GRAPH MATRICES UNDER THE MULTIVARIATE SETTING

by

IMRAN HOSSAIN

Submitted in partial fulfillment of the requirements

for the degree of Master of Science

Department of Computing and Information Sciences

CASE WESTERN RESERVE UNIVERSITY

May 2022

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis of

Imran Hossain

candidate for the degree of **Master of Science***.

Committee Chair

Dr. Harold Connamacher

Committee Member

Dr. Vincenzo Liberatore

Committee Member

Dr. Mark Meckes

Date of Defense

March 31, 2022

*We also certify that written approval has been obtained

for any proprietary material contained therein.

Contents

List of Figures

List of Acronyms & Symbols

$C(\phi)$ constraint graph under mapping ϕ

N_i i-th ground set

U_α Left vertex set of shape α

V_α Right vertex set of shape α

W_α Middle vertex set of shape α

$\mathcal{C}_{(\alpha,2q)}$ Class of constraint graphs on $H(\alpha, 2q)$

Ω input distribution

$\mathbb{E}_\Omega[f(x)]$ expectation of $f(x)$ over distribution Ω

$\widetilde{O}$ big O notation ignoring polylogarithmic factors

α shape

χ_E Fourier character of edge set E

$\boldsymbol{M}_\alpha$ graph matrix with shape α

$\mu_\Omega(j)$ j-th moment of distribution Ω

ϕ mapping of a vertex set onto ground sets

d dimension of input distribution

f_Ω density function of distribution Ω

$h_k(x)$ k-th basis polynomial

n size of largest ground set

$p_k(x)$ canonical basis monomial of degree k

iid independent and identically distributed

w.h.p. with high probability

Graph Matrices under the Multivariate Setting

Abstract

by

IMRAN HOSSAIN

We expand on the framework of graph matrices first introduced by Ahn et al. [1], which are a class of random matrices whose entries' dependence can be described by a small graph. While Ahn et al. assume that a univariate distribution underlies this dependence, we relax this assumption and introduce graph matrices whose input structure is derived from a multivariate probability distribution. We then show spectral norm bounds on these graph matrices as being consistent with those under the univariate setting using the trace power method. Our result expands Ahn et al's work by allowing for random matrices with more complicated dependencies between elements. We present potential applications that have such dependencies under the multivariate setting in fields such as graph theory.

Chapter 1

Introduction

In this thesis, we expand on the work of Ahn, Medarametla, and Potechin [1], who introduce graph matrices, a class of matrices whose random-valued entries depend on some input probability distribution. In particular, this dependence can be characterized by a small graph. While the theory of random matrices has been studied for many decades, with applications in nuclear physics [2], statistics [3], and quantum physics [4], typically these applications make the assumption that entries of these random matrices are independent. The framework that Ahn et al. introduce using graph matrices allow for the analysis of large dependent random structures (whose size is described by a parameter n) in a way that makes their analysis easier and more mechanical. These structures can come in the form of (but are not limited to) matrices, polynomials, or graphs, e.g. the Erdős-Rényi model of random graphs on n vertices, which we discuss in Section 2.2.1. This work arose out of Potechin's work involving the sum-of-squares hierarchy [5], [6], and so the applications discussed often involve problems that end up analyzing this hierarchy. In their work, Ahn et al. showed spectral norm bounds on graph matrices and have used them to more easily reproduce upper bound analyses of the sum-of-squares hierarchy. This sum-of-squares hierarchy is often used to show bounds on refutation algorithms in constraint satisfaction problems [7], a core topic of interest in theoretical computer science, as well as in inference problems in machine learning such as tensor completion [8]. The proofs of some of these bounds often required clever arguments, whereas graph matrices makes their analysis more mechanical and straightforward.

One core assumption Ahn et al. made in their work is that the entries of a graph matrix are expressed in terms of random variables that are sampled from an input distribution on one variable, i.e. a univariate probability distribution. For example, the presence of edges in a random graph on $[n]$ vertices are events sampled from the Rademacher distribution. As another example, the coefficients of a polynomial in n variables are products of random variables sampled from the standard normal

distribution [1, Section 9.2]. Notably, this assumption does not generally allow for random variables that have nonzero covariance, since samples are independent of one another.

In our work, we relax this univariate assumption and port the graph matrix framework to the multivariate setting, letting these random variables instead be sampled from a probability distribution in $d \geq 2$ dimensions. This allows for distributions that follow a stricter set of rules that either cannot be expressed or are much more difficult to express using a univariate distribution. By consequence, this increases the framework's versatility, allowing its use in more applications. We begin in Chapter 2 with preliminary backgrounds in linear algebra, graph theory, and probability theory. We then introduce the graph matrix framework in Section 2.2, starting with a warm-up example from [1, Section 2.1] of a clique indicator matrix for Erdős-Rényi random graphs, then moving on to more generalized definitions. Towards the end of the chapter, we reproduce the main theorem of Ahn et al. [1], bounding the spectral norm of a graph matrix.

In Chapter 3, as a preface to our contributions, we present the techniques and theorems used to prove this bound, namely the trace power method, constraint graphs, and the handling of an input distribution. We then move on to the multivariate setting and modify these techniques in Chapter 4. In particular, we modify theorems used to bound the input distribution and properties of constraint graphs. Some of these modifications required straightforward proofs that were essentially modified from Ahn et al's work, while others required more sophisticated arguments. This ultimately culminates in the same upper bound on the spectral norm on a graph matrix in the multivariate setting as in the univariate setting, though these bounds can be refined given a specific input distribution.

Chapter 5 then presents a few potential applications that would either be much more difficult or impossible to work with in the univariate setting. First in Chapter 5,

we extend Ahn et al's application of bounding polynomials over the unit sphere. While in [1, Section 9] the coefficients of these polynomials are random variables from the standard normal distribution, we modify one of their examples to use a multivariate normal distribution with non-zero covariance and show a bound on this polynomial over the unit sphere. We then revisit the clique indicator matrix in Section 5.2, modify it by introducing a new random graph model with edge coloring and creating a monochromatic clique probability matrix, then show spectral norm bounds on this matrix. Finally, in Section 5.3 we discuss a potential application in quantum physics, using graph matrices to represent expressions involving pure states and density matrices.

Chapter 2

Background

In this chapter we provide the necessary background on top of which our work is built. Section 2.1 provides a basic introduction to matrices and graphs as well as related concepts and notation used in the rest of the paper. Section 2.2 introduces the concept of graph matrices, a type of random matrix whose dependencies can be characterized by a graph. Section 2.3 defines some properties of probability distributions, specifically those that are multivariate.

2.1 Preliminary Definitions & Notation

For some $n \in \mathbb{N}$, we write $[n]$ to mean the set $\{1, 2, \ldots, n\}$.

Let M be some matrix over some field $\mathbb{F}$. We use the notation $\boldsymbol{M}(i, j)$ to signify the element in $\boldsymbol{M}$ at index (i, j), and the notation $\boldsymbol{M}(i)$ to signify the ith column vector in $\boldsymbol{M}$. Similarly, for some order-d tensor $\boldsymbol{T}$ where $d > 2$, we use $\boldsymbol{T}(i_1, \ldots, i_d)$ to signify a specific element in $\boldsymbol{T}$. We also use the notation $\boldsymbol{T}(i_1, \ldots, i_{d-1})$ to denote a "subtensor" of $\boldsymbol{T}$ taken by fixing the first $d-1$ indices. For example, for a $n_1 \times n_2 \times n_3$ tensor $\boldsymbol{T}$ (visually, a "cube" of elements), $\boldsymbol{T}(i, j)$ is the vector (or "rod") of elements at position (i, j) in the tensor.

In the field of linear algebra, there are several matrix norms (all denoted with $\|\cdot\|$), all of which obey the properties listed below.

Definition 2.1 (Properties of matrix norms)**.** Given matrices $\boldsymbol{M}, \boldsymbol{N}$ over $\mathbb{F}$ and scalar $k \in \mathbb{F}$, a matrix norm $\|\cdot\|$ must satisfy the following properties:

1. (Non-negativity). $\|\boldsymbol{M}\| \geq 0$
2. (Definiteness). $\|\boldsymbol{M}\| = 0$ if and only if $\boldsymbol{M}$ is the zero matrix
3. (Scaling). $\|k\boldsymbol{M}\| = |k|\|\boldsymbol{M}\|$
4. (Triangle inequality). $\|\boldsymbol{M} + \boldsymbol{N}\| \leq \|\boldsymbol{M}\| + \|\boldsymbol{N}\|$

In this thesis we focus on the spectral norm, which is roughly how much a vector is scaled by a matrix. Throughout the rest of this paper, we write $\|\cdot\|$ to denote the spectral norm. We give a concrete definition below.

Definition 2.2 (Spectral norm)**.** Given a matrix $\boldsymbol{M}$ over a field $\mathbb{F}$, the *spectral norm* of $\boldsymbol{M}$, written as $\|\boldsymbol{M}\|$, is defined as

$$\|\boldsymbol{M}\| = \sup_{\boldsymbol{v}\neq\boldsymbol{0}} \frac{\|\boldsymbol{M}\boldsymbol{v}\|_2}{\|\boldsymbol{v}\|_2} = \sup_{\|\boldsymbol{v}\|_2=1} \|\boldsymbol{M}\boldsymbol{v}\|_2,$$

where $\boldsymbol{v}$ is some vector over $\mathbb{F}$ and $\|\boldsymbol{v}\|_2$ is the Euclidean norm, or 2-norm, of $\boldsymbol{v}$. The spectral norm is also known as the operator norm or induced 2-norm.

We also define some additional terminology concerning matrices which we use throughout the paper.

Definition 2.3 (Trace)**.** For an $n \times n$ matrix $\boldsymbol{M}$, the trace of $\boldsymbol{M}$, written as $\mathrm{Tr}(\boldsymbol{M})$, is the sum of the entries on its main diagonal. That is,

$$\mathrm{Tr}(\boldsymbol{M}) = \sum_{i=1}^{n} \boldsymbol{M}(i,i).$$

Definition 2.4 (Definiteness)**.** For some real-valued square matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, we say that $\boldsymbol{M}$ is *positive semi-definite*, written $\boldsymbol{M} \succeq 0$, if $\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Given a graph G, we write its vertex set as $V(G)$ and edge set as $E(G)$. Each edge $e = \{v_1, v_2\} \in E(G)$ is a set of two vertices $v_1, v_2 \in V(G)$. In a *hypergraph* H, an edge can connect any number of vertices. That is, $E(H)$ is made up of (non-empty) subsets of $V(H)$ of any size. We introduce some properties of graphs below.

Definition 2.5 (Clique)**.** We say that a graph G is a *clique* if any two vertices in G are adjacent.

Definition 2.6 (U, V-separator)**.** For a graph G and subsets $U, V \subseteq V(G)$ of its vertex set, we say that $S \subseteq V(G)$ is a *vertex separator* between U and V, or a U, V*-separator*,

if all paths between U and V go through S. We call S a *minimum U, V-separator* if no other U, V-separators are smaller. Note that paths of length zero are allowed, so $S \supseteq U \cap V$.

Definition 2.7 (Matching)**.** For a graph G, a *matching* is a set of edges $M \subseteq E(G)$ such that no two edges are incident to the same vertex. We call M a *maximum matching* if no other matching on G is larger. For subsets $U, V \subseteq V(G)$, a U, V-*matching* is a set of edges $M \subseteq E(G)$ such that no two edges are incident to the same vertex, and each edge has an endpoint in U and another in V.

Our work also deals with probability distributions, so we provide some definitions.

Definition 2.8 (Support)**.** Let Ω be some probability distribution. The *support* of Ω is defined as $\text{supp}(\Omega) = \{x \mid X \sim \Omega,\ p(X = x) > 0\}$, the set of all values with nonzero probability under Ω. Intuitively, this is the set of all values which can be sampled from Ω.

Definition 2.9 (Moments)**.** Let $X \sim \Omega$ be some random variable drawn from a probability distribution, and $\mathbb{E}_\Omega(f(X))$ be the expectation of $f(X)$ over Ω. For all $j \in \mathbb{N}$, the j-th *moment* of Ω is defined as $\mu_\Omega(j) = \mathbb{E}_\Omega[X^j]$.

Definition 2.10 (Polynomial vector space)**.** The vector space of polynomials $\mathbb{F}[x]$ is the set of all expressions of the form

$$c_0 + c_1 x + \ldots + c_k x^k$$

where coefficients c_i are elements of the field $\mathbb{F}$. Further, we define the *canonical basis* of $\mathbb{F}[x]$ to be the set of all monomials $\{1, x, x_2, x_3, \ldots\}$.

Lastly, we provide a definition regarding asymptotic complexity.

Definition 2.11 (Soft-O notation)**.** We define the notation $\widetilde{O}(f(n))$, called *soft-O*, as shorthand for big O notation ignoring polylogarithmic factors: $O(f(n) \log^c n)$ for

some $c \geq 0$. That is, a function $g(n) = \widetilde{O}(f(n))$ when it grows asymptotically no faster than a polylogarithmic factor of $f(n)$.

2.2 Graph Matrices

Many problems in fields such as quantum physics [9] or sum-of-squares [10] boil down to analyzing some class of matrices. In many cases, entries of this matrix may be random variables drawn from some distribution. We call these *random matrices*. The properties of random matrices whose entries are *independent* random variables have been well-studied in random matrix theory, a field with many applications in physics, mathematics, and computer science. Some examples of such properties are as follows:

- For a symmetric random $n \times n$ matrix whose entries are iid random variables from the standard normal distribution $\mathcal{N}(0, 1)$, Wigner's Semicircle Law [11] describes the limit of the distribution of eigenvalues, or the *limiting spectral distribution*, as $n \to \infty$ stating that its probability density function is a semicircle.

- For a random $n \times n$ matrix whose entries are iid complex random variables with unit variance, Girko's circular law [12] states that the limiting distribution of eigenvalues as $n \to \infty$ follows the uniform distribution over the unit disc on the complex plane.

- Given a symmetric $n \times n$ matrix whose off-diagonal entries are iid random variables drawn from a distribution with unit variance and whose diagonal entries are iid from another such distribution, the spectral norm of the matrix is bounded from above by $2(1 + o(1))\sqrt{n}$. [13], [14]

As seen in these examples, random matrix theory is often interested how properties of a random matrix behave asymptotically as the size of the matrix increases.

Now, rather than a matrix whose entries are iid, consider a matrix whose entries are *dependent* on some underlying random input distribution such that entries are no longer independent of each other. Ahn et al. [1] introduce such a class of random matrices, called *graph matrices*. As we will discuss in this section, we can express this dependence with a small graph we call a *shape*. In their paper, Ahn et al. prove a rough bound on the spectral norm of these matrices and demonstrate applications of these bounds by reproducing results of various theorems concerning the sum-of-squares hierarchy.

2.2.1 Warm-up: Clique Indicator Matrix

To introduce the concept of a graph matrix, we consider a simple example excerpted from [1, Section 2.1] throughout this subsection. We will revisit this example and modify it later in Section 5.2.

Example 2.12. *Consider the distribution of graphs G whose vertex sets $V(G) = [n]$, and each possible edge has independent probability $1/2$ of appearing in G. This is known as an Erdős-Rényi distribution of random graphs, denoted $G(n, 1/2)$.*

Suppose we want to analyze the following $n \cdot (n-1) \times n \cdot (n-1)$ indicator matrix CLIQUE*, defined as*

$$\mathsf{CLIQUE}((i_1, i_2), (j_1, j_2)) := \begin{cases} 1 & \text{if } i_1, i_2, j_1, j_2 \text{ are distinct and form a clique in } G \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Note that the matrix indices are tuples of 2 vertices or *monomials*, and that each of the indices are elements of the vertex set $[n]$. Also note that the entries of this matrix are random, but not independent, as they each depend on varying factors of the same underlying graph $G \sim G(n, 1/2)$. We provide the following definitions to formalize the indices of a matrix.

Definition 2.13 (Ground set)**.** Since each monomial takes on a value from $[n]$, we call this set a *ground set*.

Definition 2.14 (Matrix index)**.** A matrix index $A = (a_1, \ldots, a_m)$ is an m-tuple of distinct indices for some $m \in \mathbb{N}$, where each a_i is drawn from the ground set $[n]$. We define $V(A)$ to be the set of vertices $\{a_1, \ldots, a_m\}$.

One way we could try to analyze this matrix is by decomposing each of its entries $\boldsymbol{M}((i_1, i_2), (j_1, j_2))$ into all possible graphs on the 4 vertices $\{i_1, i_2, j_1, j_2\}$, then querying the existence of each edge in those graphs. We call this analysis by decomposition *Fourier analysis*.

To perform this analysis, we first need a way to encode the existence of some edge $\{k_1, k_2\}$ in G. We encode these as variables $\chi_{\{k_1,k_2\}}$, where

$$\chi_{\{k_1,k_2\}} := \begin{cases} +1 & \text{if } \{k_1, k_2\} \in E(G) \\ -1 & \text{if } \{k_1, k_2\} \notin E(G). \end{cases}$$

We can then decompose the CLIQUE matrix into a sum over each possible subgraph on four vertices as in the following proposition.

Proposition 2.15 ([1, Equation 2.2])**.** *For distinct* $i_1, i_2, j_1, j_2 \in [n]$*,*

$$\mathsf{CLIQUE}((i_1, i_2), (j_1, j_2)) = \frac{1}{2^6} \cdot \sum_{R\text{: graph on } \{i_1,i_2,j_1,j_2\}} \prod_{e \in E(R)} \chi_e. \tag{2.2}$$

If a clique exists for a particular index of CLIQUE, then each factor $\chi_e = +1$ in the product, and this is true across the sum over all graphs over those indices. Otherwise some edge e is missing, and $\chi_e = -1$ in half of the summands, resulting in a perfect cancellation.

Note how each summand is a product of all the χ_e variables on the edge set of each possible graph on the given indices. We call this the *Fourier character* of the

edge set.

Definition 2.16 (Fourier characters)**.** Given some edge set E whose elements are pairs of distinct indices from the ground set $[n]$, the *Fourier character* of E is defined as $\chi_E = \prod_{e \in E} \chi_e$.

Consider a specific entry of CLIQUE, taken by mapping indices i_1, i_2, j_1, j_2 to distinct values from $[n]$. We call this a *realization* of the variables in the matrix index. For example,

$$\mathsf{CLIQUE}((1,2),(4,5)) = \frac{1}{2^6} \cdot \sum_{R\text{: graph on } \{1,2,4,5\}} \chi_{E(R)}(G)$$

We call each graph R corresponding to a summand in the entry a *ribbon*, for which we give a precise definition in Definition 2.17. Each ribbon has a set of left and right vertices, signified by the matrix index, which we call A_R and B_R. Ribbons may also have middle vertices distinct from the matrix index, which we call C_R. In this case, $A = (1,2)$ and $B = (4,5)$. We define the Fourier character of a ribbon R as that of its edge set $E(R)$. That is, $\chi_R := \chi_{E(R)}$. Thus, we can express the above entry using ribbons and Fourier characters as:

$$\mathsf{CLIQUE}((1,2),(4,5)) = \frac{1}{2^6} \cdot \sum_{\substack{\text{ribbon } R \\ V(R)=\{1,2,4,5\} \\ A_R=(1,2),\ B_R=(4,5)}} \chi_R.$$

Consider one of the terms in Equation (2.2). For example, one such term is $\chi_{\{1,5\}}\chi_{\{2,4\}}\chi_{\{2,5\}}$. Figure 2.1 shows the ribbon corresponding to this term.

We give a formal definition for ribbons below.

Definition 2.17 (Ribbons)**.** A ribbon R consists of (possibly intersecting) matrix indices A_R and B_R, a set of additional distinct indices C_R, and a set of pairs of indices $E(R)$. We represent R as a graph with left vertices $V(A_R)$, right vertices $V(B_R)$,

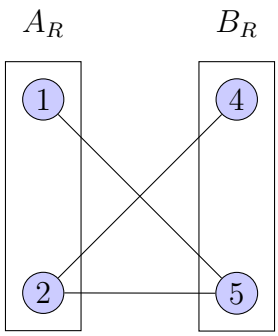


Figure 2.1: Ribbon R where $V(R) = \{1, 2, 4, 5\}$, $A_R = (1, 2)$, $B_R = (4, 5)$ and $E(R) = \{\{1, 5\}, \{2, 4\}, \{2, 5\}\}$.

middle vertices $V(C_R)$, and edge set $E(R)$. The left and right vertices may intersect, but the middle vertices are distinct. The Fourier character χ_R of a ribbon is defined as that of its edge set, $\chi_{E(R)}$.

Note how, in Figure 2.1, the left and right vertices in R correspond to the first and second matrix indices in $\mathsf{CLIQUE}((1, 2), (4, 5))$, respectively.

We can then consider all terms across all entries that have the same configuration, or *shape*, as this ribbon. Call this shape α. To do this, we replace the exact assignments $1, 2, 4, 5$ with free variables u_1, u_2, v_1, v_2, respectively. Each of these variables, when realized, will take on values from the ground set $[n]$. Additionally, we replace the exact matrix indices A_R and B_R with tuples $U_\alpha = (u_1, u_2)$ and $V_\alpha = (v_1, v_2)$, respectively. We call these tuples *index shapes*, and say that a matrix index A has index shape U if they are the same size. Figure 2.2 shows the shape of the ribbon R in the above example.

Shapes then are essentially an abstraction of ribbons, where indices are replaced by free variables. To formalize this, we give a definition for shapes.

Definition 2.18 (Shapes)**.** A shape α consists of index shapes U_α and V_α, a set of additional distinct indices W_α, and a set of pairs of variables $E(\alpha)$ taken from any of those sets. We represent α as a graph with left vertices $V(U_\alpha)$, right vertices $V(V_\alpha)$, middle vertices $V(W_\alpha)$, and edge set $E(\alpha)$ between these vertices. The left and right

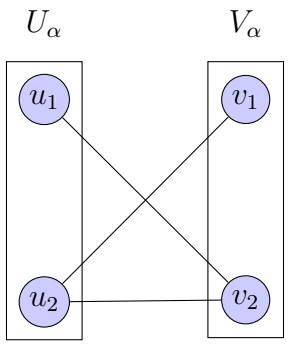


Figure 2.2: Shape α where $V(\alpha) = \{u_1, u_2, v_1, v_2\}$, $U_\alpha = (u_1, u_2)$, $V_\alpha = (v_1, v_2)$ and $E(R) = \{\{u_1, v_2\}, \{u_2, v_1\}, \{u_2, v_2\}\}$, abstracting the ribbon in Figure 2.1.

vertices may intersect, but the middle vertices are distinct. Variables in $U_\alpha, V_\alpha, W_\alpha$ are written as u_i, v_i, w_i, respectively.

Now, we can rewrite the CLIQUE decomposition in Equation (2.2) using shapes. Given a specific shape, we take all ribbons with that shape together to construct a matrix. In other words, for each shape α, we take all ribbons such that we can map the shape variables to the ribbon's indices. We then construct a matrix $\boldsymbol{M}_\alpha$ from this as follows:

$$\boldsymbol{M}_\alpha((i_1, i_2), (j_1, j_2)) := \begin{cases} \chi_R & \text{if } i_1, i_2, j_1, j_2 \text{ are distinct} \\ 0 & \text{otherwise} \end{cases},$$

where R is a ribbon with shape α such that $A_R = (i_1, i_2)$ and $B_R = (j_1, j_2)$. We call this matrix $\boldsymbol{M}_\alpha$ a *graph matrix*, and give a precise definition below.

Definition 2.19 (Graph matrices)**.** Given a shape α, the graph matrix $\boldsymbol{M}_\alpha$ is a $\frac{n!}{(n-|U_\alpha|)!} \times \frac{n!}{(n-|V_\alpha|)!}$ matrix where, for indices $A, B \subseteq [n]$ such that $|A| = |U_\alpha|$ and $|B| = |V_\alpha|$,

$$\boldsymbol{M}_\alpha(A, B) = \sum_{\text{ribbon } R\text{: } R \text{ has shape } \alpha} \chi_R.$$

We can then rewrite the entire CLIQUE indicator matrix using graph matrices over each applicable shape. Recall that CLIQUE is indexed using 2-tuples. This means we

sum over all shapes with 2 left vertices and 2 right vertices (i.e. $|U_\alpha| = |V_\alpha| = 2$), giving us the following equation:

$$\mathsf{CLIQUE} = \frac{1}{2^6} \cdot \sum_{\substack{\text{shape } \alpha \\ V(\alpha)=\{u_1,u_2,v_1,v_2\} \\ U_\alpha=(u_1,u_2),\ V_\alpha=(v_1,v_2)}} \boldsymbol{M}_\alpha \tag{2.3}$$

By the triangle inequality, the norm of CLIQUE then involves bounding the norm of each of its parts $\boldsymbol{M}_\alpha$, which Ahn et al. demonstrate in [1, Section 2.5]. For completeness, we restate their informal norm bounds on a graph matrix with some shape α.

Theorem 2.20 ([1, Theorem 2.24])**.** *Let α be a shape, u be the number of isolated middle vertices in W_α, and $s_{\min}$ be the minimum size of a U_α, V_α-separator. Then, w.h.p., $\|\boldsymbol{M}_\alpha\| \leq \widetilde{O}\left(n^{\frac{1}{2}(|V(\alpha)|+u-s_{\min})}\right)$.*

We show a more formal bound on the norm of a graph matrix in Theorem 2.35.

Next, we show rough norm bounds on CLIQUE.

Proposition 2.21. $\|\mathsf{CLIQUE}\| = \widetilde{O}(n^2)$.

Proof. Given the expression for CLIQUE in Equation 2.3, all shapes α have 4 vertices, i.e. $|V(\alpha)| = 4$ and no middle vertices. To show a norm bound, we split these shapes into three groups:

- For shapes where there are no edges between U_α and V_α, the size of a minimum U_α, V_α-separator is $s_{\min} = 0$, so $\|\boldsymbol{M}_\alpha\| = \widetilde{O}(n^2)$.
- For shapes with a maximum U_α, V_α-matching of size 1, the size of a minimum U_α, V_α-separator is $s_{\min} = 1$, so $\|\boldsymbol{M}_\alpha\| = \widetilde{O}\left(n^{\frac{3}{2}}\right)$.
- Lastly, for shapes with a maximum U_α, V_α-matching of size 2, the size of a minimum U_α, V_α-separator is $s_{\min} = 2$, so $\|\boldsymbol{M}_\alpha\| = \widetilde{O}(n)$.

Because matrix norms follow the triangle inequality as in Definition 2.1, $\|\mathsf{CLIQUE}\|$ is bounded from above by the sum of these individual norms. Adding these all together, the $\widetilde{O}(n^2)$ terms dominates all others, so we obtain a norm bound of $\widetilde{O}(n^2)$, concluding the proof. □

2.2.2 Generalized Definitions

Ahn et al. generalize their technique for bounding the norm of graph matrices beyond the Erdős-Rényi model, to bounding the norm for random matrices whose dependence can be described using a small hypergraph. This generalization extends the potential applications of graph matrices. For instance, Ahn et al. use graph matrices to replicate various proofs in papers that use the sum-of-squares hierarchy, such as in the work of Hopkins et al. [15]. To show this, we introduce an example of a homogeneous polynomial of even degree and demonstrate how we can upper bound the value of this polynomial over the unit sphere using graph matrices. We discuss this application in more detail in Section 5.1.

Example 2.22. *Rather than a random graph, suppose the input is an $m \times n$ matrix $\boldsymbol{A}$ with off-diagonal entries drawn from the standard normal distribution $\mathcal{N}(0,1)$, and that we would like to bound polynomials whose coefficients are factors of entries from $\boldsymbol{A}$. For example, suppose row index $i \in [m]$ and column indices $j_1, j_2 \in [n]$, and the polynomial*

$$f(x_1, \ldots, x_n) = \sum_{j_1 \neq j_2 \in [n]} \left[\sum_{i \in [m]} \boldsymbol{A}(i, j_1)(\boldsymbol{A}(i, j_2)^2 - 1) \right] x_{j_1} x_{j_2}. \tag{2.4}$$

If we construct another matrix $\boldsymbol{M}$ whose entries are coefficients of this polynomial, it turns out that bounding the norm of this matrix will upper bound the value of the polynomial over the unit sphere, and we can derive this bound using a generalized graph matrix. To do this, we need to:

1. Represent entries $\boldsymbol{A}(i,j)$ by an edge between some vertices that represent indices $i \in [m]$ and $j \in [n]$. Since these indices are taken from different ground sets, we represent them by two different types of vertices: row and column vertices.

2. Represent varying-degree polynomials of the input entries, such as $\boldsymbol{A}(i,j)^2$. To achieve this, we project these polynomials onto a basis for the input distribution $\mathcal{N}(0,1)$, then create labeled edges for each element of the basis that appears in the projection.

Towards these ends, we provide the generalized framework for graph matrices over a given input distribution Ω as given in [1]. We will return to this specific example and bound its value over the unit sphere later in Example 5.5. First, we generalize definitions from the previous section. Index monomials can now consist of variables of multiple types (e.g. row and column indices), so we assume there are t such types of variables. Thus, we define notation for the ground sets of different types.

Definition 2.23 (Ground sets [1, Definition 7.6])**.** For all $i \in [t]$, we write N_i to mean the i-th ground set of indices for variables of type i. Usually, $N_i = [n_i]$ for some $n_i \in \mathbb{N}$.

In representing Equation (2.4), we have two types of variables: those representing row indices and column indices, so our ground sets would be $[m]$ and $[n]$ (the dimensions of $\boldsymbol{A}$), respectively.

Next, we generalize the matrix indices in Definition 2.14. Note that entries of the input structure (e.g. a polynomial or a matrix like CLIQUE) can now be expressed in terms of higher-degree polynomials such as $(\boldsymbol{A}_{ij}^4 - 1)\boldsymbol{A}_{ij'}^2$ instead of simply the product of ± 1 entries as in $\prod_e \chi_e$. We therefore need a way to encode these expressions in the indices of the graph matrix. Below, we define a new type of matrix index by decomposing expressions by type and exponent.

Definition 2.24 (Generalized matrix indices [1, Definition 7.10])**.** A generalized matrix index $A = \{A_i\}$ is a set of disjoint matrix index pieces $A_i = ((a_1, \ldots, a_m), t_i, p_i)$, where each $a_j \in N_{t_i}$ represents the jth monomial of type t_i raised to the power p_i. All pieces (for $i < j$) are ordered by type ($t_i < t_j$) then exponent ($t_i = t_j \implies p_i < p_j$).

The size of A is a tuple $|A| = (|A_1|, |A_2|, \ldots)$, where $|A_i|$ is the number of indices in A_i.

To clarify the definition above, we provide an example of a matrix index.

Example 2.25. *Suppose we want to represent the monomial $x_1x_2x_1^2x_3^2y_1y_2$, where x_i and y_i are variables of two different types. To represent this monomial as a matrix index, we perform the following:*

1. *Decompose the expression by type, then power: x_1x_2, $x_1^2x_3^2$, y_1y_2.*
2. *Create index pieces for each decomposition: $A_1 = ((1, 2), 1, 1)$, $A_2 = ((1, 3), 1, 2)$, $A_3 = ((1, 2), 2, 1)$.*
3. *Compose the pieces together to create the matrix index $A = \{A_1, A_2, A_3\}$.*

We can generalize the index shapes similarly as $U = \{U_i\}$, each piece U_i carrying a type and exponent. We say that a matrix index A has index shape U if they have the same number of pieces of the same type and power, and all pieces are of the same size.

Next, we generalize the input distribution Ω. An example of Ω is the standard normal distribution, $\mathcal{N}(0, 1)$. For an input structure with some underlying Ω, we need to define a way to perform Fourier analysis on the structure similar to how we analyzed the CLIQUE matrix. To do this, we need to use some orthonormal basis for Ω so we can analyze expressions in terms of that basis.

Definition 2.26 (Orthonormal basis for Ω [1, Definition 7.14])**.** Given an input probability distribution Ω and its associated probability function f_Ω, we define $\{h_i\}$

as a set of polynomials found by applying the Gram-Schmidt process to the canonical basis of the polynomial space $\mathbb{R}[x]$. These polynomials have the following properties:

1. $\forall i$, $\mathbb{E}_\Omega[h_i^2(x)] = 1$,
2. $\forall i \neq j$, $\mathbb{E}_\Omega[h_i(x)h_j(x)] = 0$,
3. $\forall i$, the leading coefficient of h_i is positive.

Note that, if $|\text{supp}(\Omega)|$ is finite, there are a finite number of basis polynomials $\{h_i : i \in [|\text{supp}(\Omega)| - 1]\}$.

See Definition 4.2 for a definition of the Gram-Schmidt process as applied to a canonical basis. We also provide the orthonormal basis for the standard normal distribution in Appendix A.1.

Obtaining a basis then lets us *project* expressions such as Equation (2.4) onto that basis, i.e. represent expressions as a linear combination of the basis polynomials $\{h_i\}$. Recall that each edge e in a shape is associated with a random variable, such as χ_e from the Rademacher distribution (i.e. taking on $+1$ or -1 with probability $1/2$) in the case of the CLIQUE matrix. Additionally, recall how in Equation (2.4) there were multiple terms of varying degrees in each summand for entries $\boldsymbol{A}(i, j_1), \boldsymbol{A}(i, j_2)$, and how we represent such entries by edges between vertices that represent such indices. When projecting this expression onto a basis, we will end up with elements of that basis in each summand instead. Thus, when coming up with a shape representation for this polynomial (which we define in Section 5.1), we assign labels to each edge in the shape denoting which element of the basis to use for the random variable associated with that edge. We use these to define the Fourier character of a set of edges below.

Definition 2.27 (Fourier characters [1, Definition 7.17])**.** For a hyperedge e, we assign label l_e to denote which element of the orthonormal basis to use for the random variable associated with e. For a set E of hyperedges, we define the Fourier character

χ_E as

$$\chi_E = \prod_{e \in E} h_{l_e}(x_e),$$

where x_e is the random variable associated with e.

We make the following additional definitions for an input distribution which we use in Theorem 2.35.

Definition 2.28 (Bound function of Ω [1])**.** For a given input distribution Ω, we say that a function $B_\Omega : \mathbb{N} \to \mathbb{R}$ is a *bound function* of Ω if

1. B_Ω is non-decreasing.

2. $\forall j \in \mathbb{N}$, $|\mu_\Omega(j)| \leq B_\Omega(j)^j$.

Definition 2.29. For a given input distribution Ω and a basis $\{h_i\}$, let $h_k(x) = \sum_{j=0}^{k} c_j x^j$ be the polynomial expansion of the degree k polynomial in that basis. We define

$$h_k^+(x) = \sum_{j=0}^{k} |c_j| x^j.$$

We are now ready to generalize ribbons and shapes.

Definition 2.30 (Generalized ribbons [1, Definition 7.18])**.** A generalized ribbon R consists of (possibly intersecting) generalized matrix indices A_R and B_R, a set of additional distinct indices C_R, and a set of subsets of indices $E(R)$. We represent R as a graph with left vertices $V(A_R)$, right vertices $V(B_R)$, middle vertices $V(C_R)$, and hyperedge set $E(R)$. The left and right vertices may intersect, but the middle vertices are distinct. The Fourier character χ_R is defined as $\chi_{E(R)}$.

Definition 2.31 (Generalized shapes [1, Definition 7.19])**.** A shape α consists of generalized index shapes U_α and V_α, a set of additional distinct variables W_α, and a set of subsets of variables $E(\alpha)$. We represent α as a hypergraph with left vertices $V(U_\alpha)$, right vertices $V(V_\alpha)$, middle vertices $V(W_\alpha)$, and hyperedge set $E(\alpha)$ between

these vertices. The left and right vertices may intersect, but the middle vertices are distinct.

Because these shapes can now consist of vertices of different types, we assign weights to these vertices depending on their type, which will be used later in the norm bound.

Definition 2.32 (Weights of vertices)**.** Let $n = \max_{i \in [t]} |N_i|$, the size of the largest ground set. Given a generalized shape α, the weight of a vertex v of type i is defined as $w(v) = \log_n |N_i|$. The weight of any subset of vertices $S \subseteq V(\alpha)$ is defined as the sum of the weights of each vertex in S.

This gives us a way to define some ordering on the types based on the size of their ground set.

Recall that shapes are simply abstracted ribbons with index assignments replaced by free variables. We generalize what it means for a ribbon R to have shape α.

Definition 2.33 (Realizations of shapes)**.** Given some realization $\sigma : U \to \cup_{i \in [t]} N_i$ (i.e. mapping an index shape to elements of all ground sets), we say σ is *type-respecting* if vertices of type i in the index shape are mapped to elements of N_i, the i-th ground set.

For a shape α and ribbon R, we say that R has shape α if there exists a type-respecting realization σ such that $\sigma(\alpha) = R$.

Next, we define the graph matrix of a generalized shape α.

Definition 2.34. For a generalized shape α, we define the graph matrix $\boldsymbol{M}_\alpha$ as

$$\boldsymbol{M}_\alpha(A, B) = \sum_{\substack{\sigma : \sigma \text{ is type-respecting for } \alpha \\ \sigma(U_\alpha) = A,\ \sigma(V_\alpha) = B}} \chi_{\sigma(E(\alpha))},$$

where A and B are generalized matrix indices such that $|A| = |U_\alpha|$ and $|B| = |V_\alpha|$.

Lastly, we state the main theorem of [1] showing probabilistic bounds on the norm of a generalized graph matrix.

Theorem 2.35 (Norm bound on generalized graph matrices, [1, Theorem 8.4])**.** *For a given generalized shape α, let $S_{\min}$ be a minimum weight U_α, V_α-separator, and let W_{iso} be the set of isolated middle vertices in W_α. For each $i \in [t]$, let m_i be the number of vertices of type i not in $U_\alpha \cap V_\alpha$. Then, with probability $1 - \epsilon$,*

$$\|\boldsymbol{M}_\alpha\| \leq 2\left(\prod_{i=1}^{t} m_i^{m_i}\right) \cdot n^{\frac{1}{2}(w(V(\alpha))-w(S_{\min})+w(W_{iso}))} \min_{q\geq 3}\left\{(2q)^{|V(\alpha)\setminus(U_\alpha\cap V_\alpha)|}\left(\prod_{e\in E(\alpha)} h_{l_e}^{+}(B_\Omega(2ql_e))\right)\sqrt[2q]{\frac{n^{w(S_{\min})}}{\epsilon}}\right\}.$$

We refer readers to [1, Section 8.1] for a formal proof of this theorem, the techniques for which are described in Chapter 3.

2.3 Multivariate Distributions

In Section 2.2.2, we made the assumption that the input distribution Ω was univariate, i.e. single-valued. However, in this thesis we seek to expand the graph matrix framework by letting Ω be multivariate. To that end, we provide some definitions concerning multivariate random variables and distributions. Let d be the dimension of the multivariate distribution.

Definition 2.36 (Multivariate moment)**.** For all $\boldsymbol{j} = \langle j_1, j_2, \ldots, j_d\rangle \in \mathbb{N}^d$, the $\boldsymbol{j}$-th moment of Ω is defined as

$$\mu_\Omega(\boldsymbol{j}) = \mu_\Omega(j_1, j_2, \ldots, j_d) = \mathbb{E}_\Omega\left[\prod_{i=1}^{d} x_i^{j_i}\right],$$

where $\boldsymbol{x} = \langle x_1, x_2, \ldots, x_d\rangle \sim \Omega$.

The multivariate moment is a generalization of $\mathbb{E}_\Omega[X^j]$ for a univariate distribution.

We also define the marginal distribution of an individual random variable in a multivariate distribution.

Definition 2.37 (Marginal distribution)**.** Suppose some d-variate distribution Ω with density function $f_\Omega(x_1, \ldots, x_d)$, and let $\boldsymbol{X} = \langle X_1, X_2, \ldots, X_d \rangle \sim \Omega$. The marginal distribution of the random variable X_i is defined as a probability distribution in X_i with density function

$$f_{X_i}(x) = \int f_\Omega(x_1, \ldots, x_d)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_d.$$

As an example of a multivariate distribution, recall the univariate normal distribution $\mathcal{N}(\mu, \sigma^2)$, with probability density function $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$. The standard normal distribution is defined as $\mathcal{N}(0, 1)$. We can generalize these distributions to the multivariate case, which we define below.

Definition 2.38 (Multivariate normal distribution)**.** The d-variate normal distribution $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and $d \times d$ covariance matrix $\boldsymbol{\Sigma} \succeq 0$ is defined with the following probability density function:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right),$$

where $\boldsymbol{x} \in \mathbb{R}^d$ and $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix.

We can define the parameters of a multivariate normal distribution: Suppose $\boldsymbol{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, we define the mean vector $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{x}] = \begin{bmatrix} \mathbb{E}[x_1] & \mathbb{E}[x_2] & \ldots & \mathbb{E}[x_d] \end{bmatrix}^\top,$$

and each entry of the covariance matrix $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma}(i,j) = \mathbf{\Sigma}(j,i) = \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)] = \mathrm{Cov}[x_i, x_j].$$

Chapter 3

Techniques for Proving Bounds of Graph Matrices

In this chapter, we discuss techniques used to bound the spectral norm of a graph matrix. We begin with the trace power method in Section 3.1 as a means to bound the spectral norm of a general matrix. Section 3.2 then introduces constraint graphs, a concept used to apply the trace power method to the spectral norm of a graph matrix based on its shape. We then discuss how to handle the input distribution in the norm bound of a graph matrix in Section 3.3.

3.1 The Trace Power Method

As explained by van Handel in [16], the spectral norm as a function $\boldsymbol{M} \mapsto \|\boldsymbol{M}\|$ is nonlinear, since it is a complicated function in terms of entries $\boldsymbol{M}(i,j)$. Recall the definition of the spectral norm given in Definition 2.2. We can rewrite this norm as

$$\|\boldsymbol{M}\| = \sigma_{\max}(\boldsymbol{M}) = \sqrt{\lambda_{\max}(\boldsymbol{M}\boldsymbol{M}^\top)},$$

where $\sigma_{\max}(\boldsymbol{M})$ is the largest singular value of $\boldsymbol{M}$, and $\lambda_{\max}(\boldsymbol{M})$ is the largest eigenvalue of $\boldsymbol{M}$. This is indeed a complicated function in terms of entries of a matrix.

Now consider the trace of a matrix, which by contrast is very much linear in the entries of $\boldsymbol{M}$, as seen in Definition 2.3. We have the linear algebra identity that $\mathrm{Tr}(\boldsymbol{M}) = \sum_i \lambda_i(\boldsymbol{M})$, the sum of $\boldsymbol{M}$'s eigenvalues. More generally, we can also say that

$$\mathrm{Tr}(\boldsymbol{M}^q) = \sum_{i=1}^{n} \lambda_i^q.$$

We then have the following lemma, derived from the above, relating $\|\boldsymbol{M}\|$ to $\mathrm{Tr}\big(\boldsymbol{M}\boldsymbol{M}^\top\big)$.

Lemma 3.1 ([1], [16])**.** *For some matrix $\boldsymbol{M}$ and $q \in \mathbb{N}$, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues*

of $\boldsymbol{M}\boldsymbol{M}^\top$*. Then,*

$$\|\boldsymbol{M}\|^{2q} = \lambda_{\max}(\boldsymbol{M}\boldsymbol{M}^\top)^q = \left\|\boldsymbol{M}\boldsymbol{M}^\top\right\|^q \leq \sum_{i=1}^{n} \lambda_i^q = \mathrm{Tr}\big((\boldsymbol{M}\boldsymbol{M}^\top)^q\big).$$

This lemma lets us bound the spectral norm with the trace, which is a simpler expression to analyze. This is known as the *trace power method* or the moment method as described in other texts.

Thus, if we can bound the expected trace power of a random matrix for each q with some function, it follows that we can bound the spectral norm with those bounds. Ahn et al. give the following variant of the trace power method using this reasoning.

Lemma 3.2 (The trace power method [1])**.** *Suppose* $\boldsymbol{M}$ *is a random matrix, and that we have bounds* $\{B(2q) \mid q \in \mathbb{N}\}$ *such that* $\forall q \in \mathbb{N}$, $\mathbb{E}\big[\mathrm{Tr}\big((\boldsymbol{M}\boldsymbol{M}^\top)^q\big)\big] \leq B(2q)$*. Then, for all* $\epsilon > 0$*,*

$$\Pr\left[\|\boldsymbol{M}\| > \min_{q\in\mathbb{N}}\left\{\sqrt[2q]{\frac{B(2q)}{\epsilon}}\right\}\right] < \epsilon.$$

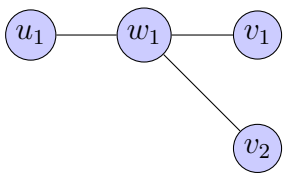
The proof for this lemma uses Markov's inequality: $\Pr\Big(x \geq \frac{\mathbb{E}[x]}{c}\Big) \leq c$.

Thus, bounding the spectral norm of a graph matrix $\boldsymbol{M}_\alpha$ involves finding a bound on the expected trace power of $\boldsymbol{M}_\alpha\boldsymbol{M}_\alpha^\top$ for some $q \in \mathbb{N}$.

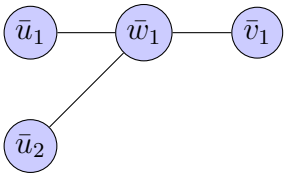
3.2 Constraint Graphs

Ahn et al. then derive a bound using the trace power method by encoding the terms of the expectation in Lemma 3.2 for some shape α using a concept called a constraint graph. To introduce these graphs, we first provide a definition of a multi-graph based on a shape α.

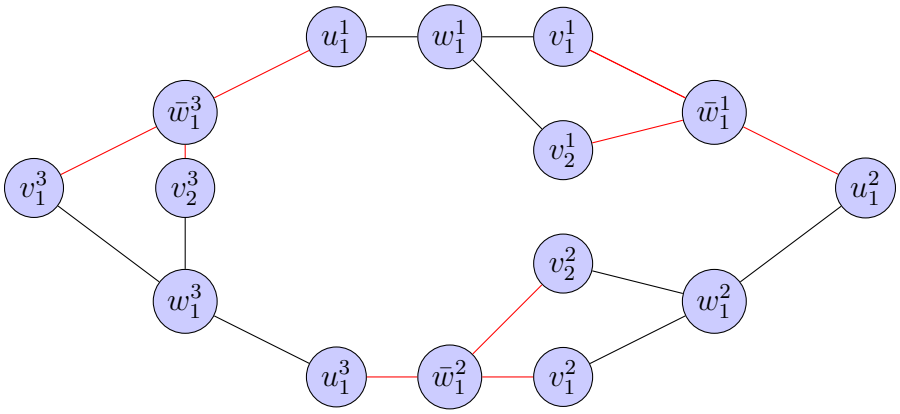
Definition 3.3. Given a shape α, we define its "transpose" shape $\alpha^\top$ by taking α and swapping its index shapes U_α and V_α. Next, given a $q \in \mathbb{N}$, we define the multi-graph $H(\alpha, 2q)$ by doing the following:



(a) Shape α



(b) $\alpha^\top$: Transpose of shape α



(c) $H(\alpha, 2q)$ for $q = 3$. Edges in copies of $\alpha^\top$ are colored red.

Figure 3.1: Example shape α, its transpose $\alpha^\top$, and $H(\alpha, 2q)$ for $q = 3$.

1. Stitch q copies of α and $\alpha^\top$ together by their left/right vertices, forming a cyclic structure. That is, for copies $\alpha_1, \alpha_1^\top, \ldots \alpha_q, \alpha_q^\top$, set $V_{\alpha_i} = U_{\alpha_i^\top}$ and $V_{\alpha_i^\top} = U_{\alpha_{i+1}}$, where i is modulo q.

2. Keep any overlapping edges as multi-edges.

We provide an example shape α, its transpose $\alpha^\top$, and $H(\alpha, 2q)$ for a fixed q in Figure 3.1, using the following notation:

- Each unique vertex in U_{α_i} and V_{α_i} is labeled u_j^i and v_j^i, respectively.
- Each vertex in W_{α_i} is labeled w_j^i. The same copy of the vertex in the transpose shape is labeled $\bar{w}_j^i$.

Having defined these graphs, the reader might notice the symmetry between the $(\boldsymbol{M}_\alpha \boldsymbol{M}_\alpha^\top)^q$ terms and the $H(\alpha, 2q)$ graphs. In particular, expanding the former expression yields copies of a matrix and its transpose strung together q times, while

graphically the latter expression shows its associated shape and *its* transpose strung together q times. Indeed, this gives us one of the tools needed to map terms in $\mathbb{E}\big[(\boldsymbol{M}_\alpha \boldsymbol{M}_\alpha^\top)^q\big]$ to different graphs that build upon $H(\alpha, 2q)$. To that end, we consider mappings on the vertex set of $H(\alpha, 2q)$, which we write as $V(\alpha, 2q)$, that assigns indices from the ground sets $\bigcup_{i\in[t]} N_i$ to those vertices. We restrict these mappings to only those such that no two vertices in the same copy of α (or $\alpha^\top$) receive the same index. Additionally, because vertices can be of different types, mappings need to respect those types. The following definitions and proposition formalize these restrictions.

Definition 3.4 (Type-respecting mappings [1])**.** We say that a mapping $\phi : V \to \bigcup_i N_i$ is type-respecting if and only if, for each $i \in [t]$, vertices of type i are mapped to the ground set N_i.

Definition 3.5 (Piecewise injectivity [1])**.** A mapping $\phi : V(\alpha, 2q) \to \bigcup_{i\in[t]} N_i$ is piecewise injective if ϕ is injective on each copy of α_i and $\alpha_i^\top$ for all $i \in [q]$. That is, no two vertices in the same copy of a shape (or its transpose) receive the same mapping.

Ahn et al. rephrase the expected trace power as a summation over such mappings using the $H(\alpha, 2q)$ graphs.

Proposition 3.6 ([1])**.** *For all shapes α and $q \in \mathbb{N}$,*

$$\mathbb{E}\big[\mathrm{Tr}\big((\boldsymbol{M}_\alpha \boldsymbol{M}_\alpha^\top)^q\big)\big] = \sum_{\substack{\phi: V(\alpha,2q)\to\bigcup_i N_i \\ \phi \text{ is piecewise injective} \\ \phi \text{ is type-respecting}}} \mathbb{E}\big[\chi_{\phi(E(\alpha,2q))}\big],$$

where $E(\alpha, 2q)$ denotes the edge set of $H(\alpha, 2q)$.

We note that the expectation of the terms of both sides are over the input distribution Ω.

Ahn et al. then observe that the expectations on the right-hand side of the above proposition only depend on sets of vertex pairs in $V(\alpha, 2q)$ (i.e. the vertex set of

$H(\alpha, 2q)$) that receive the same mapping from each ϕ. We therefore define constraint graphs for a given mapping on a set of vertices, which capture this concept by linking such vertices together by constraint edges.

Definition 3.7 (Constraint graphs [1])**.** Given a set of vertices V and a mapping $\phi : V \to \bigcup_{i\in[t]} N_i$, the constraint graph $C(\phi)$ on V is constructed as follows:

1. Give $C(\phi)$ the vertex set V.
2. For each pair of vertices $u, v \in V$ that receive the same mapping $\phi(u) = \phi(v)$, add a *constraint edge* $\{u, v\}$ to $C(\phi)$.
3. Remove one constraint edge from any cycles to remove redundancies.

We say that two constraint graphs on V are equivalent, written $C \equiv C'$, if for each pair $u, v \in V$, there exists a u, v-path in C if and only if there exists a u, v-path in C'.

Next, we provide some definitions concerning constraint graphs specifically on $H(\alpha, 2q)$.

Definition 3.8. Let $\mathcal{C}_{(\alpha,2q)} = \{C(\phi) \mid \phi : V(\alpha, 2q) \to \bigcup_i N_i\}$ be the class of all possible constraint graphs on $H(\alpha, 2q)$ with some type-respecting, piecewise injective mapping ϕ. We define the following:

1. $N(C)$ to be the size of the equivalence class on C under the set of all possible such mappings ϕ.
2. $\mathrm{val}(C) = \mathbb{E}_\Omega\big[\chi_{\phi(E(\alpha,2q))}\big]$ for some ϕ such that $C(\phi) \equiv C$.

ϕ in the above definitions are assumed to be type-respecting and piecewise injective.

Then we simplify Proposition 3.6 as follows:

Proposition 3.9. *For all shapes α and $q \in \mathbb{N}$,*

$$\mathbb{E}\big[\mathrm{Tr}\big((\boldsymbol{M}_\alpha \boldsymbol{M}_\alpha^\top)^q\big)\big] = \sum_{C\in\mathcal{C}_{(\alpha,2q)}} N(C)\,\mathrm{val}(C).$$

Ahn et al. then analyze the expected trace power using combinatorial arguments to bound $N(C)$ and $\text{val}(C)$. After coming up with these bounds on the expected trace power, Ahn et al. use another technique called the *vertex partitioning lemma*, which is a way to break up a more complicated graph matrix $\boldsymbol{M}_\alpha$ into well-behaved pieces $\boldsymbol{M}_{\alpha,P}$ by partitioning the ground set of indices. These pieces are bounded individually using some bounding function, then recombined to form a norm bound on the graph matrix itself. As this technique does not change when relaxing the univariate assumption, we refer readers to [1, Section 5] for more details.

3.3 Handling the Input Distribution

The analysis of these constraint graphs is done for any general (univariate) input distribution Ω, over which expectations throughout the previous section are computed. Bounding $\text{val}(C)$ however requires bounding moments of the input distribution. For completeness, we re-state these definitions and refer readers to [1, Section 7.4] for the proofs of the given lemmas. We modify such lemmas and give our own proofs under the multivariate setting later in Chapter 4.

Recall the definition of basis polynomials in Definition 2.26 and their absolute value counterparts in Definition 2.29. The latter set of polynomials have the following bound on the former:

Lemma 3.10. *For all $k, r \in \mathbb{N}$, $\mathbb{E}_\Omega[h_k(x)^r] \leq h_k^+(B_\Omega(kr))^r$.*

The h^+ polynomials are then used to finally bound $\text{val}(C)$ for some *well-behaved* constraint graph $C \in \mathcal{C}_{(\alpha,2q)}$. We define this well-behaved property below, then state the bound on $\text{val}(C)$.

Definition 3.11 (Well-behaved constraint graph)**.** We say that a constraint graph $C(\phi) \in \mathcal{C}_{(\alpha,2q)}$ is *well-behaved* if, for every constraint edge $(u, v) \in E(C(\phi))$, u and v

are both copies of the same vertex in α across different copies of α (or $\alpha^\top$). That is, vertices that receive the same mapping under ϕ are copies of the same vertex.

Lemma 3.12. *For a generalized shape α with edge labels $\{l_e\}_{e\in E(\alpha)}$ (as in Definition 2.27) and constraint graph $C \in \mathcal{C}_{(\alpha,2q)}$,*

$$\mathrm{val}(C) \leq \prod_{e\in E(\alpha)} h^+_{l_e}(B_\Omega(2ql_e))^{2q}.$$

Proving the norm bound then just requires bounding the expected trace power terms along with the vertex partitioning lemma mentioned in Section 3.2. The bound on the expected trace power terms is given by the following lemma:

Lemma 3.13 ([1, Lemma 8.7])**.** *Let $S_{\min}$ be a minimum weight U_α, V_α-separator. Then, for all vertex partitions P and $q \in \mathbb{N}$,*

$$\mathbb{E}\big[\mathrm{Tr}\big((\boldsymbol{M}_{\alpha,P}\boldsymbol{M}_{\alpha,P}^\top)^q\big)\big] \leq \left(\prod_{e\in E(\alpha)} h^+_{l_e}(B_\Omega(2ql_e))^{2q}\right)(2q)^{2q|V(\alpha)|} n^{q(w(\alpha)-w(S_{\min}))+w(S_{\min})}.$$

The proof for this lemma involves combinatorial arguments that examine the weight of constraint edges (i.e. the weight of one of their endpoints, well-defined under the well-behaved constraint) and number of constraint graphs in $\mathcal{C}_{(\alpha,2q)}$.

Chapter 4

Graph Matrices in the Multivariate Setting

In this chapter we extend the graph matrix framework to accomodate multivariate input distributions. From here on, we assume the input distribution Ω has dimension d. That is, values sampled from Ω are vectors of size d. For example, Ω could be the standard d-variate normal distribution $\mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$, where $\mathbf{0}_d$ is the zero vector of size d, and $\boldsymbol{I}_d$ is the $d \times d$ identity matrix.

Previous sections as well as theorems throughout [1] assumed a univariate input distribution Ω, so we must retrofit some of these theorems to the multivariate case. We begin by redefining an orthonormal basis for Ω in Section 4.1. Section 4.2 will then modify the techniques introduced in Chapter 3 to fit the multivariate case.

4.1 Orthonormal Bases

Recall the definition of an orthonormal basis in Definition 2.26. An orthonormal basis in the univariate case is taken by performing the Gram-Schmidt process on the canonical basis of the polynomial vector space $\mathbb{F}[x]$, where $\mathbb{F}$ is usually $\mathbb{R}$. We generalize this vector space to the d-variate case $\mathbb{F}[\boldsymbol{x}]$ in the following definition.

Definition 4.1 (Multivariate polynomial space)**.** We write a monomial in d variables as

$$\boldsymbol{x}^{\boldsymbol{a}} = x_1^{a_1} x_2^{a_2} \cdots x_d^{a_d},$$

where $\boldsymbol{a} = \langle a_1, \ldots, a_d \rangle$ is a vector of exponents we call the *degree* of the monomial. We also call $\|\boldsymbol{a}\|_1$ the *degree sum* of the monomial. The vector space of polynomials $\mathbb{F}[\boldsymbol{x}_d]$ is the set of all expressions of the form

$$\sum_{\boldsymbol{a} \in \mathbb{N}^d} c_{\boldsymbol{a}} \boldsymbol{x}^{\boldsymbol{a}},$$

where $c_{\boldsymbol{a}} = c_{a_1,\ldots,a_d}$ uses the multi-index notation to denote the coefficient of monomial

$\boldsymbol{x}^{\boldsymbol{a}}$. Further, we define the *canonical basis* of $\mathbb{F}[\boldsymbol{x}_d]$ as

$$\{1, x_1, \ldots, x_d, x_1^2, \ldots, x_d^2, x_1x_2, x_1x_3, \ldots, x_{d-1}x_d, \ldots\}$$

or, more succinctly, $\bigcup_{k=0}^{\infty} \{\boldsymbol{x}^{\boldsymbol{a}} \mid \|\boldsymbol{a}\|_1 = k\}$, the union of all monomials with degree sum k and unit coefficient for $k = \{0, 1, 2, \ldots\}$.

Unless otherwise noted, we use $\mathbb{R}$ as our field.

Next, we define the Gram-Schmidt process as used in deriving an orthonormal basis. All distributions Ω have a probability mass or density function associated with them, which we will write as $f_\Omega(\boldsymbol{x}_d)$. The Gram-Schmidt orthogonalization process with respect to this density function is as follows.

Definition 4.2 (Gram-Schmidt process for multivariate Ω)**.** Consider the polynomial vector space $\mathbb{F}[\boldsymbol{x}_d]$ equipped with the following inner product:

$$\langle p_i, p_j \rangle = \mathbb{E}_\Omega[p_i(\boldsymbol{x})p_j(\boldsymbol{x})] = \int_{\text{supp}(\Omega)} p_i(\boldsymbol{x})p_j(\boldsymbol{x})f_\Omega(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,, \tag{4.1}$$

where $p_i(\boldsymbol{x})$ and $p_j(\boldsymbol{x})$ are polynomials in $\boldsymbol{x}_d$ (shortened to p_i and p_j for clarity). Define the projection operator proj as

$$\text{proj}_{p_i}(p_j) = \frac{\langle p_i, p_j \rangle}{\langle p_i, p_i \rangle} p_i.$$

Let $\{p_0, p_1, p_2, \ldots, p_n\}$ be a basis for $\mathbb{F}[\boldsymbol{x}_d]$, possibly with infinitely many elements in the case where Ω has infinite support. The Gram-Schmidt process for obtaining an

orthonormal basis is defined as first calculating the polynomials

$$
\begin{aligned}
q_0 &= p_0 \\
q_1 &= p_1 - \mathrm{proj}_{q_0}(p_1) \\
q_2 &= p_2 - \mathrm{proj}_{q_0}(p_2) - \mathrm{proj}_{q_1}(p_2) \\
&\vdots \\
q_k &= p_k - \sum_{j=0}^{k-1} \mathrm{proj}_{q_j}(p_k) \\
&\vdots
\end{aligned}
$$

for $k \in [|\mathrm{supp}(\Omega)|]$. We then normalize each q_k to get polynomials $h_k = \frac{q_k}{\sqrt{\langle q_k, q_k \rangle}}$ so that $\langle h_k, h_k \rangle = 1$.

Finally, we are ready to define an orthonormal basis for Ω using this process.

Definition 4.3 (Multivariate orthonormal basis for Ω)**.** Given an input distribution Ω in d variables and its associated probability function f_Ω, we define $\{h_{\boldsymbol{i}}\}$ as a set of polynomials bound by applying the Gram-Schmidt process in Definition 4.2 to the canonical basis of the polynomial space $\mathbb{F}[\boldsymbol{x}_d]$. The set of basis polynomials is indexed by a vector $\boldsymbol{i} \in \mathbb{N}^d$. These polynomials have the following properties:

1. (Normality). $\forall \boldsymbol{i}$, $\mathbb{E}_\Omega[h_{\boldsymbol{i}}^2(\boldsymbol{x})] = 1$.

2. (Orthogonality). $\forall \boldsymbol{i}, \boldsymbol{j}$ such that $\boldsymbol{i} \neq \boldsymbol{j}$, $\mathbb{E}_\Omega[h_{\boldsymbol{i}}(\boldsymbol{x}) h_{\boldsymbol{j}}(\boldsymbol{x})] = 0$.

3. (Positive leading coefficient). $\forall \boldsymbol{i}$, the monomial $\boldsymbol{x}^{\boldsymbol{i}}$ of $h_{\boldsymbol{i}}$ has a positive coefficient.

Note that the first two properties can be restated as: the polynomials form an orthonormal basis of the polynomial vector space equipped with the inner product defined in Equation (4.1).

A careful reader might notice that the indices of the basis polynomials in Definition 4.2 are scalars while those in Definition 4.3 are vectors. For our purposes, since

we will use the canonical basis to obtain an orthonormal basis, we associate a basis polynomial $h_{\boldsymbol{i}}$ with the polynomial obtained using the canonical basis monomial $\boldsymbol{x}^{\boldsymbol{i}}$ during the Gram-Schmidt process.

With these polynomials, we can then modify the techniques described in Chapter 3 to apply the graph matrix framework in the multivariate case.

4.2 Modified Techniques

Recall that bounding the norm of a graph matrix involves bounding the expected trace power of that matrix as explained in Section 3.1, and that this expression is bounded using properties of constraint graphs, specifically the functions $N(C)$ and $\text{val}(C)$ for a class of constraint graphs $\mathcal{C}_{(\alpha,2q)}$. In [1], this involved bounding $\text{val}(C)$ using the orthonormal bases of the input distribution Ω, which we showed in Section 3.3. Because Ω is now a d-variate distribution, we must modify the lemmas given in Section 3.3 that were used to bound $\text{val}(C)$.

First, we make the following definitions for bounding functions for a distribution Ω.

Definition 4.4 (Bound function of Ω)**.** For any input distribution Ω, a bound function $B_\Omega : \mathbb{N}^d \to \mathbb{R}$ for Ω satisfies the following

1. It is component-wise nondecreasing.
2. $\forall j_1, j_2, \ldots, j_d \in \mathbb{N}$, $|\mu_\Omega(j_1, j_2, \ldots, j_d)| \leq B_\Omega(j_1, j_2, \ldots)^{\sum_{i=1}^d j_i}$.

Note that this function is not unique for each Ω, but finding a bound function that produces tighter bounds on the moment of the distribution will produce a better refined bound on $\text{val}(C)$.

Next, to bound each polynomial in the orthonormal basis $\{h_{\boldsymbol{k}}\}$, we define a new set of polynomials that are a "flattening" and normalization of the basis polynomials, similar to Definition 2.29:

Definition 4.5. For some vector $\boldsymbol{k} = \langle k_1, k_2, \ldots, k_d \rangle$, let $h_{\boldsymbol{k}}(x_1, x_2, \ldots, x_d)$ be the polynomial taken by applying the Gram-Schmidt process for Ω to the element of the canonical basis with degree k_1 in x_1, degree k_2 in x_2, and so on. Consider the polynomial expansion of $h_{\boldsymbol{k}}$ as

$$h_{\boldsymbol{k}}(\boldsymbol{x}) = \sum_{\boldsymbol{j} \in \mathbb{N}^d} c_{j_1, j_2, \ldots, j_d} \boldsymbol{x}^{\boldsymbol{j}}.$$

We define $h^+_{\boldsymbol{k}}(x)$ to be

$$h^+_{\boldsymbol{k}}(x) = \sum_{j=0}^{\|\boldsymbol{k}\|_1} \left(\sum_{\substack{j_1, \ldots, j_d \\ j_1 + \cdots j_d = j}} |c_{j_1, \ldots, j_d}| \right) x^j \tag{4.2}$$

That is, we group all monomials with a given degree sum j and sum the absolute value of their coefficients, giving us the coefficient for the monomial x^j. This is done for all j from 0 to $\|\boldsymbol{k}\|_1$, since no monomials will have a higher degree sum due to the way we index each $h_{\boldsymbol{k}}$.

With these, we can state the following theorem.

Theorem 4.6. *For all $\boldsymbol{k} \in \mathbb{N}^d$, $r \in \mathbb{N}$,*

$$\mathbb{E}_\Omega[h_{\boldsymbol{k}}(\boldsymbol{x})^r] \leq h^+_{\boldsymbol{k}}(B_\Omega(r\boldsymbol{k}))^r,$$

where $r\boldsymbol{k} = \langle rk_1, rk_2, \ldots, rk_d \rangle$.

Proof. Consider the polynomial expansion of $h_{\boldsymbol{k}}(\boldsymbol{x})^r$ and $h^+_{\boldsymbol{k}}(x)^r$. That is, write

$$h_{\boldsymbol{k}}(\boldsymbol{x})^r = \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} a_{j_1, \ldots, j_d} x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d}$$

and

$$h^+_{\boldsymbol{k}}(x)^r = \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} b_{j_1,\ldots,j_d} x^j,$$

where $j = \sum_{i=1}^{d} j_i$. Then, we can write the expectation as

$$\begin{aligned}
\mathbb{E}_\Omega[h_{\boldsymbol{k}}(\boldsymbol{x})^r] &= \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} a_{j_1,\ldots,j_d} \, \mathbb{E}_\Omega \left[x_1^{j_1} x_2^{j_2} \cdots x_d^{j_d} \right] \\
&= \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} a_{j_1,\ldots,j_d} \mu_\Omega(j_1, \ldots, j_d) \\
&\leq \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} |a_{j_1,\ldots,j_d}| \mu_\Omega(j_1, \ldots, j_d)
\end{aligned}$$

By Definition 4.4, B_Ω bounds the moment of Ω, so

$$\leq \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} |a_{j_1,\ldots,j_d}| B_\Omega(j_1, \ldots, j_d)^j$$

And since B_Ω is component-wise nondecreasing,

$$\leq \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} |a_{j_1,\ldots,j_d}| B_\Omega(k_1 r, \ldots, k_d r)^j.$$

Before we proceed any further, we make the following proposition.

Proposition 4.7. *Using the polynomial expansions above, each coefficient $a_{\boldsymbol{j}}$ in $h_{\boldsymbol{k}}(\boldsymbol{x})^r$ is matched by a coefficient $b_{\boldsymbol{j}}$ of equal or greater value in $h^+_{\boldsymbol{k}}(x)^r$.*

Proof. Fix the vector $\boldsymbol{j}$. Each $a_{j_1,\ldots,j_d} = a_{\boldsymbol{j}}$ is a product of multiple coefficients $\{c_{\boldsymbol{l}}\}$ and a multinomial coefficient $\binom{\|\boldsymbol{k}\|_1 r}{j_1,\ldots,j_d}$. Similarly, each $b_{\boldsymbol{j}}$ is a product of the same coefficients $\{|c_{\boldsymbol{l}}| \geq c_{\boldsymbol{l}}\}$ and the same multinomial coefficient $\binom{\|\boldsymbol{k}\|_1 r}{j_1,\ldots,j_d}$.

To show this, consider how monomials in $h_{\boldsymbol{k}}(\boldsymbol{x})$ are grouped by degree sum in $h^+_{\boldsymbol{k}}(x)$ as in Equation (4.2), so that each monomial x^l in the latter has coefficients equal to the sum of monomials $\boldsymbol{x}^{\boldsymbol{l}}$ in the former where $\|\boldsymbol{l}\|_1 = l$. We can separate this

sum of coefficients out so Equation (4.2) becomes

$$h_{\boldsymbol{k}}^{+}(x) = \sum_{l=0}^{\|\boldsymbol{k}\|_1} \left(\sum_{\substack{l_1,\ldots,l_d \\ l_1+\cdots+l_d=l}} |c_{l_1,\ldots,l_d}| x^l \right).$$

(Note how the x_l term is distributed to the coefficients in the inner sum.) Each of these monomials thus corresponds to a monomial $c_{l_1,\ldots,l_d}\boldsymbol{x}^{\boldsymbol{l}}$ in $h_{\boldsymbol{k}}(\boldsymbol{x})$.

So when expanding $h_{\boldsymbol{k}}^{+}(x)^r$, to calculate the coefficient $b_{\boldsymbol{j}}$, we would multiply the monomials in $h_{\boldsymbol{k}}^{+}(x)$ that correspond to those in $h_{\boldsymbol{k}}(\boldsymbol{x})$, the only difference being the coefficients in the former are the absolute value of those in the latter. The resulting product of coefficients in the former, $b_{\boldsymbol{j}}$, is therefore the same as the absolute value of that in the latter, $a_{\boldsymbol{j}}$. In other words, $b_{\boldsymbol{j}} = |a_{\boldsymbol{j}}|$. We repeat this for all $\boldsymbol{j}$ as necessary, proving the proposition. □

Returning to the proof for Theorem 4.6, using the above proposition, we can then bound the left-hand side of the theorem's inequality by

$$\begin{aligned}
\mathbb{E}_\Omega[h_{\boldsymbol{k}}(\boldsymbol{x})^r] &\leq \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} |a_{j_1,\ldots,j_d}| B_\Omega(k_1 r, \ldots, k_d r)^j \\
&\leq \sum_{j_1=0}^{k_1 r} \sum_{j_2=0}^{k_2 r} \cdots \sum_{j_d=0}^{k_d r} b_{j_1,\ldots,j_d} B_\Omega(k_1 r, \ldots, k_d r)^j.
\end{aligned}$$

But this is just the polynomial expansion of $h_{\boldsymbol{k}}^{+}(x)$ for $x = B_\Omega(r\boldsymbol{k})$, so

$$= h_{\boldsymbol{k}}^{+}(B_\Omega(r\boldsymbol{k}))^r,$$

which concludes the proof. □

With this theorem, we can state our bound on $\mathrm{val}(C)$.

Theorem 4.8. *For a shape α containing hyperedges labeled with $\{\boldsymbol{l}_e\}_{e\in E(\alpha)}$, let*

$C \in \mathcal{C}_{(\alpha,2q)}$ be a well-behaved constraint graph. Then

$$\mathrm{val}(C) \leq \prod_{e \in E(\alpha)} h^+_{\boldsymbol{l}_e}(B_\Omega(2q\boldsymbol{l}_e))^{2q}.$$

Proof. Let ϕ be the mapping associated with $C \in \mathcal{C}_{(\alpha,2q)}$. Recall from Definition 3.8 that $\mathrm{val}(C) = \mathbb{E}_\Omega\left[\chi_{\phi(E(\alpha,2q))}\right]$. In the case where $\mathrm{val}(C) = 0$, this bound is trivial.

Consider $\mathrm{val}(C) \neq 0$. We can say that

$$\begin{aligned}\mathrm{val}(C) &= \mathbb{E}_\Omega\left[\chi_{\phi(E(\alpha,2q))}\right] \\ &= \mathbb{E}_\Omega\left[\prod_{e \in \phi(E(\alpha,2q))} h_{\boldsymbol{l}_e}(\boldsymbol{x})\right] \qquad \text{(by definition of } \chi_E\text{)}\end{aligned}$$

Partition $H(\alpha, 2q)$ into $\{C_e\}_{e \in E(\alpha)}$, where each C_e is the subset consisting of only the $2q$ copies of e. Since C is well-behaved, two hyperedges $e_1, e_2 \in E(\alpha, 2q)$ correspond to the same hyperedge e under the mapping $\phi(E(\alpha, 2q))$ if and only if e_1 and e_2 are copies of each other, and are thus in the same set C_e. The random variable corresponding to each hyperedge is independently drawn from the input distribution Ω, so the expectation becomes

$$\mathrm{val}(C) = \prod_{e \in E(\alpha)} \mathbb{E}_\Omega\left[\chi_{\phi(C_e)}\right]. \tag{4.3}$$

Now suppose each of the p unique hyperedges $e_i \in C_e$ appears r_i times for $i \in [p]$. Then, each term in the above product becomes

$$\mathbb{E}_\Omega\left[\chi_{\phi(C_e)}\right] = \prod_{i=1}^{p} \mathbb{E}_\Omega\left[h_{\boldsymbol{l}_e}(\boldsymbol{x})^{r_i}\right],$$

and using Theorem 4.6,

$$\leq \prod_{i=1}^{p} h^+_{\boldsymbol{l}_e}(B_\Omega(r_i \boldsymbol{l}_e))^{r_i}.$$

Since $r_i \leq 2q$, and by Definition 4.4 B_Ω is component-wise nondecreasing,

$$\begin{aligned} &\leq \prod_{i=1}^{p} h^+_{\boldsymbol{l}_e}(B_\Omega(2q\boldsymbol{l}_e))^{r_i} \\ &\leq h^+_{\boldsymbol{l}_e}(B_\Omega(2q\boldsymbol{l}_e))^{2q} \qquad (\textstyle\sum_{i=1}^{p} r_i = 2q) \end{aligned}$$

Returning to Equation (4.3), we can then apply this bound to each factor in the product (i.e. all edges $e \in E(\alpha)$), which gives us

$$\mathrm{val}(C) = \prod_{e \in E(\alpha)} \mathbb{E}_\Omega\big[\chi_{\phi(C_e)}(X)\big] \leq \prod_{e \in E(\alpha)} h^+_{\boldsymbol{l}_e}(B_\Omega(2q\boldsymbol{l}_e))^{2q}.$$

□

Because the bound mostly stays the same with a few adjustments to account for the multivariate setting, the bound on the expected trace power of a graph matrix piece is mostly the same as well:

Lemma 4.9. *Let $S_{\min}$ be a minimum weight U_α, V_α-separator. Then, for all vertex partitions P and $q \in \mathbb{N}$,*

$$\mathbb{E}\big[\mathrm{Tr}\big((\boldsymbol{M}_{\alpha,P}\boldsymbol{M}_{\alpha,P}^\top)^q\big)\big] \leq \left(\prod_{e \in E(\alpha)} h^+_{\boldsymbol{l}_e}(B_\Omega(2q\boldsymbol{l}_e))^{2q}\right)(2q)^{2q|V(\alpha)|} n^{q(w(\alpha)-w(S_{\min}))+w(S_{\min})}.$$

The most notable change to the bound is that the labels of the shape's edges $e \in E(\alpha)$ are now vectors $\boldsymbol{l}_e$ of length d rather than scalars, since basis polynomials for Ω are now in d variables. As a result, the main theorem bounding the norm of a graph matrix $\boldsymbol{M}_\alpha$ in Theorem 2.35 also is unchanged aside from the same change to

the labels. For completeness, we state the modified theorem here.

Theorem 4.10 (Norm bound on graph matrices for multivariate Ω)**.** *Given a generalized shape α, let $S_{\min}$ be a minimum weight U_α, V_α-separator, and let W_{iso} be the set of isolated middle vertices in W_α. For each $i \in [t]$, let m_i be the number of vertices of type i not in $U_\alpha \cap V_\alpha$, and let $n = \max_{i\in[t]} |N_i|$, the size of the largest ground set. Then, with probability $1 - \epsilon$,*

$$\|\boldsymbol{M}_\alpha\| \leq 2\left(\prod_{i=1}^{t} m_i^{m_i}\right) \cdot n^{\frac{1}{2}(w(V(\alpha))-w(S_{\min})+w(W_{iso}))} \min_{q\geq 3}\left\{(2q)^{|V(\alpha)\setminus(U_\alpha\cap V_\alpha)|}\left(\prod_{e\in E(\alpha)} h^+_{\boldsymbol{l}_e}(B_\Omega(2q\boldsymbol{l}_e))\right)\sqrt[2q]{\frac{n^{w(S_{\min})}}{\epsilon}}\right\}.$$

Again, we refer readers to the proof in [1, Section 8.1] as the proof is largely unchanged. As a brief summary, proving Theorem 4.10 roughly involves

1. Bounding $N_P(C)$ and $\mathrm{val}(C)$ for a well-behaved constraint graph from the class $\mathcal{C}_{(\alpha,2q)}$,
2. Using these bounds and Lemma 3.10 to bound the expected trace power of a partitioned graph matrix $\boldsymbol{M}_{\alpha,P}$, and
3. Using the bounds over all partitions P of the ground sets with the vertex partitioning lemma ([1, Lemma 5.14]) to derive a bound on the spectral norm of a graph matrix $\boldsymbol{M}_\alpha$.

Ahn et al. also remark that a nearly optimal minimum value for q is often $q = 3\left\lceil \frac{\log\left(n^{w(S_{\min})}/\epsilon\right)}{3|V(\alpha)\setminus(U_\alpha\cap V_\alpha)|}\right\rceil$ (the factor 3 is included to ensure $q \geq 3$).

Chapter 5

Applications using the Multivariate Setting

This chapter discusses the various applications that we can analyze with graph matrices under a multivariate distribution.

5.1 Representing Polynomials as a Shape

One of the primary motivations for developing graph matrices was analyzing the sum-of-squares hierarchy, used to solve some special optimization problems. Some proofs for these problems involved proving an upper bound on the maximum value of some polynomial over the unit sphere. Ahn et al. show that, while the proofs for some of these bounds involved complex and technical reasoning, reproducing the same bounds using graph matrices (within a polylogarithmic factor) is fairly straightforward and mechanical [1, Section 9]. For a given polynomial, we first find a *shape representation* of this polynomial, then apply the theorem bounding the norm of the graph matrix of this shape. Bounding this norm will end up bounding the norm of the polynomial over the unit sphere.

These polynomials involve coefficients whose factors are drawn from a univariate distribution, usually $\mathcal{N}(0,1)$. We expand this by letting several of these factors be grouped into samples from a multivariate distribution instead.

For completeness, we first introduce what it means to have a shape representation of a polynomial. We consider only the set of homogeneous polynomials of even degree, defined below.

Definition 5.1 (Homogeneous polynomial)**.** We say that a polynomial is *homogeneous* if all its terms have the same degree sum r. Further, we say that the polynomial is a homogeneous polynomial of degree r. For example, $x^4 + x^2y^2 + z^3x$ is a homogeneous polynomial of degree 4. On the other hand, $x^4 + xy^2$ is not a homogeneous polynomial, since its terms have different degree sums.

These polynomials are scalar-valued and will take some number n of input variables

drawn from some ground set, so we need some operator which takes a vector of inputs $\boldsymbol{x} = \langle x_1, x_2, \ldots, x_n \rangle$ and some matrix $\boldsymbol{M}$ and outputs a scalar. In this case we can use the vector-matrix-vector operator $\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x}$. For example, consider $n = 2$ and $\boldsymbol{M} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$. This gives us

$$\begin{aligned} \boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 + 2x_2 & x_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= x_1^2 + 2x_1 x_2 + x_2^2. \end{aligned} \tag{5.1}$$

The reader might notice that this only allows us to represent homogeneous polynomials of degree 2. We would like to be able to express polynomials of arbitrary (even) degree using this method. To achieve this, we need to somehow expand the vector of inputs so they contain every possible monomial of a certain degree. To that end, we use the tensor product over vectors, which we define below.

Definition 5.2 (Tensor product of vectors)**.** Suppose two vectors $\boldsymbol{x}, \boldsymbol{y}$, where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}.$$

The *tensor product* of these vectors, written $\boldsymbol{x} \otimes \boldsymbol{y}$, is the vector containing the product of each pairing of the elements of $\boldsymbol{x}$ and $\boldsymbol{y}$. Formally, this is the vector

$$\boldsymbol{x} \otimes \boldsymbol{y} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_d & x_2 y_1 & x_2 y_2 & \cdots & x_2 y_d & \cdots & x_d y_d \end{bmatrix}^\top.$$

We use the notation $\boldsymbol{x}^{\otimes k}$ for some integer k to denote

$$\boldsymbol{x}^{\otimes k} = \boldsymbol{x} \underbrace{\otimes \cdots \otimes}_{k \text{ times}} \boldsymbol{x}.$$

Trivially, when $k = 1$, this is equivalent to $\boldsymbol{x}$.

Note that $\boldsymbol{x}^{\otimes k}$ gives all monomials of degree k in $\boldsymbol{x}$. Given this definition, we can then proceed to define the shape representation of a homogeneous polynomial of even degree.

Definition 5.3 (Matrix and shape representations of polynomials)**.** Let r be some even natural number. For some degree r polynomial f in n variables $x_1, \ldots, x_n$, we say that an $n^{r/2} \times n^{r/2}$ matrix $\boldsymbol{M}$ is a *matrix representation* of f if, $\forall \boldsymbol{x} = \langle x_1, \ldots, x_n \rangle \in \mathbb{R}^n$,

$$f(x_1, \ldots, x_n) = \left(\boldsymbol{x}^{\otimes r/2}\right)^\top \boldsymbol{M} \boldsymbol{x}^{\otimes r/2}.$$

Additionally, we say that a matrix representation $\boldsymbol{M}$ is a *graph matrix representation* if $\boldsymbol{M} = \boldsymbol{M}_\alpha$ is a graph matrix with corresponding shape α. In this case we also call α a *shape representation* of f.

We provide a few illustrative warm-up examples below.

Example 5.4. *As a brief warm-up example, consider the polynomial*

$$f(x_1, x_2, \ldots, x_n) = \sum_{j_1 \neq j_2 \in [n]} x_{j_1} x_{j_2} = 2x_1x_2 + 2x_1x_3 + \ldots 2x_{n-1}x_n,$$

where the last equality is due to the fact that each pair of monomials $x_{j_1}x_{j_2}$ appears twice with indices in different order. Clearly this is a homogeneous polynomial of degree 2, so we do not need to use the tensor product on our input vector $\boldsymbol{x}$.

To determine the matrix representation $\boldsymbol{M}$ of this polynomial, we can work backwards. If we expand each term $2x_{j_1}x_{j_2}$ into $x_{j_1}x_{j_2} + x_{j_2}x_{j_1}$, we can observe that

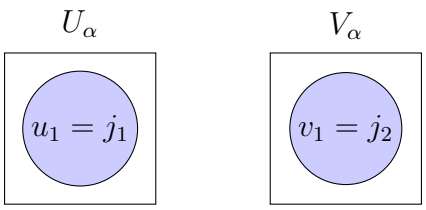


Figure 5.1: Shape representation of the polynomial in Example 5.4.

these entries correspond to the entries $\boldsymbol{M}(j_1, j_2) = \boldsymbol{M}(j_2, j_1) = 1$, since each term only occurs once in the sum, and the term has unit coefficient. Additionally, since the summation is over all pairs of *distinct* indices j_1, j_2, terms like $x_{j_1} x_{j_1}$ do not appear in the polynomial, so the diagonal entries of $\boldsymbol{M}$ are zero.

To obtain a shape representation α that corresponds to this matrix, recall in Definition 2.18 how each vertex corresponds an index in the underlying structure. In this case, we need two vertices in our shape, which we put in the left and right sets U_α and V_α, respectively. Also note that this polynomial is deterministic, i.e. its coefficients do not follow a probability distribution. It follows that no elements of any basis appears in the polynomial, and therefore our shape representation does not need any edges. Thus, we are left with the empty graph on two vertices, shown in Figure 5.1. As this shape has no edges, the left and right vertex sets U_α and V_α are already separated.

Thus, using Theorem 4.10, the norm of the graph matrix has a bound of $\widetilde{O}(n)$. To relate this to a bound on the actual polynomial, recall the definition of a spectral norm in Definition 2.2. As mentioned in Chapter 2, one intuition for the spectral norm $\|\boldsymbol{M}\|$ is the maximum amount that a vector is scaled by that matrix. Similarly, we can view the vector-matrix-vector multiplication operator $\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x}$ as a "generalized dot product" $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$, distorted by $\boldsymbol{M}$ as a basis matrix (or alternatively as the dot product between $\boldsymbol{x}$ and $\boldsymbol{M}\boldsymbol{x}$). It should then follow that, for a unit vector $\boldsymbol{x} = \langle x_1, \ldots, x_n \rangle$,

$$\max_{\|\boldsymbol{x}\|=1} f(x_1, \ldots, x_n) = \boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x} \leq \|\boldsymbol{M}\|.$$

(Note that the same will apply in general we replace $\boldsymbol{x}$ by $\boldsymbol{x}^{\otimes r/2}$ for a homogeneous polynomial of degree r, since our matrix indices will have size $r/2$.) So we can conclude this example by saying $\max_{\|\boldsymbol{x}\|=1} f(\boldsymbol{x}) = \widetilde{O}(n)$.

Let us now modify this polynomial to add some random-valued coefficients. Specifically, we revisit the introductory polynomial used in Example 2.22 to introduce generalized graph matrices.

Example 5.5. *Let $\boldsymbol{A}$ be an $m \times n$ matrix whose values are drawn from the standard normal distribution, where $m \leq n$. Consider the following homogeneous polynomial:*

$$f(x_1, x_2, \ldots, x_n) = \sum_{j_1 \neq j_2 \in [n]} x_{j_1} x_{j_2} \left(\sum_{i \in [m]} \boldsymbol{A}(i, j_1) \big(\boldsymbol{A}(i, j_2)^2 - 1 \big) \right).$$

Here we now have two ground sets: one for row indices $[m]$ (from which i is drawn) and one for column indices $[n]$ (from which j_1, j_2 are drawn). The matrix representation for this polynomial will then be filled with products of normally-distributed entries, because the coefficients themselves are products of normally-distributed random variables.

To come up with the shape representation α for this polynomial, we first observe that the variables of the polynomial have indices j_1, j_2, both of which are column indices. This implies that the two indices I_1, I_2 of our matrix representation $\boldsymbol{M}(I_1, I_2)$ should each be of size 1: one to represent the variable x_{j_1} and another to represent x_{j_2}, as in the previous example. The vertex set of our shape, which represent the indices used in each summand, is then comprised of three vertices: two of the type for column indices, and one of the type for row indices. Since the left and right vertex sets U_α and V_α correspond with the matrix indices, each get one "column index" vertex $u_1 = j_1$ and $v_1 = j_2$, respectively. The remaining vertex for the row index i is placed in the middle vertex set W_α, so we call it $w_1 = i$.

Next, recall that edges in a shape represent the random variables drawn from the

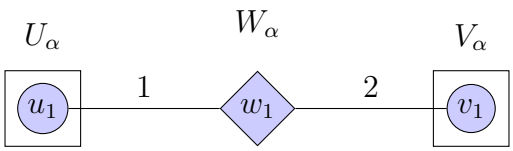


Figure 5.2: Shape representation of the polynomial in Example 5.5. Vertex shaping is used to resemble vertices of different types.

input distribution. In this case, the input distribution $\Omega = \mathcal{N}(0, 1)$, and its basis polynomials are the normalized Hermite polynomials, as shown in Appendix A.1. We give the first few below:

$$\begin{aligned} h_0(x) &= 1 \\ h_1(x) &= x \\ h_2(x) &= \frac{1}{\sqrt{2}}\big(x^2 - 1\big) \\ h_3(x) &= \frac{1}{\sqrt{6}}\big(x^3 - 3x\big) \end{aligned} \tag{5.2}$$

To project the coefficients of our polynomial onto this basis, we take note of the terms for each random variable, in this case $\boldsymbol{A}(i, j_1)$ and $\boldsymbol{A}(i, j_2)^2 - 1$. These are equal to $h_1(\boldsymbol{A}(i, j_1))$ and $\sqrt{2} \cdot h_2(\boldsymbol{A}(i, j_2))$, respectively. This translates to two edges in our shape: one connecting u_1 to w_1 with label 1, and one connecting v_1 to w_1 with label 2. We show the final shape in Figure 5.2.

Finally, to calculate the norm bound using Theorem 4.10, we need to find a minimum-weight U_α, V_α-separator. We have three different choices for a separator: $\{u_1\}$, $\{v_1\}$, or $\{w_1\}$. To determine the best to use, recall that in Definition 2.32 we assign weights according to the size of the ground set of the vertex. In this case, since $m \leq n$, we let $N = \max\{m, n\} = n$ and assign weights as follows: $w(u_1) = w(v_1) = \log_N n$ and $w(w_1) = \log_N m$. This implies that $\{w_1\}$ is our minimum

weight separator, giving us a bound of:

$$\max_{\|\boldsymbol{x}\|=1} f(\boldsymbol{x}) = \widetilde{O}\Big(n^{\frac{1}{2}(2\log_n n + \log_n m - \log_n m)}\Big) = \widetilde{O}(n).$$

Now let's introduce a polynomial whose coefficients are drawn from a multivariate distribution.

Example 5.6 (Adapted from [1, Example 9.2] and [15])**.** *Let $d = 2$, and suppose the following input distribution:*

$$\Omega = \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & 1\end{bmatrix}\right).$$

Note how each sample drawn from this distribution is a vector of length 2 where the two entries have non-zero covariance. Next, suppose some $m \times n \times 2$ tensor $\boldsymbol{B}$ for $n \leq m$, where each "rod" of the tensor $\boldsymbol{B}(i,j) = \langle a_{i,j}, b_{i,j}\rangle \sim \Omega$. That is, each pair of entries $\boldsymbol{B}(i,j,1), \boldsymbol{B}(i,j,2)$ is drawn from the distribution Ω, and thus the entries themselves are dependent on each other. Consider the following polynomial:

$$f(x_1, \ldots, x_n) = \sum_{j_1 \neq j_2 \in [n]} x_{j_1} x_{j_2} \left[\sum_{\substack{i \in [m],\ j_3 \in [n] \\ j_3 \notin \{j_1, j_2\}}} a_{i,j_1} b_{i,j_1} (a_{i,j_3}^2 - 1) b_{i,j_2} \right].$$

Here we have one row index i and three column indices j_1, j_2, j_3, implying that our shape representation α will have one row vertex and three column vertices. We also note that the column indices are *distinct*. Similarly to Example 5.5, the left and right vertex sets will have one column vertex each for j_1 and j_2, since those indices are used in the variables of the polynomial. Call these vertices u_1 and v_1, respectively since they go into U_α and V_α. The remaining vertices for indices i and j_3 go to the middle vertex set W_α, call them w_1 and w_2, respectively. Given the indices of the coefficients,

we can also write our shape's edge set $E(\alpha) = \{\{u_1, w_1\}, \{v_1, w_1\}, \{w_1, w_2\}\}$.

However, we also need to assign labels to these edges. To do this, we first need to compute the basis polynomials of Ω so that we may project the coefficient terms onto that basis. As mentioned in Definition 4.3, we begin with the canonical basis, in this case the basis in two variables (for consistency, we use a and b). If we separate the coefficients by index, we have three terms: $a_{i,j_1}b_{i,j_1}$, $a_{i,j_3}^2 - 1$, and b_{i,j_2}. Because these coefficients are only up to degree two, we only need to consider the canonical basis monomials up to degree two: $\{1, a, b, a^2, ab, b^2\}$. We also take note of the density function of Ω:

$$f_\Omega(a, b) = \frac{1}{\pi\sqrt{3}} \exp\left(-\frac{2}{3}\left(a^2 - ab + b^2\right)\right).$$

Using this density function, we apply the Gram-Schmidt process on the canonical basis to obtain the following basis polynomials up to degree two:

$$\begin{aligned}
h_{0,0}(a, b) &= 1 \\
h_{1,0}(a, b) &= a \\
h_{0,1}(a, b) &= \frac{-a + 2b}{\sqrt{3}} \\
h_{2,0}(a, b) &= \frac{a^2 - 1}{\sqrt{2}} \\
h_{1,1}(a, b) &= \frac{2ab - a^2}{\sqrt{3}} \\
h_{0,2}(a, b) &= \frac{4b^2 - 4ab + a^2 - 3}{3\sqrt{2}}
\end{aligned}$$

See Appendix A.2 for the derivation of these polynomials using the Gram-Schmidt process. To project our coefficients onto this basis, i.e. express each coefficient term as a linear combination of the basis polynomials, for each of the coefficient terms, we can devise a system of equations $\boldsymbol{H}\boldsymbol{c} = \boldsymbol{b}$ and solve for $\boldsymbol{c}$, where $\boldsymbol{H}$ is the matrix of coefficients of each basis polynomial (call this the "basis coefficient matrix"), $\boldsymbol{c}$ is our desired vector of coefficients to apply to our basis polynomial, and $\boldsymbol{b}$ is the vector of

coefficients of the polynomial we are projecting. The matrix $\boldsymbol{H}$ is given by populating each column with the coefficients of each term in the basis polynomial.

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & -\frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & \frac{2}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{3\sqrt{2}} \\ 0 & 0 & 0 & 0 & \frac{2}{\sqrt{3}} & \frac{4}{3\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix}.$$

Each row corresponds to the coefficients of a certain monomial in each basis polynomial. For example, the first row corresponds to the constants in each polynomial, the second to the coefficients of the monomial a, then b, a^2, ab, and finally b^2.

Now consider the first coefficient term $a_{i,j_1}b_{i,j_1}$. Because it only involves the monomial ab, we need only to set $\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}^\top$ and solve $\boldsymbol{Hc} = \boldsymbol{b}$ for $\boldsymbol{c}$. Solving gives us $\boldsymbol{c} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{\sqrt{3}}{2} & 0 \end{bmatrix}^\top$, which tells us that

$$a_{i,j_1}b_{i,j_1} = \frac{1}{2}h_{0,0} + \frac{1}{\sqrt{2}}h_{2,0} + \frac{\sqrt{3}}{2}h_{1,1}.$$

Repeating this process for the rest of the coefficients gives us

$$\begin{aligned} a_{i,j_3}^2 - 1 &= \sqrt{2}h_{2,0} \\ b_{i,j_2} &= \frac{1}{2}h_{1,0} + \frac{\sqrt{3}}{2}h_{0,1}. \end{aligned}$$

Note that $h_{k_1,k_2}(a,b)$ has been shortened to h_{k_1,k_2} for clarity. Additionally, observe that expressing something even as simple as b_{i,j_2} requires more than one basis polynomial due to the covariance between a_{i,j_2} and b_{i,j_2}. This implies that, under the inner product space equipped with the density function f_Ω, it is not possible to express one variable

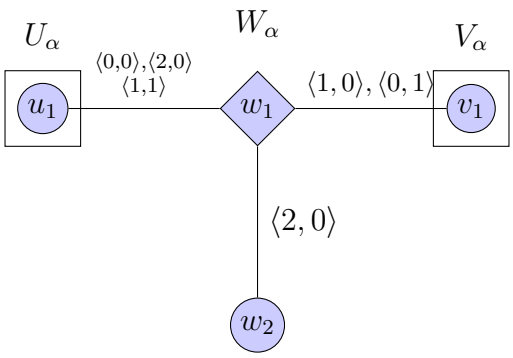


Figure 5.3: Shape representation of the polynomial in Example 5.6. Coloring is used to resemble vertices of different types, and multiple labels are shown to denote arbitrary choice.

b without using the other variable a, which we see here.

These equations tell us the labels that we can assign to each edge in $E(\alpha)$, in particular using the indices of the polynomials on the right-hand sides that have non-zero coefficients. For example, for edge $\{u_1, w_1\}$ representing coefficient $a_{i,j_1}b_{i,j_1}$, we can assign any of $\langle 0, 0\rangle, \langle 2, 0\rangle, \langle 1, 1\rangle$. Since our choice of label for all three edges does not affect the minimum weight separator, we pick some arbitrary label. The final shape representation α is show in Figure 5.3.

Finally, we can compute the minimum weight U_α, V_α-separator and bound the norm of the graph matrix representation, thereby bounding the polynomial over the unit sphere. Recall that $n \leq m$ in our tensor of coefficients $\boldsymbol{B}$, so a minimum weight U_α, V_α separator would be either $\{u_1\}$ or $\{v_1\}$. The size of the largest ground set is m, so using Theorem 4.10, this gives us a bound of

$$\max_{\|\boldsymbol{x}\|=1} f(\boldsymbol{x}) = \widetilde{O}\Big(m^{\frac{1}{2}(3\log_m n+\log_m m-\log_n n)}\Big) = \widetilde{O}(n\sqrt{m}),$$

where the second equality above is due to the following:

$$m^{\frac{1}{2}(3\log_m n+\log_m m-\log_n n)} = m^{\frac{1}{2}(2\log_m n+\log_m m)} = m^{\log_m n+\frac{1}{2}} = n\sqrt{m}.$$

As we have shown, we are able to bound the norm of random polynomials over the unit sphere, where the polynomials' coefficients are sampled from a multivariate distribution (especially one whose random variables are dependent). This can easily be extended for $d > 2$. Importantly, if we had refined bounds for our choice of Ω, i.e. if we had a sufficient bounding function B_Ω and a bound on $h^+_{\boldsymbol{k}}(x)$ as defined in Definition 4.4 and Definition 4.5, we would be able to derive a refined version of Theorem 4.10 specific to Ω by picking a proper minimum over q. This refined bound may give us insight into an optimal choice of label to use for each edge.

5.2 Monochromatic Cliques in Randomly Colored Graphs

Let us return to the CLIQUE indicator matrix example introduced in Section 2.2.1 and introduce more complex dependencies between its entries. Suppose now that, instead of regular cliques in a random graph, we were interested in monochromatic cliques in a randomly-colored complete graph. Specifically, suppose we label (or "color") each edge of a complete graph according to some discrete probability distribution, called the categorical distribution. We would like to compute the expected number of monochromatic 4-cliques in this graph, i.e. the number of 4-cliques in which all edges in the clique have the same color. Before we formalize this problem, we first go over a few definitions. First, we briefly define the categorical distribution.

Definition 5.7 (Categorical distribution)**.** The categorical distribution with parameters $p_1, p_2, \ldots, p_d$ is a discrete probability distribution describing the probability of drawing each of the d possible outcomes. Its support is $[d]$, and its probability mass function is $p(x = i) = p_i$ for $i \in [d]$.

Recall that in the CLIQUE matrix, the distribution of graphs used was $G(n, 1/2)$, the Erdős-Rényi random graph model. The random variables in that example were

appearances of each possible edge in the graph, which followed the Rademacher distribution (each sample taking on $+1$ or -1 with equal probability $1/2$). Clearly this model does not fit our new problem, so we need to define a new model of random graphs.

One of the advantages of working in the multivariate setting is that distributions are able to accommodate more complicated supports. For example, rather than just a single random variable, we can use a distribution whose sample is a d-vector with values in the interval $(0, 1)$ that all sum to 1. In fact, samples from this distribution can themselves be treated as *categorical distributions*. This distribution is called the Dirichlet distribution, and we define it below.

Definition 5.8 (Dirichlet distribution)**.** The Dirichlet distribution with shape parameter $\boldsymbol{s}$, denoted $\mathrm{Dir}(\boldsymbol{s})$ for d-vector with positive entries $\boldsymbol{s}$, is a d-variate probability distribution whose support is the $(d-1)$-simplex. That is for $\langle x_1, \ldots, x_d \rangle$, $x_i \in (0, 1)$ and $\sum_{i=1}^{d} x_i = 1$. The distribution has density function

$$f_{\mathrm{Dir}(\boldsymbol{s})}(x_1, \ldots, x_d) = \frac{1}{B(\boldsymbol{s})} \prod_{i=1}^{d} x_i^{s_i - 1},$$

where $B(\boldsymbol{s})$ is a normalization constant defined as the multivariate beta function

$$B(\boldsymbol{s}) = \frac{\prod_{i=1}^{d} \Gamma(s_i)}{\Gamma(\|\boldsymbol{s}\|_1)}, \quad \Gamma(s) = \int_0^\infty t^{s-1} e^{-t} \mathrm{d}t.$$

For the remainder of this section, we set $\Omega = \mathrm{Dir}(1, 1, \ldots, 1)$, akin to the uniform distribution on the $(d-1)$-simplex. This is also known as the "flat" Dirichlet distribution, as its density function is flat across the simplex. We can now define our random graph model. Our vertex set is the same as that of the Erdős-Rényi model, i.e. $V(G) = [n]$. However, our base graph is the complete graph K_n on the vertex set $[n]$. For each edge, we sample a random variable $\langle p_1, \ldots, p_d \rangle$ from the Dirichlet distribution

Ω. Each p_i is the probability that the edge receives the color i. To encapsulate this into a structure, we let $\boldsymbol{C}$ be an $n \times n \times d$ tensor where

$$\boldsymbol{C}(i,j) = \begin{cases} \langle p_1, \ldots, p_d \rangle \sim \Omega & \text{if } i < j \\ \mathbf{0}_d & \text{if } i = j \\ \boldsymbol{C}(j,i) & \text{otherwise.} \end{cases}$$

Note that $\boldsymbol{C}(i,j)$ is used to denote the "rod" of elements at index (i,j) of the tensor. Then, for each edge $i, j \in G$, we assign color k with probability $\boldsymbol{C}(i,j,k)$ for $k \in [d]$.

We are thus interested in bounding the spectral norm of the following matrix:

$$\mathsf{COLORCLIQUE}((i_1,i_2),(j_1,j_2)) = \begin{cases} \Pr(i_1,i_2,j_1,j_2 \text{ form a monochromatic clique}) \\ \qquad \text{if } i_1,i_2,j_1,j_2 \in [n] \text{ are distinct} \\ 0 \quad \text{otherwise.} \end{cases} \tag{5.3}$$

For simplicity, we assume there are $d = 3$ colors, but remark that this can easily be extended to any finite number of colors. Additionally, for a sample $\langle p_1, p_2, p_3 \rangle \sim \Omega$, note that $p_3 = 1 - (p_1 + p_2)$. In this sense, p_3 is a redundant element, and for this reason, we assume that the density function $f_\Omega(x, y, (1-x-y))$ is in two variables instead, written $f_\Omega(x,y)$.

Theorem 5.9. $\|\mathsf{COLORCLIQUE}\| = \widetilde{O}(n)$.

Proof. The proof of this bound involves a few steps:

1. Split COLORCLIQUE into different "submatrices" such that they sum to the original matrix.

2. Derive the norm bound for each submatrix.

3. Compute the bound of COLORCLIQUE using the bounds of the submatrices.

Step 1 To determine how we split up COLORCLIQUE, we first write an explicit form for each of its entries in the following proposition.

Proposition 5.10. *For $i_1, i_2, j_1, j_2 \in [n]$,*

$$\mathsf{COLORCLIQUE}((i_1, i_2), (j_1, j_2)) = \sum_{k=1}^{d} \prod_{i \neq j \in \{i_1, i_2, j_1, j_2\}} \boldsymbol{C}(i, j, k).$$

Proof of Proposition 5.10. For fixed matrix indices $(i_1, i_2), (j_1, j_2)$, consider the subgraph H of G where $V(H) = \{i_1, i_2, j_1, j_2\}$. (Note that H is the complete graph on 4 vertices.) Then, the matrix entry is the probability that H is a monochromatic clique. Recall that each edge $\{i, j\}$ in G has a color k with probability $\boldsymbol{C}(i, j, k)$, and that the color of each edge is sampled from its own categorical distribution, so they are independent events. Thus, the probability that H is a monochromatic clique of color k is the product of the probabilities that each edge in H has color k, giving us the product of $\boldsymbol{C}(i, j, k)$ for each $\{i, j\} \in E(H)$. Adding this up for all $k \in [d]$, we recover the sum in the proposition. □

Looking ahead to projecting each matrix entry onto the basis for Ω, we see that each entry is the sum of products of elements from a sample of Ω. In particular, the sum is over each index k, and the product is over the kth element of each sample. To make analysis easier, we therefore break up COLORCLIQUE by color into d submatrices. That is, for $k \in [d]$, we define

$$\mathsf{CLIQUE}_k((i_1, i_2), (j_1, j_2)) = \prod_{i \neq j \in \{i_1, i_2, j_1, j_2\}} \boldsymbol{C}(i, j, k). \tag{5.4}$$

It is then clear that

$$\mathsf{COLORCLIQUE} = \sum_{k=1}^{d} \mathsf{CLIQUE}_k. \tag{5.5}$$

Step 2 Now we find the norm bound of each submatrix CLIQUE_k. To do this, we first give an orthonormal basis for Ω, project each entry of CLIQUE_k onto this basis, then derive the shape representation of each entry. Note that each entry in CLIQUE_k is made up of terms that are of degree 1 for each pair of indices i, j. For this reason, we only need to apply the Gram-Schmidt process on the canonical basis monomials up to degree 1, i.e. $\{1, x, y\}$. See Appendix A.3 for the application of this process. This gives us the three basis polynomials below:

$$\begin{aligned} h_{0,0}(x, y) &= 1 \\ h_{1,0}(x, y) &= \sqrt{2}(3x - 1) \\ h_{0,1}(x, y) &= \sqrt{6}(2y + x - 1), \end{aligned}$$

giving us the basis coefficient matrix

$$\boldsymbol{H} = \begin{bmatrix} 1 & \sqrt{2} & -\sqrt{6} \\ 0 & 3\sqrt{2} & \sqrt{6} \\ 0 & 0 & 2\sqrt{6} \end{bmatrix}.$$

If we expand Equation (5.4) for each k, we get

$$\begin{aligned} \mathsf{CLIQUE}_1((i_1, i_2), (j_1, j_2)) &= x_{i_1,i_2} x_{i_1,j_1} x_{i_1,j_2} x_{i_2,j_1} x_{i_2,j_2} x_{j_1,j_2} \\ \mathsf{CLIQUE}_2((i_1, i_2), (j_1, j_2)) &= y_{i_1,i_2} y_{i_1,j_1} y_{i_1,j_2} y_{i_2,j_1} y_{i_2,j_2} y_{j_1,j_2} \\ \mathsf{CLIQUE}_3((i_1, i_2), (j_1, j_2)) &= z_{i_1,i_2} z_{i_1,j_1} z_{i_1,j_2} z_{i_2,j_1} z_{i_2,j_2} z_{j_1,j_2}, \end{aligned} \tag{5.6}$$

where $x_{i,j} = \boldsymbol{C}(i, j, 1)$, $y_{i,j} = \boldsymbol{C}(i, j, 2)$, and $z_{i,j} = \boldsymbol{C}(i, j, 3) = 1 - x_{i,j} - y_{i,j}$.

Like in Section 5.1, we now need to project these expressions onto our basis for Ω.

In other words, we need to solve the equation $\boldsymbol{H}\boldsymbol{c} = \boldsymbol{b}$ for $\boldsymbol{c}$ given $\boldsymbol{H}$ from above and

$$\boldsymbol{b} \in \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \right\}.$$

Solving for each value of $\boldsymbol{b}$ gives us the following projections:

$$\begin{aligned} x_{i,j} = \boldsymbol{C}(i,j,1) &= -\frac{1}{3}h_{0,0} + \frac{1}{3\sqrt{2}}h_{1,0} \\ y_{i,j} = \boldsymbol{C}(i,j,2) &= \frac{2}{3}h_{0,0} - \frac{1}{6\sqrt{2}}h_{1,0} + \frac{1}{2\sqrt{6}}h_{0,1} \\ 1 - x_{i,j} - y_{i,j} = \boldsymbol{C}(i,j,3) &= \frac{2}{3}h_{0,0} - \frac{1}{6\sqrt{2}}h_{1,0} - \frac{1}{2\sqrt{6}}h_{0,1}, \end{aligned}$$

where $h_{r_1,r_2}(x_{i,j}, y_{i,j})$ is shortened to h_{r_1,r_2} for clarity. This means that, while the shape of each CLIQUE_k (call it α_k) will have the same vertex and edge sets (they use the same indices, after all), the edges potentially could have different labels. For example, while the edges in α_2 could have any of the labels $\langle 0,0\rangle, \langle 1,0\rangle, \langle 1,0\rangle$, those in α_1 could only have the label $\langle 0,0\rangle$ and $\langle 1,0\rangle$.

Next, for each k, we bound $\|\mathsf{CLIQUE}_k\|$ by deriving the shape representation. Note that there is only one ground set $[n]$, so all vertices will be of the same type. We have four indices i_1, i_2, j_1, j_2 used in the formula for each entry, so our shapes will have four vertices representing these indices. Additionally, the matrix indices are (i_1, i_2) and (j_1, j_2), so the left vertex set $U_\alpha = \{u_1 = i_1, u_2 = j_2\}$ and the right vertex set $V_\alpha = \{v_1 = j_1, v_2 = j_2\}$. Finally, recall that in Equation (5.4), the product is taken over all possible (distinct) pairs of the four indices, i.e. all edges on the subgraph H. It follows that the edge set of our shape will then have all possible edges on the vertex set. The final shape representation of CLIQUE_k is shown in Figure 5.4.

Since this is a complete graph, the minimum U_α, V_α-separator is in fact either U_α or V_α. In either case, the separator will have 2 vertices, so we can then invoke

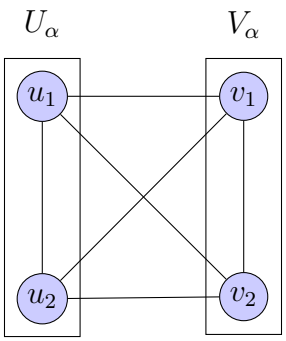


Figure 5.4: Shape representation of CLIQUE_k. Choice of labels for each edge depend on k.

Theorem 4.10 to show that

$$\|\mathsf{CLIQUE}_k\| = \widetilde{O}(n^{\frac{1}{2}(4-2)}) = \widetilde{O}(n).$$

Step 3 Recall that all matrix norms obey the triangle inequality. Therefore, using Equation (5.5),

$$\|\mathsf{COLORCLIQUE}\| = \left\| \sum_{k=1}^{d} \mathsf{CLIQUE}_k \right\| \leq \sum_{k=1}^{d} \|\mathsf{CLIQUE}_k\|.$$

Since we have a bound on $\|\mathsf{CLIQUE}_k\|$ for each k, adding these up gives a bound on $\|\mathsf{COLORCLIQUE}\|$ of $\widetilde{O}(n)$, concluding the proof. □

Again, while we only showed this for $d = 3$, it is simple to expand this to higher (but finite) values of d, since it would only leave us with more submatrices of the same kind to bound.

5.3 The Uniform Distribution on Quantum States

One potential application for the multivariate setting is representing expressions involving quantum pure states with shapes. As a brief introduction, consider the state of some system in classical computing, such as the state of a bit b in memory.

This bit can either be set to 1 or 0, so overall one number is used to specify the state of this system. In contrast, a quantum system is described in qubits $|\phi\rangle$, such as the polarization of a photon. A 2-state quantum system is described in the form $\alpha|0\rangle + \beta|1\rangle$, where $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$, and $|0\rangle$ and $|1\rangle$ are basis states. (This is sometimes written as a vector $[\alpha \quad \beta]^\top$.) In other words, the state of a quantum system can be described as a complex convex combination of basis states, Such a state is called a "pure state", and is said to be in a "superposition" of basis states. (In contrast, a "mixed state" cannot be expressed purely in terms of basis states, but for now we focus on pure states.)

This was our primary motivation for investigating graph matrices under the multivariate setting. Letting the input distribution Ω be multivariate increases the expressive power of graph matrices, since we can use them to represent structures that follow a more strict set of rules. In this case, we would like a way to uniformly sample from the space of quantum pure states in d dimensions, i.e. a quantum system whose state can be described as $\sum_{k=0}^{d-1} \alpha_k |k\rangle$, where $\alpha_k \in \mathbb{C}$, $\sum_{k=0}^{d-1} |\alpha_k|^2 = 1$, and $|k\rangle$ is a basis state.

Khatri in [17] discusses such a distribution, calling it a Haar measure, and expands on this distribution in a note [18]. Specifically, we are interested in the following procedure:

Proposition 5.11. *To sample a random pure state uniformly from distribution on pure states in d dimensions,*

1. *Let Ω be the $2d$-variate standard normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_{2d})$. From this distribution, sample $\langle a_0, a_1, \ldots, a_{d-1}, b_0, \ldots, b_{d-1}\rangle \sim \Omega$.*

2. *Let $N = \sum_{k=0}^{d-1}(a_k + b_k)^2$. Construct the pure state as*

$$|\phi\rangle = \sum_{k=0}^{d-1}\left(\frac{a_k}{N} + \frac{b_k}{N}i\right)|k\rangle\,.$$

Khatri proves this in [18] by first showing that the uniform distribution on d-dimensional pure states has density function $f_d(\phi) = \frac{(d-1)!}{\pi^{d-1}}$ using a derivation of the volume of the space of pure states. If we consider $|\phi\rangle$ in the form $\sum_{k=0}^{d-1} \sqrt{p_k} e^{\imath\varphi_k} |k\rangle$, where p_k is a probability and $\varphi \in [0, 2\pi]$ is an angle, marginalizing the angles out gives us a uniform distribution on the $(d-1)$-simplex, which happens to coincide with the density function of the Dirichlet distribution $\mathrm{Dir}_d(1, 1, \ldots, 1)$. This implies that the probabilities $\{p_k\}_{k=0}^{d-1}$ jointly follow the same distribution. Khatri then proves a theorem that shows that the above random variables $\frac{a_k}{N}$ and $\frac{b_k}{N}$ follow the same distribution.

This procedure would then allow us to use $\Omega = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{2d})$ as our input distribution and produce large random expressions involving quantum states formed using the expression in Proposition 5.11.

Another recently published article from Chen and Tumulka [19] defines a uniform distribution on density matrices, which are used to expressed mixed states, i.e. quantum states which cannot be expressed purely in terms of basis states. This mirrored prior work from Hall [20], answering a similar question of how to derive the "most random" distribution of possible states of a quantum system. In a purely mathematical sense, density matrices are simply complex matrices that are Hermitian, positive definite, and have unit trace. A distribution on such matrices could be used to form even more complicated expressions involving quantum states, such as large random quantum operators. We refer readers to Bengtsson and Życzkowski's text [21] for more details on the geometry of the space of density matrices and of quantum states more generally.

Additionally, while this thesis focused on bounding the spectral norm of graph matrices, further work has been done on analyzing other properties. Cai and Potechin in [22] determined the limiting distribution of singular values of graph matrices whose shape forms a 'Z'. An important property of quantum operators states that the possible measurement outcomes of an operator $\boldsymbol{A}$ are the set of eigenvalues of $\boldsymbol{A}$. Further

work in investigating the extent of the expressive power of graph matrices under a multivariate distribution remains, and we leave such exploration to future research.

Appendix A

Orthonormal Bases for Input Distributions

In this appendix, we show the Gram-Schmidt process for deriving orthonormal basis for some input distribution Ω. We begin with the univariate standard normal distribution as a warm-up in Appendix A.1, then move on to a multivariate distribution with covariance in Appendix A.2. Finally, in Appendix A.3 we tackle a different distribution used in Section 5.2, the "flat" Dirichlet distribution.

A.1 Standard Normal Distribution

Recall that in the space of polynomials with real coefficients $\mathbb{R}[x]$, we defined our canonical basis in Definition 2.10 as the set $\{p_i(x)\}_{i=0}^{\infty} = \{1, x, x^2, x^3, \ldots\}$. The first step in the Gram-Schmidt process is to define the inner product space over which we orthogonalize this basis. The standard normal distribution $\mathcal{N}(0,1)$ has density function:

$$f_{\mathcal{N}(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

This gives us the inner product between two polynomials under $\mathcal{N}(0,1)$:

$$\langle p_i(x), p_j(x) \rangle = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} p_i(x) p_j(x) e^{-x/2} \mathrm{d}x,$$

where the integral is taken over the support of the distribution. For clarity, throughout this chapter we use p_i, q_i, h_i as shorthand for $p_i(x), q_i(x), h_i(x)$, respectively.

Next, we use the Gram-Schmidt process to orthogonalize the canonical basis. Looking at the process as defined in Definition 4.2, we start by calculating the polynomials q_k using the projection operator, then normalizing each q_k with respect to our inner product to obtain our desired basis polynomials h_k. Trivially we have that $p_0(x) = q_0(x) = 1$. Normalizing this gives us

$$h_0(x) = \frac{q_0}{\sqrt{\langle q_0, q_0 \rangle}} = \frac{1}{(2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-x^2/2} \mathrm{d}x} = \frac{1}{\sqrt{(2\pi)^{-1/2}(2\pi)^{1/2}}} = 1.$$

We note that $\int_{-\infty}^{\infty} e^{-x^2/2}\mathrm{d}x = \sqrt{2\pi}$.

Next, we orthogonalize $p_1(x) = x$. We first use the projection operator to calculate q_1:

$$\begin{aligned} q_1 &= p_1 - \mathrm{proj}_{q_0}(p_1) = x - \mathrm{proj}_1(x) \\ &= x - \frac{\langle 1, x\rangle}{\langle 1, 1\rangle} \cdot 1 \\ &= x - \frac{\int_{\infty}^{\infty} xe^{-x/2}\mathrm{d}x}{1} \cdot 1 \\ &= x - \frac{0}{1} \\ &= x. \end{aligned}$$

Then we normalize q_1:

$$h_1 = \frac{x}{\sqrt{\langle x, x\rangle}} = \frac{x}{1} = x,$$

which turns out to have no effect, giving us $h_1(x) = x$.

Notice how the first two basis polynomials are actually equivalent to the canonical basis monomials. The first one is trivial since $q_1 = p_1$ in general, but the second one is more notable because it is explained by the fact that the first moment (more generally, odd moments) of the normal distribution, i.e. its mean is equivalent to zero:

$$\mu_{\mathcal{N}(0,1)}(j) = \mathbb{E}_{\Omega}[x^j] = \begin{cases} 0 & \text{if } j \text{ is odd} \\ (j-1)!! & \text{if } j \text{ is even,} \end{cases} \tag{A.1}$$

where $n!!$ denotes the double factorial $n(n-2)(n-4)\cdots 1$. We use this when calculating $\langle 1, x\rangle = \mathbb{E}_{\mathcal{N}(0,1)}[x] = \mu_{\mathcal{N}(0,1)}(1) = 0$. We will return to this fact later as we calculate more basis polynomials, because it leads to another observation about these basis polynomials further along the Gram-Schmidt process.

Next in the process is $p_2(x) = x^2$. We first use the projection operator to calculate q_2:

$$\begin{aligned}
q_2 &= p_2 - \text{proj}_{q_0}(p_2) - \text{proj}_{q_1}(p_2) = x^2 - \text{proj}_1(x^2) - \text{proj}_x(x^2) \\
&= x^2 - \frac{\langle 1, x^2 \rangle}{\langle 1, 1 \rangle} \cdot 1 - \frac{\langle x, x^2 \rangle}{\langle x, x \rangle} \cdot x \\
&= x^2 - \frac{\int_{-\infty}^{\infty} x^2 e^{-x/2} \mathrm{d}x}{1} - \frac{\int_{-\infty}^{\infty} x e^{-x/2} \mathrm{d}x}{1} \cdot x \\
&= x^2 - \mu_{\mathcal{N}(0,1)}(2) - \mu_{\mathcal{N}(0,1)}(1)x \\
&= x^2 - 1.
\end{aligned}$$

Then normalize it:

$$h_2 = \frac{x^2 - 1}{\sqrt{\langle x^2 - 1, x^2 - 1 \rangle}} = \frac{x^2 - 1}{\sqrt{\int_{-\infty}^{\infty} (x^2 - 1)^2 e^{x^2/2} \mathrm{d}x}} = \frac{x^2 - 1}{\sqrt{2}},$$

giving us the basis polynomial $h_2(x) = \frac{x^2 - 1}{\sqrt{2}}$.

Continuing this process for the next few canonical basis monomials gives us the following:

$$\begin{aligned}
q_3(x) &= x^3 - x & h_3(x) &= \frac{x^3 - x}{\sqrt{6}} \\
q_4(x) &= x^4 - 6x^2 + 3 & h_4(x) &= \frac{x^4 - 6x^2 + 3}{\sqrt{24}} \\
q_5(x) &= x^5 - 10x^3 + 15x & h_5(x) &= \frac{x^5 - 10x^3 + 15x}{\sqrt{120}}.
\end{aligned}$$

At this point, the reader may notice that each polynomial only uses monomials of the same parity as its index. This is due to the fact that the projection step for monomial x^j involves subtracting polynomials $q_i(x)$ with coefficients that have $\mu_{\mathcal{N}(0,1)}(i)$ as factors from x^j. In particular, when i is odd, this moment factor is zero, removing that monomial altogether from the projection. Additionally, when

orthogonalizing, the reader may notice $h_j(x) = \frac{1}{\sqrt{j!}} q_j(x)$. We will stop the process here but note that the process continues infinitely since the distribution has a support of infinite size. However, for a given expression of degree j we can truncate this process at $p_j(x) = x^j$, since the resulting projection will never use monomials of any degree greater than j.

The polynomials q_i are commonly known as the probabilist's Hermite polynomials. Rahman gives us the following formula for these Hermite polynomials based on the density function of the normal distribution [23]:

$$H_j(x) = \frac{(-1)^j}{f_{\mathcal{N}(0,1)}(x)} \frac{\mathrm{d}^j}{\mathrm{d}x^j} f_{\mathcal{N}(0,1)}(x),$$

where the normalization is given by

$$h_j(x) = \frac{H_j(x)}{\sqrt{\mathbb{E}_{\mathcal{N}(0,1)}[H_j^2(x)]}}.$$

A.2 Multivariate Normal Distribution

In this section, we set Ω to the same input distribution used in Example 5.6,

$$\Omega = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \right).$$

Notably, the two elements of random vectors sampled from this distribution have a covariance of $\frac{1}{2}$. For clarity, since there are only two elements (i.e. $d = 2$), we write them as x and y. Again, we start with defining the inner product space.

Since we are now in the multivariate setting, we move to the 2-variate canonical basis monomials $\{p_{i,j}(x, y)\}_{i,j} = \{1, x, y, x^2, xy, y^2, \ldots\}$. We note that there are many orderings of this basis since it is in multiple variables, and in this case we use the graded

lexicographic order: first comparing the degree sum, then applying a lexicographic ordering on monomials of the same degree sum. We also use the degree in x and y of each monomial as the indices i and j in $p_{i,j}(x, y)$, respectively. For example, $p_{1,2}(x, y) = xy^2$. Similarly to the previous section, we use the notation $p_{i,j}$ as shorthand for clarity.

To find the density function for this distribution, first recall the formula for the density function of a general multivariate normal distribution from Definition 2.38. We substitute our specific mean vector (the zero vector) and covariance matrix into this formula to get

$$\begin{aligned}
f_\Omega(x, y) &= \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right) \\
&= \frac{1}{\sqrt{(2\pi)^2 \begin{vmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{vmatrix}}} \exp\left(-\frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right) \\
&= \frac{1}{\sqrt{4\pi^2 \cdot \frac{3}{4}}} \exp\left(-\frac{1}{2} \cdot \left(x \cdot \left(\frac{4x}{3} - \frac{2y}{3}\right) + y \cdot \left(-\frac{2x}{3} + \frac{4y}{3}\right)\right)\right) \\
&= \frac{1}{\pi\sqrt{3}} \exp\left(-\frac{2}{3}(x^2 - xy + y^2)\right).
\end{aligned}$$

This gives us the inner product under Ω:

$$\langle p_{i_1,j_1}, p_{i_2,j_2} \rangle = \frac{1}{\pi\sqrt{3}} \int_{\mathbb{R}^2} p_{i_1,j_1}(x, y) p_{i_2,j_2}(x, y) \exp\left(-\frac{2}{3}(x^2 - xy + y^2)\right) \mathrm{d}x\mathrm{d}y.$$

Now that we have an inner product defined on our polynomial space, we can use it to orthogonalize the canonical basis monomials using the Gram-Schmidt process. This will be very similar to Appendix A.1, with the primary difference being the indices

used. Again, the trivial monomial $p_{0,0}(x,y) = q_{0,0}(x,y) = 1$. Normalizing it, we get

$$\begin{aligned} h_{0,0}(x,y) &= \frac{q_{0,0}}{\sqrt{\langle q_{0,0}, q_{0,0}\rangle}} \\ &= \frac{1}{\left(\pi\sqrt{3}\right)^{-1}\int_{\mathbb{R}^2}\exp\left(\frac{2}{3}(x^2-xy+y^2)\right)\mathrm{d}x\mathrm{d}y} \\ &= \frac{1}{\left(\pi\sqrt{3}\right)^{-1}\pi\sqrt{3}} \\ &= 1. \end{aligned}$$

After applying this process to a few of the monomials, we get the following:

$$\begin{aligned} q_{1,0}(x,y) &= x & h_{1,0}(x,y) &= x \\ q_{0,1}(x,y) &= -x+2y & h_{0,1}(x,y) &= \frac{-x+2y}{\sqrt{3}} \\ q_{2,0}(x,y) &= x^2-1 & h_{2,0}(x,y) &= \frac{x^2-1}{\sqrt{2}} \\ q_{1,1}(x,y) &= 2xy-x^2 & h_{1,1}(x,y) &= \frac{2xy-x^2}{\sqrt{3}} \\ q_{0,2}(x,y) &= 4y^2-4xy+x^2-3 & h_{0,2}(x,y) &= \frac{4y^2-4xy+x^2-3}{3\sqrt{2}} \end{aligned}$$

Again, note how each basis polynomial only uses monomials of the same parity in degree sum. For example, $h_{1,1}$, the polynomial taken from xy in the canonical basis, only uses monomials x^2, y^2, xy, and a constant, where $h_{0,1}$ uses monomials x and y (with no constant). This is similar to the basis for the univariate normal distribution, as their moment functions share a similar property. Additionally, we also note that this process continues infinitely since the support of Ω is of infinite size.

Rahman also gives an expression for multivariate Hermite polynomials for a given covariance matrix [23], which also can serve as an orthonormal basis for Ω:

$$H_{\boldsymbol{j}}^{\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{(-1)^{\|\boldsymbol{j}\|_1}}{f_{\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})}(\boldsymbol{x})}\frac{\partial^j}{\partial \boldsymbol{x}^j} f_{\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})}(\boldsymbol{x}),$$

where $\frac{\partial^{\boldsymbol{j}}}{\partial \boldsymbol{x}^{\boldsymbol{j}}}$ is shorthand for $\partial^{\|\boldsymbol{j}\|_1}/\partial x_1^{j_1} \cdots \partial x_d^{j_d}$. Plugging our covariance matrix in, we obtain the polynomials:

$$
\begin{aligned}
h_{0,0}(x,y) &= 1 \\
h_{1,0}(x,y) &= \frac{1}{\sqrt{3}}(2x-y) \\
h_{0,1}(x,y) &= \frac{1}{\sqrt{3}}(-x+2y) \\
h_{1,1}(x,y) &= \frac{1}{3\sqrt{5}}(3+10xy-4x^2-4y^2) \\
h_{2,0}(x,y) &= \frac{1}{3\sqrt{2}}(-3-4xy+4x^2+y^2) \\
h_{0,2}(x,y) &= \frac{1}{3\sqrt{2}}(-3-4xy+x^2+4y^2).
\end{aligned}
$$

These are mostly different from the basis we obtained through the Gram-Schmidt process, but we remark that different basis polynomials may be obtained by changing the ordering of the canonical basis monomials. As noted earlier, we used the graded lexicographic ordering.

A.3 Flat Dirichlet Distribution

The Dirichlet distribution $\mathrm{Dir}(\boldsymbol{s})$, defined in Definition 5.8, is a multivariate distribution with a support that is notably different from the previous two distributions we have examined. Rather than being $\mathbb{R}^d$, the support is a set of d-dimensional vectors where each element is positive and the sum of these elements is one. Alternatively, one can consider the last element x_d of these vectors as simply $1 - \sum_{i=1}^{d-1} x_j$, in which case the support can also be defined as the open interval $(0,1)^{d-1}$. In particular, we will examine the flat 3-variate Dirichlet distribution $\Omega = \mathrm{Dir}(\langle 1,1,1 \rangle)$, but for simplicity we will do it in two variables x, y for the reason mentioned above, i.e. that the third variable is simply $1 - x - y$ and the resulting support is easier to use for integration.

Having different support does not change the basis over which we apply the Gram-Schmidt process however: we still use the canonical basis monomials (specifically in graded lexicographic order). Rather, the support changes our domain of integration when defining the inner product. The density function of Ω turns out to be a constant:

$$\begin{aligned} f_\Omega(x,y) &= \frac{1}{B(\langle 1,1,1\rangle)} x^{1-1} y^{1-1} \\ &= \frac{\Gamma(1+1+1)}{\Gamma(1)\cdot\Gamma(1)\cdot\Gamma(1)} = \frac{2!}{(0!)^3} \\ &= 2, \end{aligned}$$

which gives the "flat" Dirichlet distribution its nickname. Our inner product then becomes

$$\langle p_{i_1,j_1}, p_{i_2,j_2}\rangle = 2\int_{(0,1)^{d-1}} p_{i_1,j_1}\cdot p_{i_2,j_2}\mathrm{d}x\mathrm{d}y.$$

Now we generate our orthogonal basis, starting with $h_{0,0} = q_{0,0} = p_{0,0} = 1$. Applying the Gram-Schmidt process as before gives us:

$$\begin{aligned} q_{1,0} &= 3x-1 & h_{1,0} &= \sqrt{2}(3x-1) \\ q_{0,1} &= 2x+y-1 & h_{0,1} &= \sqrt{6}(2x+y-1) \\ q_{2,0} &= 10x^2-8x-1 & h_{2,0} &= \sqrt{3}(10x^2-8x-1) \\ q_{1,1} &= 10xy+5x^2-6x-2y+1 & h_{1,1} &= 3\cdot(10xy+5x^2-6x-2y+1) \\ q_{0,2} &= 6y^2+x^2+6xy-2x-6y+1 & h_{0,2} &= \sqrt{15}(6y^2+x^2+6xy-2x-6y+1). \end{aligned}$$

Bibliography

[1] K. Ahn, D. Medarametla, and A. Potechin, "Graph Matrices: Norm Bounds and Applications," Oct. 9, 2020. arXiv: 1604.03423 [math].

[2] E. P. Wigner, "Characteristic Vectors of Bordered Matrices With Infinite Dimensions," *Annals of Mathematics*, vol. 62, no. 3, pp. 548–564, 1955, ISSN: 0003-486X. DOI: 10.2307/1970079. JSTOR: 1970079.

[3] J. WISHART, "THE GENERALISED PRODUCT MOMENT DISTRIBUTION IN SAMPLES FROM A NORMAL MULTIVARIATE POPULATION," *Biometrika*, vol. 20A, no. 1-2, pp. 32–52, Dec. 1, 1928, ISSN: 0006-3444. DOI: 10.1093/biomet/20A.1-2.32.

[4] J. Verbaarschot and T. Wettig, "Random Matrix Theory and Chiral Symmetry in QCD," *Annual Review of Nuclear and Particle Science*, vol. 50, no. 1, pp. 343–410, Dec. 2000, ISSN: 0163-8998, 1545-4134. DOI: 10.1146/annurev.nucl.50.1.343.

[5] S. B. Hopkins, P. K. Kothari, and A. Potechin, "SoS and Planted Clique: Tight Analysis of MPW Moments at all Degrees and an Optimal Lower Bound at Degree Four," Jul. 18, 2015. arXiv: 1507.05230 [cs].

[6] A. Potechin and D. Steurer, "Exact tensor completion with sum-of-squares," Jun. 23, 2017. arXiv: 1702.06237 [cs, math, stat].

[7] P. Raghavendra, S. Rao, and T. Schramm, "Strongly Refuting Random CSPs Below the Spectral Threshold," Nov. 3, 2016. arXiv: 1605.00058 [cs].

[8] B. Barak and A. Moitra, "Noisy Tensor Completion via the Sum-of-Squares Hierarchy," Feb. 18, 2016. arXiv: 1501.06521 [cs, stat].

[9] S. Aaronson and A. Arkhipov, "The Computational Complexity of Linear Optics," *Theory of Computing*, vol. 9, no. 1, pp. 143–252, Feb. 9, 2013, ISSN: 1557-2862. DOI: 10.4086/toc.2013.v009a004.

[10] B. Barak, S. B. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, "A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem," Apr. 12, 2016. arXiv: 1604.03084 [cs].

[11] E. P. Wigner, "On the Distribution of the Roots of Certain Symmetric Matrices," *Annals of Mathematics*, vol. 67, no. 2, pp. 325–327, 1958, ISSN: 0003-486X. DOI: 10.2307/1970008. JSTOR: 1970008.

[12] V. L. Girko, "Circular Law," *Theory of Probability & Its Applications*, vol. 29, no. 4, pp. 694–706, Jan. 1, 1985, ISSN: 0040-585X. DOI: 10.1137/1129095.

[13] A. S. Bandeira and R. van Handel, "Sharp nonasymptotic bounds on the norm of random matrices with independent entries," *The Annals of Probability*, vol. 44, no. 4, pp. 2479–2506, Jul. 2016, ISSN: 0091-1798, 2168-894X. DOI: 10.1214/15-AOP1025.

[14] Z. D. Bai and Y. Q. Yin, "Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix," *The Annals of Probability*, vol. 16, no. 4, pp. 1729–1741, Oct. 1988, ISSN: 0091-1798, 2168-894X. DOI: 10.1214/aop/1176991594.

[15] S. B. Hopkins, T. Schramm, J. Shi, and D. Steurer, "Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors," Feb. 3, 2016. arXiv: 1512.02337 [cs, stat].

[16] R. van Handel, "Structured Random Matrices," in *Convexity and Concentration*, E. Carlen, M. Madiman, and E. M. Werner, Eds., ser. The IMA Volumes in Mathematics and Its Applications, New York, NY: Springer, 2017, pp. 107–156, ISBN: 978-1-4939-7005-6. DOI: 10.1007/978-1-4939-7005-6_4.

[17] S. Khatri and M. M. Wilde, "Principles of Quantum Communication Theory: A Modern Approach," Nov. 9, 2020. DOI: 10.48550/arXiv.2011.04672.

[18] S. Khatri, "Random Pure States," *(unpublished)*, May 30, 2020.

[19] E. K. Chen and R. Tumulka, "Uniform probability distribution over all density matrices," *Quantum Studies: Mathematics and Foundations*, Jan. 16, 2022, ISSN: 2196-5617. DOI: 10.1007/s40509-021-00267-5.

[20] M. J. W. Hall, "Random quantum correlations and density operator distributions," *Physics Letters A*, vol. 242, no. 3, pp. 123–129, May 1998, ISSN: 03759601. DOI: 10.1016/S0375-9601(98)00190-X. arXiv: quant-ph/9802052.

[21] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, 2nd ed. Cambridge: Cambridge University Press, 2017, ISBN: 978-1-139-20701-0. DOI: 10.1017/9781139207010.

[22] W. Cai and A. Potechin, "The Spectrum of the Singular Values of Z-Shaped Graph Matrices," Jun. 24, 2020. arXiv: 2006.14144 [cs, math].

[23] S. Rahman, "Wiener–Hermite polynomial expansion for multivariate Gaussian probability measures," *Journal of Mathematical Analysis and Applications*, vol. 454, no. 1, pp. 303–334, Oct. 1, 2017, ISSN: 0022-247X. DOI: 10.1016/j.jmaa.2017.04.062.