiv preprint arXiv:1109.5193*, 2011.

[167] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990.

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

[169] Naum Zuselevich Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics*, 23(5):695–700, 1987.

[170] Steve Smale. On the average number of steps of the simplex method of linear programming. *Mathematical programming*, 27(3):241–262, 1983.

[171] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[172] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory*, pages 583–602. University of California Press, 1972.

[173] Charles Stein. Approximate computation of expectations. IMS, 1986.

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

[175] Khot Subhash, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 592–601. IEEE, 2018.

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

[177] Kean Ming Tan, Ashley Petersen, and Daniela Witten. Classification of rna-seq data. In *Statistical analysis of next generation sequencing data*, pages 219–246. Springer, 2014.

[178] Ryota Tomioka and Taiji Suzuki. Spectral norm of random tensors. *arXiv preprint arXiv:1407.1870*, 2014.

[179] Luca Trevisan. Approximation algorithms for unique games. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005.

[180] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[181] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[182] Van H Vu. Spectral norm of random matrices. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 423–430, 2005.

[183] Dong Wang, Huchuan Lu, and Ming-Hsuan Yang. Online object tracking with sparse prototypes. *IEEE transactions on image processing*, 22(1):314–325, 2012.

[184] Tengyao Wang, Quentin Berthet, Richard J Samworth, et al. Statistical and computational trade-offs in estimation of sparse principal components. *The Annals of Statistics*, 44(5):1896–1930, 2016.

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

[186] Ilias Zadik, Min Jae Song, Alexander S. Wein, and Joan Bruna. Lattice-based methods surpass sum-of-squares in clustering, 2021.

[187] Lenka Zdeborová and Stefan Boettcher. A conjecture on the maximum cut and bisection width in random regular graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(02):P02020, 2010.

[188] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.