Complex Matrix Scalings, Extremal Permanents, and the Geometric Measure of Entanglement

by

George Hutchinson

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Mathematics

Guelph, Ontario, Canada

ABSTRACT

COMPLEX MATRIX SCALINGS, EXTREMAL PERMANENTS, AND THE GEOMETRIC MEASURE OF ENTANGLEMENT

George Hutchinson

University of Guelph, 2018

Advisor:

Dr. Rajesh Pereira

An $n \times n$ matrix with complex entries is said to be doubly quasi-stochastic (DQS) if all row and column sums are equal to one. Given a positive definite matrix A and a diagonal matrix D, we say that D^*AD is a (complex matrix) scaling of A if D^*AD is doubly quasi-stochastic. Motivated by a result of Pereira and Boneng concerning the application of complex matrix scalings to the geometric measure of entanglement of certain symmetric states, we embark upon an investigation of these scalings and their properties.

We begin with an existence theorem that unifies complex matrix scalings together with other classical notions of scaling (Sinkhorn scaling, Marshall and Olkin's scaling of copositive real matrices, etc.). We then discuss the notion of double quasi-stochasticity as it pertains to tensors (ie. hypermatrices) and we extend some classical scaling results to these higher-order objects.

Returning to the study of complex matrix scalings of positive definite matrices, we disprove a conjecture of Pereira and Boneng concerning the cardinality of the set of complex matrix scalings for a given positive definite matrix A. In particular, we show that a given 3×3 real matrix has at most 6 scalings, and that when $n \geq 4$ there exist $n \times n$ matrices with infinitely many scalings. Using the application to quantum entanglement as motivation, we consider complex matrix scalings of extremal permanent; that is, given positive definite A, we investigate the scaling(s) of A that have maximal (minimal) permanent. We prove that scalings with maximal permanent satisfy a certain optimization condition and use this condition to derive topological properties of this set as well as a lower bound on the permanent of these "maximal scalings". We also arrive at an upper bound on the permanent of certain "minimal scalings" and use these results to bound the geometric measure of entanglement of symmetric states that satisfy certain conditions.

We close with a discussion of possible future work and open problems in the study of matrix scalings, including: a scaling algorithm that builds on the work of O'Leary; the permanent conjecture of Chollet and Drury; the problem of maximizing the permanent of an $n \times n$ matrix with prescribed eigenvalues; and mutually unbiased bases.

Dedication

This thesis is dedicated to my family. Without their love, support, guidance, and friendship, this work would not have been possible.

Acknowledgements

I would like to thank my advisor, Rajesh Pereira, for his guidance and support throughout my doctoral studies. He has been wonderful to work with, and I consider myself extremely lucky to have had the opportunity to learn from him over the past three and a half years.

I would also like to thank the other members of of my advisory committee, David Kribs and Laurent Marcoux, for their support. Thank you to Allan Willms for his aid improving this thesis, and thank you to my external examiner, Fuzhen Zhang, for his thoughtful comments and his thought-provoking questions. As well, I would like to express my gratitude to Susan McCormick and Carrie Tanti for their aid with administrative matters, and to Larry Banks for his assistance with all IT-related issues.

I am extremely fortunate to have a wonderful family that has always been there for me. Thank you Mom and Dad for always providing me with a home to return to. Your advice, friendship, and free food have been a constant blessing throughout my schooling. I also want to thank my brothers, Calder and Jim, for always giving me something fun to look forward to. When I start to overwork myself, I need only take a few minutes to plan for our next

"tournament" or recall the most recent exploit of Apocalypse, and I enter a much healthier (and more productive) mindset.

Thank you to my partner, Natalie, for her love and support throughout this process. Early in my doctoral studies, there were many long days that ended in frustration and disappointment. These feelings would vanish as soon as I was home with Natalie. Her friendship is the reason that I start every day happy, refreshed, and excited to test out new ideas. Thank you Natalie. I can't wait to see where life will take us next.

Contents

Chapter 1

Introduction

This thesis is a study of matrix scalings and their applications. In the first half of the dissertation (Chapters 1-3), we will develop some background and discuss the concept of matrix scaling in full generality − covering various types of scalings that have arisen in different areas of mathematics. For the second half of the thesis (Chapters 4-6), we will focus on one type of matrix scaling in particular: *complex, positive definite* scalings. This type of scaling has been shown to have applications to quantum information and, motivated by these applications, we will derive a number of results on these scalings.

We will introduce more specialized terminology and notation when they are needed, but we take this opportunity to remind the reader of some basic matricial definitions. Let us begin with some notation. We will denote the non-negative real ray as $\mathbb{R}_+$, and the strictly positive real numbers as $\mathbb{R}_{++}$. Almost all matrices that we work with will be square, and we denote the set of $n \times n$ matrices with entries from the field $\mathbb{F}$ as $\mathcal{M}_n(\mathbb{F})$. Capital letters

from the Latin alphabet will be reserved for matrices, and the (i,j)th entry of a matrix A $(B,C,...)$ will be denoted a_{ij} $(b_{ij},c_{ij},...)$. Given a matrix $A \in \mathcal{M}_n(\mathbb{C})$, we use $\sigma(A)$ to represent the spectrum of A. (As we will only be working with finite dimensional spaces, the spectrum of a matrix is just the set of its eigenvalues.)

Recall that a matrix is said to be *Hermitian* if it is self-adjoint (i.e. equal to its conjugate transpose).

Definition Let $M \in \mathcal{M}_n(\mathbb{C})$. We say that M is *positive semi-definite* if M is Hermitian, and for all $\lambda \in \sigma(M), \lambda \geq 0$. If this inequality is strict, we say that M is *positive definite*. We denote the set of all $n \times n$ complex positive semi-definite matrices as $\mathcal{PSD}_n$.

We will often write a positive definite matrix M as the sum of its "real part" and its "imaginary part": $M = Re(M) + iIm(M)$, where $Re(M)$ is the real symmetric matrix $Re(M) := \frac{M+M^T}{2}$, and $Im(M)$ is the skew symmetric real matrix $Im(M) := \frac{M-M^T}{2}$. The set of real positive definite matrices are a proper subset of a larger group of matrices, the copositive matrices:

Definition Let $A \in \mathcal{M}_n(\mathbb{R})$. We say that A is a *copositive matrix* if $\langle Ax, x\rangle = x^T Ax \geq 0$ for all non-negative vectors $x \in \mathbb{R}^n_+$. If this equality is strict whenever $x \neq 0$, then A is said to be *strictly copositive*. We denote the set of all $n \times n$ copositive matrices by $\mathcal{COP}_n$.

Another subset of the copositive matrices (and, in fact, subset of the real, positive semi-definite matrices) is the set of completely positive matrices:

Definition Let $A \in \mathcal{M}_n(\mathbb{R})$. We say that A is a *completely positive matrix* if it can be written as BB^T, for some $n \times m$ matrix B with entries in $\mathbb{R}_+$. We will denote the set of $n \times n$ completely positive matrices by $\mathcal{CP}_n$.

Completely positive matrices have been intensely studied for their connection with block designs, exchangeable probability distributions and many other areas [1]. While we will only discuss completely positive matrices briefly in Chapter 2, the interested reader is encouraged to see [1] for a thorough treatment of this set and its applications.

We end our introduction with two more sets of matrices, which will be of some interest to us in Chapter 2:

$$\mathcal{NN}_n := \{A \in M_n(\mathbb{R}) \mid a_{ij} \geq 0 \text{ for all } i, j\}$$

and

$$\mathcal{DNN}_n := \mathcal{PSD}_n \cap \mathcal{NN}_n.$$

One usually refers to elements of $\mathcal{NN}_n$ as non-negative matrices and elements of $\mathcal{DNN}_n$ as doubly non-negative matrices.

1.1 Organization of thesis

This dissertation is divided into seven chapters. For the remainder of the present chapter, we will review some well-known matrix theory results. We will mainly focus our attention on the concept of the *matrix permanent*, and discuss a characterization of the permanent that

allows us to derive a number of important results. We conclude the chapter by defining the *geometric measure of entanglement (GME)* of quantum states, and introducing a theorem of Pereira and Boneng [2] that allows us to calculate the GME by considering the permanent of a certain set of matrices known as *complex matrix scalings*.

In Chapter 2, we examine matrix scalings in generality. We discuss a few different types of scalings that have been investigated by mathematicians over the years, as well as two important existence results that were discovered in the 1960s. After a brief foray into the theory of convex cones, this chapter concludes by collecting the aforementioned existence results, along with that of Pereira (introduced in the previous chapter), into one unifying theorem that proves all three results as corollaries. This theorem is the first novel result introduced herein, though it is essentially a generalization of the arguments found in [3] and [4].

Chapter 3 considers the problem of scaling matrices as a special case of the problem of scaling *tensors* (i.e. multi-dimensional matrices, or "hypermatrices"). We begin by introducing notation and language, including two different ways to extend the concept of *doubly quasi-stochasticity* to tensors. After discussing other authors' results on tensor scalings, we close by proving two results of our own, the former of which proves the existence of a scaling for every real copositive tensor.

The remaining chapters are motivated by Pereira and Boneng's connection between complex, positive definite scalings and the geometric measure of entanglement for symmetric states. In Chapter 4, we investigate the problem of how many ways one can scale a given

positive definite matrix. We begin by providing a 3×3 counter-example to a conjecture made in [2] regarding the maximum number of possible scalings, and then we prove that the true upper bound on scalings of a 3×3 real matrices is 6. We provide conditions on when this upper bound is attained, and then show that no such upper bound exists in general for matrices of dimension 4 or greater.

Chapter 5 investigates matrix scalings that we consider to have "extremal permanent" (known as *maximal* and *minimal* scalings). After introducing necessary notation, we use the proof of our unifying theorem from Chapter 2 to arrive at a characterization of maximal scalings. Using this characterization, we prove certain topological properties of the set of maximal scalings, and arrive at permanental bounds for any maximal scaling. We then prove similar bounds for certain types of minimal scalings: 3×3 minimal scalings, $n \times n$ real minimal scalings, and minimal scalings that are also group matrices. We conclude this chapter by combining these permanental results with the work of Pereira and Boneng to arrive at bounds on the geometric measure of entanglement for certain Slater permanents.

In Chapter 6, we shift gears from completed work to discussing possible directions for future research. Firstly, we build upon an algorithm of O'Leary [5] to introduce an easily implemented algorithm that allows us to calculate multiple complex matrix scalings of positive definite matrices. After discussing the efficacy of this algorithm, we consider a few applications for matrix scalings, including their possible utility towards solving famous open problems.

Chapter 7 serves to briefly recapitulate the results from the above chapters, summarizing

all of the novel results and open questions introduced therein.

1.2 The matrix permanent

We now introduce the quantity that is the focus of Chapter 5. This is the tool that will allow us to derive a connection between matrix scalings and the geometric measure of entanglement later in this chapter (Theorem 1.3.3).

Definition Let A be an $n \times n$ matrix, and let S_n denote the symmetric group on n elements. Then the *permanent* of the matrix A, $per(A)$, is defined as:

$$per(A) = \sum_{\sigma \in S_n} \prod_{k=1}^{n} a_{k\sigma(k)}.$$

Although the matrix permanent is deceptively similar in definition to the determinant (the permanent is only missing the $(-1)^{sgn(\sigma)}$ factor on each summand), it turns out that it is far more difficult to work with. While the determinant has many nice algebraic properties (it is a multiplicative function equal to the product of the eigenvalues of its argument) the permanent does not seem to be nearly so well behaved. As such, an area of intense research (see [6], [7], [8]) has been identifying properties of the permanent and how the permanent behaves under operations such as matrix multiplication. We will begin our discussion with a major breakthrough discovered by Marvin Marcus and Morris Newman in the 1960s. Before we can introduce this result, however, we require some concepts from multilinear algebra.

1.2.1 Introduction to tensors

Much of the information contained in the next few sections is a re-wording of an argument found in [9]. It is included in this dissertation in hopes of providing the reader with a self-contained exposition that is a bit more accessible. To this end, we take this opportunity to include a few definitions and details that were omitted by the authors of [9]. We begin with a reminder of what is meant when one speaks about the tensor product of an n-dimensional Hilbert space with itself. For our purposes, it suffices to consider only the n-dimensional Euclidean space $\mathbb{C}^n$, but the definitions extend to any complex inner product space in the obvious way.

Let $\mathcal{M}_m(\mathbb{C}^n)$ denote the set of multilinear complex functionals acting on $(\mathbb{C}^n)^m$ (i.e. the functions $\phi : \underbrace{\mathbb{C}^n \times \mathbb{C}^n \times ... \times \mathbb{C}^n}_{m \text{ times}} \to \mathbb{C}$ which are linear in each variable). Then we define the associated space of *tensors*, $(\mathbb{C}^n)^{(m)}$, as the space of complex-valued linear functionals acting on $\mathcal{M}_m(\mathbb{C}^n)$ (i.e. $(\mathbb{C}^n)^{(m)} := (\mathcal{M}_m(\mathbb{C}^n))^*$, the complex dual space of $\mathcal{M}_m(\mathbb{C}^n)$). The elements of $(\mathbb{C}^n)^{(m)}$ take one of two forms: decomposable tensors and indecomposable tensors.

Definition Let $f \in (\mathbb{C}^n)^{(m)}$. We say that f is a *decomposable* tensor if there exist vectors $v_1, v_2, ..., v_m \in \mathbb{C}^n$ such that for all $\phi \in \mathcal{M}_m(\mathbb{C}^n)$:

$$f(\phi) = \phi(v_1, v_2, ..., v_m).$$

(In this case, we write $f = v_1 \otimes ... \otimes v_m$.)

A tensor $g \in (\mathbb{C}^n)^{(m)}$ is called *indecomposable* if it is not decomposable.

Remark It is useful to note that any indecomposable tensor can be written as a linear combination of decomposable tensors, whence we see that the space $(\mathbb{C}^n)^{(m)}$ is simply the span of the set of decomposable tensors.

Let us denote the usual Euclidean inner product on $\mathbb{C}^n$ by $\langle\ ,\ \rangle_E$. There is a natural inner product on our tensor space $(\mathbb{C}^n)^{(m)}$. We define it below for only the decomposable tensors, as this will extend to the entire space by linearity.

Definition Let $f = v_1 \otimes ... \otimes v_m$ and $g = w_1 \otimes ... \otimes w_m$ be decomposable tensors in $(\mathbb{C}^n)^{(m)}$. Then the inner product of f and g, denoted $\langle f, g\rangle_\otimes$ is given by:

$$\langle f, g\rangle_\otimes := \prod_{j=1}^{m} \langle v_j, w_j\rangle_E.$$

It is easy to verify that $\langle\ ,\ \rangle_\otimes$ is indeed an inner product on $(\mathbb{C}^n)^{(m)}$. This is, of course, not the only inner product on $(\mathbb{C}^n)^{(m)}$, but it will be the only one that we will use.

We close with an introduction to a type of tensor that will not only be useful for our present discussion, but will re-appear under a slightly different name in Section 1.3, when we discuss quantum states.

Definition Let $f \in (\mathbb{C}^n)^{(m)}$ be a (possibly indecomposable) tensor, i.e. $f = \sum_{j=1}^{k} v_{j1} \otimes ... \otimes v_{jm}$, where $k \in \mathbb{N}$, and $v_{11}, ..., v_{1m}, v_{21}, ..., v_{2m}, ..., v_{k1}, ..., v_{km} \in \mathbb{C}^n$. Then f is called *symmetric* if for all permutations $\sigma \in S_m$, $P_\sigma f := \sum_{j=1}^{k} v_{j\sigma(1)} \otimes ... \otimes v_{j\sigma(m)} = f$. We denote the set of all symmetric elements of $(\mathbb{C}^n)^{(m)}$ as $Sym_{(\mathbb{C}^n)^{(m)}}$.

Note that if a *decomposable* tensor $f = v_1 \otimes ... \otimes v_m$ is symmetric, then it must be the case that $v_1 = v_2 = ... = v_m$.

Now that we have a bit of background into tensors, we are in a position to show how they can help us attack the problems that arise in the study of the matrix permanent.

1.2.2 The matrix permanent as an inner product

In the early 1960s, Marcus and Newman discovered a characterization of the matrix permanent for positive definite matrices that allowed them to prove many important results (see [9]). In this section, we will illustrate the argument that they used to derive this characterization, and use it to prove a few results that will be of use to us in Chapter 5.

First we must introduce an operator that will be our main tool for illustrating the connection between the tensor space $(\mathbb{C}^n)^{(m)}$ and the permanent of complex matrices. This operator is sometimes referred to as the "symmetry operator" on $(\mathbb{C}^n)^{(m)}$, and (again) we define this operator for decomposable tensors, and extend by linearity. Let $T : (\mathbb{C}^n)^{(m)} \to (\mathbb{C}^n)^{(m)}$ be defined as follows:

$$T(v_1 \otimes v_2 \otimes ... \otimes v_m) = \frac{1}{m!} \sum_{\sigma \in S_m} v_{\sigma(1)} \otimes ... \otimes v_{\sigma(m)}.$$

Proposition 1.2.1 *T is an orthogonal projection onto a subset of $Sym_{(\mathbb{C}^n)^{(m)}}$.*

Proof We must show three things:

(1) *range*(T) is a subset of the $Sym_{(\mathbb{C}^n)^{(m)}}$.

(2) T is idempotent.

(3) T is Hermitian (with respect to $\langle\ ,\ \rangle_{\otimes}$).

(1): Let $k \in \mathbb{N}$, $v_{11}, ..., v_{1m}, ..., v_{k1}, ...v_{km} \in \mathbb{C}^n$, and let $f = \sum_{j=1}^{k} v_{j1} \otimes ... \otimes v_{jm} \in (\mathbb{C}^n)^{(m)}$ be a (possibly indecomposable) tensor. We consider the image of f under T:

$$Tf = \frac{1}{m!} \sum_{\sigma \in S_m} \sum_{j=1}^{k} v_{j\sigma(1)} \otimes ... \otimes v_{j\sigma(m)} \in (\mathbb{C}^n)^{(m)}.$$

Now we consider the image of Tf under permutation of its components. Let $\tau \in S_m$.

$$\begin{aligned} P_\tau(Tf) &= \frac{1}{m!} \sum_{\sigma \in S_m} \sum_{j=1}^{k} v_{j\tau(\sigma(1))} \otimes ... \otimes v_{j\tau(\sigma(m))} \\ &= \frac{1}{m!} \sum_{\sigma \in S_m} \sum_{j=1}^{k} v_{j\sigma(1)} \otimes ... \otimes v_{j\sigma(m)} \\ &= Tf, \end{aligned}$$

where the penultimate equality arises from the fact that the left coset τS_m is simply the whole group S_m. Thus Tf is unaffected by permutation of its components, and the image of any element of $(\mathbb{C}^n)^{(m)}$ under T is symmetric.

(2): Let $f = v_1 \otimes ... \otimes v_m$, and consider $T^2 f$:

$$T^2 f = T(\frac{1}{m!} \sum_{\sigma \in S_m} v_{\sigma(1)} \otimes ... \otimes v_{\sigma(m)}).$$

From (1), we know that Tf is symmetric, so permuting the indices will do nothing. Thus:

$$\begin{aligned} T^2 f &= \frac{1}{m!} \sum_{\tau \in S_m} \left(\frac{1}{m!} (\sum_{\sigma \in S_m} v_{\tau\sigma(1)} \otimes ... \otimes v_{\tau\sigma(m)}) \right) \\ &= \left(\frac{1}{m!} \right) (m!) \left(\frac{1}{m!} (\sum_{\sigma \in S_m} v_{\sigma(1)} \otimes ... \otimes v_{\sigma(m)}) \right) \\ &= Tf, \end{aligned}$$

as desired.

(3): We must show that $T = T^*$. Let $g = w_1 \otimes ... \otimes w_m \in (\mathbb{C}^n)^{(m)}$. Then

$$\begin{aligned} \langle Tf, g \rangle_{\otimes} &= \langle \frac{1}{m!} \sum_{\sigma \in S_m} v_{\sigma(1)} \otimes ... \otimes v_{\sigma(m)}, w_1 \otimes ... \otimes w_m \rangle_{\otimes} \\ &= \frac{1}{m!} \sum_{\sigma \in S_m} \prod_{j=1}^{n} \langle v_{\sigma(j)}, w_j \rangle_E \\ &= \frac{1}{m!} \sum_{\sigma \in S_m} \prod_{j=1}^{n} \langle v_j, w_{\sigma(j)} \rangle_E \\ &= \langle f, Tg \rangle_{\otimes} \end{aligned}$$

(where the second-to-last equality follows from the fact that we are summing over all elements of the symmetric group, and the order of summation does not matter). □

Remark The range of T does not comprise all of $Sym_{(\mathbb{C}^n)^{(m)}}$, and some authors refer to $range(T)$ as the *symmetry class of* $Sym_{(\mathbb{C}^n)^{(m)}}$. In quantum information, however, the elements of $range(T)$ are called *Slater permanents*, and these will be the objects of interest

in our work on quantum entanglement (see Section 1.3).

We are almost ready to demonstrate how the matrix permanent can be viewed as an inner product. We will need the following lemma:

Lemma 1.2.2 *Let $\mathcal{Y}$ be a countable set of non-zero vectors in a Hilbert space $\mathcal{H}$. Then there exists a vector $v \in \mathcal{H}$ such that $\langle v, y\rangle \neq 0$ for all $y \in \mathcal{Y}$.*

Proof For each $y \in \mathcal{Y}$, define $\mathcal{N}_y := \{v \in \mathcal{H} : \langle v, y\rangle \neq 0\}$. These sets are open and dense in $\mathcal{H}$, for all $y \in \mathcal{Y}$. To prove our result, it suffices to show that $\bigcap_{y\in\mathcal{Y}} \mathcal{N}_y \neq \emptyset$. Indeed, this is a countable intersection of open dense sets, which (by the Baire Category theorem) must be dense and hence nonempty. □

While Lemma 1.2.2 is the result that we need moving forward, it is worth noting that the conclusion still holds in incomplete spaces, provided $\mathcal{Y}$ is finite.

Lemma 1.2.3 *Let $\mathcal{Y} = \{y_j\}_{j=1}^m$ be some finite set of non-zero vectors in an inner product space $\mathcal{V}$. Then there exists a vector $v \in \mathcal{V}$ such that $\langle v, y_j\rangle \neq 0$ for all $j = 1...m$.*

Proof Suppose not. That is, suppose that there exists a finite set $\mathcal{Y} = \{y_j\}_{j=1}^m$, $m \in \mathbb{N}$ such that for all $v \in \mathcal{V}$ there exists an j such that $\langle v, y_j\rangle = 0$. Further, suppose that $\mathcal{Y}$ is the smallest such set (so no proper subset has this orthogonality property). Now consider the proper subset $\mathcal{A} = \{y_j\}_{j=2}^m$. Then there exists $w \in \mathcal{V}$ such that $\langle w, y\rangle \neq 0$ for all $y \in \mathcal{A}$.

By our assumption on $\mathcal{Y}$, $\langle w, y_1\rangle = 0$. Now define k to be the positive real number

$$k := \begin{cases} 2\frac{\max\limits_{y_j \in \mathcal{A}} |\langle y_1, y_j \rangle|}{\min\limits_{y_j \in \mathcal{A}} |\langle w, y_j \rangle|} & \text{if } \max\limits_{y_j \in \mathcal{A}} |\langle y_1, y_j \rangle| \neq 0, \\ 1 & \text{if } \max\limits_{y_j \in \mathcal{A}} |\langle y_1, y_j \rangle| = 0, \end{cases}$$

and consider the vector $y_1 + kw$. We claim that this vector is not orthogonal to any vector in $\mathcal{Y}$. Indeed, we have $\langle y_1 + kw, y_1 \rangle = \|y_1\|^2 > 0$, so it remains to be shown that $\langle y_1 + kw, y \rangle \neq 0$ for all $y \in \mathcal{A}$.

Suppose that $\langle y_1, y \rangle = 0$ for all $y \in \mathcal{A}$. Then $k = 1$ and $\langle y_1 + kw, y \rangle = \langle w, y \rangle \neq 0$ for all $y \in \mathcal{Y}$, as desired.

Now suppose that $\langle y_1, y \rangle \neq 0$ for some $y \in \mathcal{A}$. Then for any $y \in \mathcal{A}$, we have:

$$\langle y_1 + kw, y \rangle = \langle y_1, y \rangle + k\langle w, y \rangle = \langle y_1, y \rangle + 2\frac{\langle w, y \rangle}{\min\limits_{y_j \in \mathcal{A}} |\langle w, y_j \rangle|} |\max_{y_j \in \mathcal{A}} \langle y_1, y_j \rangle|,$$

which cannot be zero, as the latter term has absolute value greater than or equal to $|2\langle y_1, y \rangle|$. We have seen that none of the elements of $\mathcal{Y}$ are orthogonal to $y_1 + kw$, contradicting our assumption on $\mathcal{Y}$. □

We may now illustrate the connection between symmetric tensors and the matrix permanent. We begin with an important result on the permanent of the product of two matrices.

Theorem 1.2.4 ([9], Theorem 5) *Let A and B be $m \times m$ matrices. Then*

$$|per(AB)|^2 \leq per(AA^*)per(B^*B),$$

with equality if and only if either of the following conditions holds:

(1) A has a zero row, or B has a zero column.

(2) There exists an $m \times m$ diagonal matrix D and an $m \times m$ permutation matrix P such that $A^* = BDP$.

In the proof to follow, the reader is encouraged to notice that our first few steps transform the permanent of AB into an inner product of two symmetric tensors.

Proof Let the transpose of the i-th row of A be denoted a_i, the i-th column of B be denoted b_i, and let e_k denote the vector with k-th entry 1 and zeroes elsewhere. Then

$$
\begin{aligned}
|per(AB)|^2 &= |\sum_{\sigma\in S_m}\prod_{j=1}^{m}(AB)_{j\sigma(j)}|^2 \\
&= |\sum_{\sigma\in S_m}\prod_{j=1}^{m}\langle (AB)e_j, e_{\sigma(j)}\rangle_E|^2 \\
&= |\sum_{\sigma\in S_m}\prod_{j=1}^{m}\langle Be_j, A^*e_{\sigma(j)}\rangle_E|^2 \\
&= |\sum_{\sigma\in S_m}\prod_{j=1}^{m}\langle b_j, \overline{a_{\sigma(j)}}\rangle_E|^2 \\
&= |\sum_{\sigma\in S_m}\langle b_1\otimes ...\otimes b_m, \overline{a_{\sigma(1)}}\otimes ...\otimes \overline{a_{\sigma(m)}}\rangle_\otimes|^2 \\
&= (m!)^2|\langle b_1\otimes ...\otimes b_m, T(\overline{a_1}\otimes ...\otimes \overline{a_m})\rangle_\otimes|^2 \\
&= (m!)^2|\langle T(b_1\otimes ...\otimes b_m), T(\overline{a_1}\otimes ...\otimes \overline{a_m})\rangle_\otimes|^2 \\
&\leq (m!)^2\langle T(b_1\otimes ...\otimes b_m), T(b_1\otimes ...\otimes b_m)\rangle_\otimes\langle T(\overline{a_1}\otimes ...\otimes \overline{a_m}), T(\overline{a_1}\otimes ...\otimes \overline{a_m})\rangle_\otimes \\
&= (m!)^2\langle (b_1\otimes ...\otimes b_m), T(b_1\otimes ...\otimes b_m)\rangle_\otimes\langle (\overline{a_1}\otimes ...\otimes \overline{a_m}), T(\overline{a_1}\otimes ...\otimes \overline{a_m})\rangle_\otimes \\
&= (m!)^2\frac{1}{m!}\left(\sum_{\sigma\in S_m}\prod_{j=1}^{n}\langle b_j, b_{\sigma(j)}\rangle_E\right)\frac{1}{m!}\left(\sum_{\sigma\in S_m}\prod_{j=1}^{n}\langle \overline{a_j}, \overline{a_{\sigma(j)}}\rangle_E\right) \\
&= \sum_{\sigma\in S_m}\prod_{j=1}^{n}\langle Be_j, Be_{\sigma(j)}\rangle_E\sum_{\sigma\in S_m}\prod_{j=1}^{n}\langle A^*e_j, A^*e_{\sigma(j)}\rangle_E \\
&= \sum_{\sigma\in S_m}\prod_{j=1}^{n}\langle B^*Be_j, e_{\sigma(j)}\rangle_E\sum_{\sigma\in S_m}\prod_{j=1}^{n}\langle AA^*e_j, e_{\sigma(j)}\rangle_E \\
&= per(B^*B)per(AA^*),
\end{aligned}
$$

where the inequality arises upon application of the Cauchy-Schwarz inequality.

Now suppose the inequality is, in fact, equality. Then (by the equality condition in the Cauchy-Schwarz inequality) we know that $T(b_1\otimes ...\otimes b_m) = k(T(\overline{a_1}\otimes ...\otimes \overline{a_m}))$, for some

constant k, or $T(\overline{a_1}\otimes...\otimes\overline{a_m}) = 0$. Suppose the former holds and $k = 0$, so $T(b_1\otimes...\otimes b_m) = 0$. Now choose any $v \in \mathbb{C}^m$ and let f be the symmetric decomposable tensor $f = v\otimes v\otimes...\otimes v \in (\mathbb{C}^m)^{(m)}$, and consider $\langle f, T(b_1 \otimes ... \otimes b_m)\rangle = 0$. By the properties of T, we have:

$$\begin{aligned} 0 &= \langle f, T(b_1 \otimes ... \otimes b_m)\rangle_\otimes \\ &= \langle Tf, (b_1 \otimes ... \otimes b_m)\rangle_\otimes \\ &= \langle f, (b_1 \otimes ... \otimes b_m)\rangle_\otimes \\ &= \prod_{i=1}^{m}\langle v, b_i\rangle_E. \end{aligned}$$

As the above equality holds for all $v \in \mathbb{C}^m$, we must have that one (or more) b_i is zero (by Lemma 1.2.2), and hence B has a zero column. We can similarily argue that when $T(\overline{a_1} \otimes ... \otimes \overline{a_m}) = 0$, A must have a zero row.

Now suppose that there exists a $k \neq 0$ such that $T(b_1 \otimes ... \otimes b_m) = kT(\overline{a_1} \otimes ... \otimes \overline{a_m})$, and that $T(\overline{a_1} \otimes ... \otimes \overline{a_m}) \neq 0$. Defining f as above, we have

$$\langle f, T(b_1 \otimes ... \otimes b_m)\rangle_\otimes = \overline{k}\langle f, (T(\overline{a_1} \otimes ... \otimes \overline{a_m}))\rangle_\otimes,$$

which (applying the same steps as above) yields:

$$\prod_{i=1}^{m}\langle v, b_i\rangle_E = \overline{k}\prod_{i=1}^{m}\langle v, \overline{a_i}\rangle_E.$$

Suppose that we chose v to be perpendicular to b_1. Then the above is equal to 0, and

v must be perpendicular to $\overline{a_i}$ for some i. We consider the orthogonal decomposition of $\overline{a_i}$ into $\overline{a_i} = d_i b_1 + w_i$, where $w_i \in span\{b_1\}^{\perp}$ and $d_i \in \mathbb{C}$.

As noted above, for all $v \in span\{b_1\}^{\perp}$, we know there exists an i such that $0 = \langle v, \overline{a_i} \rangle_E = \langle v, w_i \rangle_E$. By Lemma 1.2.2 (where our Hilbert space is $span\{b_1\}^{\perp}$), we see that w_i must be zero for some i. This means that $\overline{a_i} = d_i b_1$ or, after an appropriate permutation, $b_1 = d_1 \overline{a_1}$. Repeating this argument for all columns of B, we see that there must be $D = diag(d_1, ...d_m)$ and an $m \times m$ permutation matrix P such that $A^* = BDP$. □

We have the following well-known corollary:

Corollary 1.2.5 ([10]) *Let A be a positive semi-definite matrix. Then $per(A) \geq 0$.*

Proof Let S be an $n \times n$ matrix satisfying $A = SS^*$. Theorem 1.2.4 yields:

$$|per(SI_n)|^2 \leq per(SS^*)per(I_n I_n) = per(A),$$

and the result follows. □

If we apply Theorem 1.2.4 to the pair of matrices (A, I_n) (where A is a positive semidefinite matrix), then we can see that $per(A) \leq (per(A^2))^{1/2}$, with equality if and only if A has a zero row or is diagonal. Applying this result iteratively, we see that $g(n) = (per(A^{2n}))^{1/2n}$ is a non-decreasing function. It turns out that we can do better. (In what follows, recall that for any positive semi-defininte matrix A with eigendecomposition $A = U^* DU$, and real number $t \in (0, \infty)$, the matrix power A^t is given by $A^t = U^* D^t U$, where D^t is the matrix obtained by raising all entries of $D = (d_{ij})$ to the exponent t, i.e. $D^t = (d_{ij}^t)$.)

Theorem 1.2.6 ([9], Theorem 11) *Let A be an $n \times n$ positive semi-definite matrix that is neither diagonal, nor possesses a zero row. Then*

$$f(t) = (per(A^t))^{1/t}$$

is a strictly increasing function on $t \in (0, \infty)$. Further, when $t \in (0, \infty)$, $f(t) \leq \lambda_1^n$, where λ_1 is the largest eigenvalue of A.

We will need the following definition:

Definition Given a matrix $A \in \mathcal{M}_n(\mathbb{C})$, we define the *$m$-th induced power of A*, $P_m(A)$, to be the function acting on the symmetry class of $Sym_{(\mathbb{C}^n)^{(m)}}$ in the following way:

$$P_m(A)(T(x_1 \otimes ... \otimes x_m)) = T(Ax_1 \otimes ... \otimes Ax_m).$$

It is well known that the eigenvalues of $P_m(A)$ are simply all products $\lambda_{j_1}\lambda_{j_2}...\lambda_{j_m}$, where $\lambda_{j_k} \in \sigma(A)$, and the indices $1 \leq j_k \leq n$ are not necessarily distinct. We will use a result of Ryser relating this operator to the permanent of A:

Lemma 1.2.7 ([11]) *Let $A \in \mathcal{M}_n(\mathbb{C})$ in the standard order basis $\{e_1, ..., e_n\}$. Then*

$$per(A) = \langle P_n(A)(T(e_1 \otimes ... \otimes e_n)), T(e_1 \otimes ... \otimes e_n)\rangle_{\otimes}.$$

Proof of Theorem 1.2.6 We have shown above that $f(1) < f(2)$. For the purposes of

induction, suppose that $f(m-1) < f(m)$, for some $m \in \mathbb{N}$. Then Theorem 1.2.4 yields:

$$|per(A^m)|^2 = |per(A^{\frac{m+1}{2}} A^{\frac{m-1}{2}})|^2 \leq per(A^{m+1}) per(A^{m-1}).$$

This can be rewritten as $(f(m))^{2m} \leq (f(m+1))^{m+1}(f(m-1))^{m-1}$. Applying our inductive hypothesis

$$\begin{aligned} (f(m))^{2m} &\leq f(m+1)^{m+1} f(m-1)^{m-1} \\ &< (f(m+1))^{m+1}(f(m))^{m-1}. \end{aligned}$$

Dividing both sides by $(f(m))^{m-1}$ yields: $(f(m))^{m+1} < (f(m+1))^{m+1}$ and thus

$$f(m) < f(m+1) \text{ for all } m \in \mathbb{N}. \tag{1}$$

We now prove that if r_1, r_2 are positive rational numbes with $r_1 < r_2$, then $f(r_1) < f(r_2)$. Let $r_1 = a/b$ and $r_2 = c/d$, for $a, b, c, d \in \mathbb{N}$. Then we see that $f(r_1) < f(r_2)$ if and only if $(per(A^{\frac{ad}{bd}}))^{\frac{bd}{ad}} < (per(A^{\frac{cb}{bd}}))^{\frac{bd}{cb}}$ if and only if $(per((A^{\frac{1}{bd}})^{ad}))^{\frac{1}{ad}} < (per((A^{\frac{1}{bd}})^{cb}))^{\frac{1}{cb}}$, which must be true, as (re-defining f accordingly) (1) holds for the positive semi-definite matrix $A^{\frac{1}{bd}}$ and $ad < cb$ (and $ad, cb \in \mathbb{N}$). Extending to positive irrational numbers by continuity, we have that $f(t)$ is monotone increasing on $(0, \infty)$.

To obtain our upper bound on $f(t)$, we note that $(f(t))^t = per(A^t)$ and, by Lemma 1.2.7,

$per(A^t) = \langle P_n(A^t)(T(e_1 \otimes ... \otimes e_n), T(e_1 \otimes ... \otimes e_n)\rangle_{\otimes}$. Now, this is of the form $\langle P_n(A^t)v, v\rangle_{\otimes}$ and hence must be bounded above by the largest eigenvalue of $P_n(A^t)$. As the eigenvalues of $P_n(A^t)$ are given by all products of the eigenvalues of A^t (as described above Lemma 1.2.7), the largest eigenvalue of $P_n(A^t)$ is simply the product of the largest eigenvalue, λ_1^t, with itself n times. We conclude that $(f(t))^t \leq (\lambda_1^t)^n$, whence our result follows. □

Remark In fact, there is an even stronger result proven in [9], where it is shown that $\lim_{t\to 0^+} f(t) = det(A)$, and $\lim_{n\to\infty} f(n) = \lambda_1^n$.

We close this section with one last result that this inner product representation of the permanent allows us to prove. Until the early 1980s (when it was proven in [12] and [13], simultaneously), much of the research done on the permanent was in attempts to prove the famous *van der Waerden inequality*, conjectured by Bartel Leendert van der Waerden in 1926 [14]. Before we can state the inequality, we first introduce a set of matrices often referred to as the *Birkhoff polytope*:

Definition Let $S \in \mathcal{M}_n(\mathbb{R}_{++})$ be a real matrix with all entries non-negative. We say that S is *doubly stochastic* if the entries in any given row or column sum to 1. We denote the set of all $n \times n$ doubly stochastic matrices by Ω_n.

Theorem 1.2.8 ([12], [13], The van der Waerden inequality) *Let $S \in \Omega_n$. Then $per(S) \geq \frac{n!}{n^n}$, with equality if and only if $S = J_n$, the matrix with all entries equal to $\frac{1}{n}$.*

Our inner product characterization of the permanent allows us to prove this inequality in the case of positive semi-definite doubly stochastic matrices.

Theorem 1.2.9 ([9], Theorem 6) *Let $A \in \Omega_n$ be positive semidefinite. Then $per(A) \geq \frac{n!}{n^n}$, with equality if and only if $A = J_n$.*

Proof As S is positive semi-definite, there exist vectors $x_1, ...x_n \in \mathbb{C}^n$ such that $a_{ij} = \langle x_i, x_j \rangle_E$.

$$\begin{aligned} per(A) &= \sum_{\sigma \in S_n} \prod_{j=1}^{n} a_{j\sigma(j)} \\ &= \sum_{\sigma \in S_n} \prod_{j=1}^{n} \langle x_j, x_{\sigma(j)} \rangle_E \\ &= \sum_{\sigma \in S_n} \langle x_1 \otimes ... \otimes x_n, x_{\sigma(1)} \otimes ... \otimes x_{\sigma(n)} \rangle_{\otimes} \\ &= n! \langle x_1 \otimes ... \otimes x_n, T(x_1 \otimes ... \otimes x_n) \rangle_{\otimes} \\ &= n! \| T(x_1 \otimes ... \otimes x_n) \|_{\otimes}^2, \end{aligned} \tag{2}$$

where the last equality comes from the idempotency and Hermiticity of T.

Define the vector $v := \sum_{j=1}^{n} x_j$, and consider $\|v\|_E^2$.

$$\begin{aligned} \|v\|_E^2 &= \langle \sum_{i=1}^{n} x_i, \sum_{j=1}^{n} x_j \rangle_E \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle x_i, x_j \rangle_E \\ &= n, \end{aligned}$$

(where the last equality arises from the fact that A is doubly-stochastic and $\sum_{j=1}^{n}\langle x_i, x_j\rangle_E$ is just the i-th row sum). The tensor $w := \frac{v}{\sqrt{n}} \otimes \frac{v}{\sqrt{n}} \otimes ... \otimes \frac{v}{\sqrt{n}}$ must then have norm 1, and we have $|\langle T(x_1 \otimes ... \otimes x_n), w\rangle_{\otimes}| \leq \|T(x_1 \otimes ... \otimes x_n)\|_{\otimes}$ (as $\|w\|_{\otimes} = 1$) by the Cauchy-Schwarz-inequality. Applying this to (2) above, we have:

$$\begin{aligned} per(A) &\geq n!|\langle T(x_1 \otimes ... \otimes x_n), w\rangle_{\otimes}|^2 \\ &= n!|\langle (x_1 \otimes ... \otimes x_n), T(\frac{v}{\sqrt{n}} \otimes \frac{v}{\sqrt{n}} \otimes ... \otimes \frac{v}{\sqrt{n}})\rangle_{\otimes}|^2 \\ &= \frac{n!}{n^n}|\langle x_1 \otimes ... \otimes x_n, T(v \otimes v \otimes ... \otimes v)\rangle|^2 \\ &= \frac{n!}{n^n}\prod_{j=1}^{n}|\langle x_j, v\rangle|_E^2. \end{aligned}$$

Noting again that $\langle x_j, v\rangle_E = \sum_{i=1}^{n}\langle x_j, x_i\rangle_E$ is just the jth row sum, we see that our product is equal to 1, and we arrive at the desired inequality.

By the equality conditions of the Cauchy-Schwarz inequality, the above inequality is equality precisely when $T(x_1 \otimes ... \otimes x_n) = kw$, for some constant k. Following the same reasoning as the discussion of equality in Theorem 1.2.4, we see that we must have $x_1 \otimes ... \otimes x_n = 0$ (which can't be the case, as A is doubly stochastic) or for all i, $x_i = lv$ for some nonzero constant l. But this means that all x_i are multiples of the same vector, and thus A has rank one.

If we have that A is a real, rank-one symmetric matrix, then we can write it as zz^T, where z is the normalized eigenvector of A that does not vanish under A. As A is doubly

stochastic, it must certainly have the (eigenvalue,eigenvector)-pair $\left(1, (\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}})^T\right)$. Thus $z = (\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}})^T$, and $A = J_n$, as desired. □

The inner product characterization of Marcus and Newman was used prove many more permanental properties of positive semi-definite matrices. The results in our next section are no exception, but to give the reader a sense of some other techniques for working with the permanent, we will prove the next set of results using more elementary methods.

1.2.3 A few more results on the permanent

No treatment of the permanent is complete without the pair of inequalities that we introduce in this section. These two results will conclude our introduction to the permanent, and they will be very useful to us for our investigations in Chapter 5. We will include proofs in hopes of making the reasoning a bit clearer than it was in the original papers. The first of these inequalities that we will discuss is the so-called *the Hadamard permanental inequality*, and the second inequality will follow as Corollary 1.2.13. We first remind the reader of the famous determinantal analogue, originally appearing in [15]:

Proposition 1.2.10 (Hadamard determinant inequality) *Let A be a positive definite matrix. Then $det(A) \leq \prod_{j=1}^{n} a_{jj}$.*

It turns out that the opposite inequality is true for the permanent:

Theorem 1.2.11 (The Hadamard permanent inequality, [16]) *Let A be an $n \times n$ positive definite matrix. Then $per(A) \geq \prod_{j=1}^{n} a_{jj}$, the product of the diagonal entries of A. This*

inequality becomes equality if and only if A is a diagonal matrix, or A has a zero row.

While this was originally proven by Marcus, using the tensor arguments from the previous section, we will actually elaborate on a proof of Bapat [17]. This method of proof is more concise and uses only elementary arguments.

We will need the following well-known result (see [18] for a proof of this theorem), the analogue of a more famous theorem concerning determinants.

Theorem 1.2.12 (Binet-Cauchy Formula for Permanents) *Let A, B be $m \times n$ and $n \times m$ matrices, respectively. Then the permanent of the product, $per(AB)$ is given by:*

$$per(AB) = \sum_{S \in \mathcal{S}_{m,n}} \frac{1}{\mu(S)} per(A_{[S]}) per(B_{<S>}),$$

where $\mathcal{S}_{m,n}$ is the set of all m-tuples $\{k_1, ..., k_m\} \in \mathbb{Z}^m$ satisfying $k_j \geq 0$ for all j and $k_1 + k_2 + ... + k_m = n$; $\mu(S) = \frac{1}{k_1! ... k_m!}$; $A_{[S]}$ is the $n \times n$ matrix given by taking k_j copies of the j-th column of A; and $B_{<S>}$ is the matrix given by taking k_j copies of the j-th row of B.

We may now prove the Hadamard permanent inequality. Following the argument in [17], we will actually prove a stronger inequality first, and our lemma will follow.

Proof of Theorem 1.2.11 We begin with some notation. Let A be our $n \times n$ positive definite matrix, let A_{ij} denote the i,j-th minor of A (i.e. the $(n-1) \times (n-1)$ principal submatrix of A obtained by deleting the i-th row and j-th column of A) and let M be the Schur complement of the (1,1)-entry of A (so $M = A_{11} - \frac{1}{a_{11}} xx^*$, where $x = (a_{21} \ ... \ a_{n1})^T$).

Following the argument from [17], we will show that $per(A) \geq per(A_{11}) - per(M)$. To this end, we examine $per(A)$, $per(A_{11})$, and $per(M)$, separately.

***per*(*A*) :** Let $A = LL^*$ be the Cholesky decomposition of A (so L is lower triangular). Noting that $L^*_{<S>} = (L_{[S]})^*$, we see that

$$\begin{aligned} per(A) = per(LL^*) &= \sum_{S \in \mathcal{S}_{n,n}} \frac{1}{\mu(S)} per(L_{[S]}) per(L^*_{<S>}) \\ &= \sum_{S \in \mathcal{S}_{n,n}} \frac{1}{\mu(S)} |per(L_{[S]})|^2. \end{aligned}$$

Further, as $L_{1j} = 0$ for all $j \neq 1$, wee see that $L_{[S]}$ has a row of zeroes whenever $k_1 = 0$. This means that when $k_1 = 0$, $per(L_{[S]}) = 0$ and the summand vanishes. It suffices, then, to sum over all elements of $\mathcal{S}_{n,n}$ that have $k_1 > 0$. Consider the form of $L_{[S]}$, when $k_1 > 0$:

$$L_{[S]} = \begin{pmatrix} l_{11} & \dots & l_{11} & 0 & \dots & 0 \\ l_{21} & \dots & l_{21} & & & \\ \vdots & \dots & \vdots & & * & \\ l_{n1} & \dots & l_{n1} & & & \end{pmatrix},$$

where the first column has been repeated k_1 times. We proceed by Laplace expansion along the first row:

$$\begin{aligned} per(L_{[S]}) &= l_{11} per((L_{[S]})_{11}) + l_{11} per((L_{[S]})_{12}) + \dots + l_{11} per((L_{[S]})_{1k_1}) + 0 + \dots + 0 \\ &= k_1 l_{11} per((L_{[S]})_{11}), \end{aligned}$$

where the latter equality follows from the fact that $(L_{[S]})_{11} = (L_{[S]})_{12} = ... = (L_{[S]})_{1k_1}$.

We now substitute this into the equation for $per(A)$. Note that if $k_1 > 1$ and $k_2 > 0$, the first and second columns of $(L_{[S]})_{11}$ are the same. Using this fact, we have:

$$\begin{aligned}
per(A) &= \sum_{S\in\mathcal{S}_{n,n}} \frac{1}{\mu(S)} |per(L_{[S]})|^2 \\
&= \sum_{\substack{S\in\mathcal{S}_{n,n} \\ k_1\neq 0}} \frac{1}{\mu(S)} |k_1 l_{11} per((L_{[0,k_1+k_2,k_3,\ldots,k_n]})_{11})|^2 \\
&= \sum_{\substack{S\in\mathcal{S}_{n,n} \\ k_1=1}} \frac{1}{\mu(S)} a_{11} |per((L_{[0,k_2+1,k_3,\ldots k_n]})_{11})|^2 + 2 \sum_{\substack{S\in\mathcal{S}_{n,n} \\ k_1=2}} \frac{2}{\mu(S)} a_{11} |per((L_{[0,2+k_2,k_3,\ldots k_n]})_{11})|^2 \\
&\quad + \ldots + (n-1) \sum_{\substack{S\in\mathcal{S}_{n,n} \\ k_1=n-1}} \frac{n-1}{\mu(S)} a_{11} |per((L_{[0,n-1+k_2,k_3,\ldots k_n]})_{11})|^2 \\
&\quad + n \frac{n}{\mu(S)} a_{11} |per((L_{[0,n,0,\ldots]})_{11})|^2 \\
&= \sum_{j=1}^{n} j \left(\sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=j-1}} \frac{1}{\mu(S)} a_{11} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]})_{11})|^2 \right). \qquad (1)
\end{aligned}$$

Note that in the last step, we have changed our indexing set $\mathcal{S}_{n,n}$ to $\mathcal{S}_{n-1,n}$, and lost the numerator (k_1) above $\mu(S)$. This follows as we have ostensibly replaced k_1 with $k_1 - 1$. In doing so, we now have $k_1 + ... + k_n = n - 1$, and we have made use of the fact that $\frac{k_1}{\mu(S)} = \frac{1}{(k_1-1)!k_2!\ldots k_n!}$.

***per*(A_{11}) :** Let L_1 denote the $(n-1) \times n$ matrix obtained by removing the 1st row of L. Then it is easy to verify that $A_{11} = L_1 L_1^*$. Thus, by the Binet-Cauchy Theorem, we have:

$$\begin{aligned} per(A_{11}) &= \sum_{S\in\mathcal{S}_{n-1,n}} \frac{1}{\mu(S)} per((L_1)_{[S]}) per((L_1)^*_{<S>}) \\ &= \sum_{S\in\mathcal{S}_{n-1,n}} \frac{1}{\mu(S)} |per((L_1)_{[S]})|^2. \end{aligned}$$

Using the facts that the first two columns of L_1 are identical and equal to the first column of $(L_{[S]})_{11}$ where $k_1 + k_2 > 1$)

$$\begin{aligned} per(A_{11}) &= \sum_{S\in\mathcal{S}_{n-1,n}} \frac{1}{\mu(S)} |per((L_1)_{[S]})|^2 \\ &= \sum_{S\in\mathcal{S}_{n-1,n}} \frac{1}{\mu(S)} |per((L_1)_{[0,k_1+k_2,k_3,\ldots,k_n]})|^2 \\ &= \sum_{S\in\mathcal{S}_{n-1,n}} \frac{1}{\mu(S)} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]}))_{11}|^2. \end{aligned} \tag{2}$$

***per*(M) :** Consider the i,j-th element of M:

$$\begin{aligned} m_{ij} &= a_{i+1j+1} - \frac{1}{a_{11}} a_{(i+1)1} a_{1(j+1)} \\ &= a_{i+1j+1} - \frac{1}{a_{11}} a_{(i+1)1} \overline{a_{(j+1)1}} \\ &= \sum_{k=1}^{n} l_{(i+1)k} \overline{l_{(j+1)k}} - \frac{1}{a_{11}} \left(\sum_{s=1}^{n} l_{(i+1)s} \overline{l_{1s}} \right) \left(\sum_{t=1}^{n} \overline{l_{(j+1)t}} l_{1t} \right). \end{aligned}$$

But l_{1s} (l_{1t}) is 0 unless $s = 1$ ($t = 1$), and $l_{11}^2 = a_{11}$, so:

$$\begin{aligned} m_{ij} &= \sum_{k=1}^{n} l_{(i+1)k}\overline{l_{(j+1)k}} - \frac{1}{a_{11}}(a_{11}l_{(i+1)1}\overline{l_{(j+1)1}}) \\ &= \sum_{k=2}^{n} l_{(i+1)k}\overline{l_{(j+1)k}} \\ &= \sum_{k=1}^{n-1} l_{i+1k+1}\overline{l_{j+1k+1}}, \end{aligned}$$

which is the (i, j)-th entry of $L_{11}L_{11}^*$. Thus we have $M = L_{11}L_{11}^*$, and the Binet-Cauchy Theorem yields:

$$\begin{aligned} per(M) &= \sum_{S\in\mathcal{S}_{n-1,n-1}} \frac{1}{\mu(S)}|per((L_{11})_{[S]})|^2 \\ &= \sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=0}} \frac{1}{\mu(S)}|per((L_{[0,k_1+k_2+1,\ldots,k_n]})_{11})|^2. \end{aligned} \tag{3}$$

(where we have written it in terms of L_{11} and introduced a variable k_1 that is always zero to be able to sum over $\mathcal{S}_{n-1,n}$ rather than $\mathcal{S}_{n-1,n-1}$, without changing the value of the sum).

Now we combine (1), (2) and (3) to obtain our desired result. (1) yields:

$$
\begin{aligned}
per(A) &= \sum_{j=1}^{n} j \left(\sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=j-1}} \frac{1}{\mu(S)} a_{11} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]})_{11})|^2 \right) \\
&\geq a_{11} \sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=0}} \frac{1}{\mu(S)} |per((L_{[0,1+k_2,k_3,\ldots k_n]})_{11})|^2 \\
&\quad + \sum_{j=2}^{n} j \left(\sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=j-1}} \frac{1}{\mu(S)} a_{11} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]})_{11})|^2 \right) \\
&= a_{11} per(M) + \sum_{j=2}^{n} j \left(\sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=j-1}} \frac{1}{\mu(S)} a_{11} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]})_{11})|^2 \right). \qquad (4)
\end{aligned}
$$

Further,

$$
\begin{aligned}
per(A_{11}) - per(M) &= \sum_{S\in\mathcal{S}_{n-1,n}} \frac{1}{\mu(S)} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]}))_{11}|^2 \\
&\quad - \sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1=0}} \frac{1}{\mu(S)} |per((L_{[0,k_1+k_2+1,k_3,\ldots,k_n]})_{11})|^2 \\
&= \sum_{\substack{S\in\mathcal{S}_{n-1,n} \\ k_1\neq 0}} \frac{1}{\mu(S)} |per((L_{[0,k_1+k_2+1,k_3,\ldots k_n]}))_{11}|^2.
\end{aligned}
$$

Compare this with the second term of (4), and we see that

$$\begin{aligned} per(A) &\geq a_{11}per(M) + 2a_{11}(per(A_{11}) - per(M)) \\ &= 2a_{11}per(A_{11}) - a_{11}per(M) \\ &= a_{11}per(A_{11}) + a_{11}(per(A_{11}) - per(M)) \\ &\geq a_{11}per(A_{11}), \end{aligned}$$

where the last inequality follows from observing that our above expression for $per(A_{11}) - per(M)$ is always nonnegative.

We have seen that $per(A) \geq a_{11}per(A_{11})$. Apply the same argument to the matrix A_{11}, and we get $per(A) \geq a_{11}a_{22}per(A_{22})$. Iterating $n-2$ more times completes the proof. □

The Hadamard determinant inequality and the Hadamard permanent inequality combine to yield the following.

Corollary 1.2.13 *Let A be a positive definite matrix. Then $per(A) \geq det(A)$.*

Remark Theorem 1.2.11 and Corollary 1.2.13 are, in fact, true for all positive semi-definite matrices (as shown in [16]). This can be shown by a similar argument as above, though we must be careful to account for the case where $a_{ii} = 0$.

This concludes our discussion of the historical results on the matrix permanent. We now illustrate how, by considering entities known as "complex matrix scalings", we can use

the permanent to calculate an important quantity in quantum information: the geometric measure of entanglement.

1.3 The connection between the matrix permanent and the geometric measure of entanglement

As mentioned in the introduction, our interest in matrix scalings and their permanent arose from an application to quantum information; a connection discovered by Pereira and Boneng in [2]. We take this opportunity to walk the reader through this result. This section is essentially a summary of the main argument found in Section 4 of [2]. As this is a foray into the land of quantum information, we will adopt the associated notation and language. For the benefit of the reader, we provide a brief rundown of the terms associated with quantum information (the reader looking for a proper discussion of the mathematical formulation of quantum mechanics should see [19]).

In quantum information, we refer to (normalized) column vectors as *pure states*. We denote a pure state by $|\phi\rangle$, and the conjugate transpose of this vector by $\langle\phi|$. *The inner product of two pure states* $|\phi\rangle$, $|\psi\rangle$ is denoted $\langle\phi|\psi\rangle$.

All states that we discuss are assumed to be pure states (we will drop the "pure" at times for brevity's sake).

Definition Given two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, a pure state $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is called a *separable state* (or *product state*) if it can be written as the tensor product of two pure

states, $\phi_A \otimes \phi_B$, where $|\phi_A\rangle \in \mathcal{H}_A$, and $|\phi_B\rangle \in \mathcal{H}_B$.

Remark (1) Inductively, the above definition extends to a state that is the tensor product of finitely many pure states.

(2) Separable states are precisely the set of decomposable tensors, as defined in Section 1.2.1 (though now we allow the components to arise from different Hilbert spaces).

Recall from our discussion in Section 1.2 that the tensor product of two Hilbert spaces $\mathcal{H}_\mathcal{A}, \mathcal{H}_\mathcal{B}$ is the (closed) linear span of all states of the form $|\phi_A\rangle \otimes |\phi_B\rangle$. Thus one can think of the pure separable states as the states that "generate" $\mathcal{H}_A \otimes \mathcal{H}_B$. We say that such states are *unentangled*, and it is the process of taking these linear combinations that introduces entanglement. An *entangled* pure state is a pure state that is not separable (i.e. an indecomposable tensor).

Example *Let $|\phi\rangle_1, |\phi\rangle_2 \in \mathcal{H}_\mathcal{A}$, and $|\psi\rangle_1, |\psi\rangle_2 \in \mathcal{H}_\mathcal{B}$ be states in their respective Hilbert spaces, with $|\phi\rangle_1 \neq |\phi\rangle_2$ and $|\psi\rangle_1 \neq |\psi\rangle_2$. Then the state given by $\frac{1}{\sqrt{2}}(\phi_1 \otimes \psi_1 + \phi_2 \otimes \psi_2)$ cannot be written as $|\phi\rangle_A \otimes |\phi\rangle_B$ where $|\phi\rangle_A \in \mathcal{H}_A, |\phi\rangle_B \in \mathcal{H}_B$. Thus, this state is entangled.*

In physical terms, a system is said to be entangled if measuring one component of the system affects the other components. Identifying the extent to which a given state's components are affected in this way (i.e. how entangled the state is) is a problem of utmost importance in quantum information.

One method of quantifying the entanglement of a state was introduced by Shimony in [20]. He showed that we could express the entanglement of a state by examining the

distance between that state and the set of separable states (the states that experience no entanglement). This "geometric measure of entanglement" was a little difficult to work with until Wei and Goldbart [21] found a way to express this same idea in a manner that is a bit easier to calculate. This yielded the following formula:

Definition Let $\{\mathcal{H}_k\}_{k=1}^n$ be complex Hilbert spaces and let *Sep* denote the set of unit length separable states in $\mathcal{K} = \bigotimes_{k=1}^{n} H_k$. Let $|\phi\rangle \in \mathcal{K}$ be any element of unit length. Then the *geometric measure of entanglement (GME)* of $|\phi\rangle, E(|\phi\rangle)$, is defined as

$$E(|\phi\rangle) = 1 - \max_{|\psi\rangle \in Sep} |\langle\phi|\psi\rangle|^2.$$

The geometric measure of entanglement has been applied to a multitude of fields, such as the study of entanglement witnesses [22], LOCC state discrimination [23], and phase transitions in spin models [24].

We will be considering the GME of states known as *Slater permanents*. As is more common in quantum information, we will denote the tensor product of a Hilbert space $\mathcal{H}$ with itself n times as $\mathcal{H}^{\otimes n}$ (rather than $(\mathcal{H})^{(n)}$, as we did in Section 1.2).

Definition Let $\mathcal{K} = \mathcal{H}^{\otimes n}$ be the tensor product of a complex Hilbert space with itself n times. A state $|\phi\rangle \in \mathcal{K}$ (satisfying $\langle\phi|\phi\rangle = 1$) is said to be a *Slater permanent* if it can be written as the tensor product

$$|\phi\rangle = \frac{1}{n!} \sum_{\sigma \in S_n} \bigotimes_{k=1}^{n} |\phi_{\sigma(k)}\rangle, \text{ where } \phi_k \in \mathcal{H} \text{ for all } k = 1...n.$$

Remark This definition should look familiar. These are precisely the tensors that we called the "symmetry class of $Sym_{(\mathbb{C}^n)^{(m)}}$" in Section 1.2.2, and they arose as the range of our operator T.

It was shown in [25] that when calculating the geometric measure of entanglement of a Slater permanent, we only need to consider the separable states that are themselves Slater permanents. As mentioned in Section 1.2, a separable Slater permanent in $\mathcal{H}^{\otimes n}$ must be of the form $\frac{1}{n!}\bigotimes_{k=1}^{n}|\psi\rangle$, for some $|\psi\rangle \in \mathcal{H}$ This means that for a given Slater permanent $|\phi\rangle = \sum_{\sigma\in S_n}\bigotimes_{k=1}^{n}|\phi_{\sigma(k)}\rangle \in \mathcal{H}^{\otimes n}$,

$$E(|\phi\rangle) = 1 - \max_{\substack{|\psi\rangle\in\mathcal{H}\\ \langle\psi|\psi\rangle=1}} |\langle\phi|\bigotimes_{k=1}^{n}\psi\rangle|^2 = 1 - \max_{\substack{|\psi\rangle\in\mathcal{H}\\ \langle\psi|\psi\rangle=1}} \frac{1}{n!}\sum_{\sigma\in S_n}\prod_{k=1}^{n}|\langle\phi_{\sigma(k)}|\psi\rangle|^2.$$

Now let us fix $|\phi\rangle$ for the moment and define $B_{|\phi\rangle}$ as the matrix whose ith column is $|\phi_i\rangle$. Further, suppose that $B_{|\phi\rangle}$ is invertible. Then $\langle\psi|B_{|\phi\rangle} = (\langle\psi|\phi_1\rangle, \langle\psi|\phi_2\rangle, ..., \langle\psi|\phi_n\rangle)$. Let this vector matrix product $\langle\psi|B_{|\phi\rangle} =: v_\psi$. Then $\prod_{k=1}^{n}\langle\psi|\phi_{\sigma(k)}\rangle = \prod_{k=1}^{n}(v_\psi)_k$, and $\langle\psi|\psi\rangle = 1$ if and only if $\langle v_\psi B_{|\phi\rangle}^{-1}|v_\psi B_{|\phi\rangle}^{-1}\rangle = 1$. Thus, given $|\phi\rangle$, we now have following expression for the geometric measure of entanglement of $|\phi\rangle$:

$$E(|\phi\rangle) = 1 - \max_{\substack{v\in\mathbb{C}^n\\ \langle vB_{|\phi\rangle}^{-1}|vB_{|\phi\rangle}^{-1}\rangle=1}} \frac{1}{n!}\sum_{\sigma\in S_n}\prod_{i=1}^{n}|v_i|^2 = 1 - \max_{\substack{v\in\mathbb{C}^n\\ \langle vB_{|\phi\rangle}^{-1}|vB_{|\phi\rangle}^{-1}\rangle=1}} \prod_{i=1}^{n}|v_i|^2.$$

From this point forward, we will assume that $\mathcal{H} = \mathbb{C}^n$. In this case, $\langle vB_{|\phi\rangle}^{-1}|vB_{|\phi\rangle}^{-1}\rangle = v(B_{|\phi\rangle})^{-1}(B_{|\phi\rangle}^*)^{-1}v^*$. Let $(G_{|\phi\rangle})^{-1} := (B_{|\phi\rangle})^{-1}(B_{|\phi\rangle}^*)^{-1}$. Then our problem reduces to maximizing $\prod_{i=1}^{n}|v_i|^2$ over all vectors satisfying $v(G_{|\phi\rangle})^{-1}v^* = 1$.

We now introduce a result from [2], the proof of which we will not state here, as it will actually follow from Theorem 2.4.1 in the next Chapter.

Theorem 1.3.1 ([2], Corollary 2.1) *Let A be an $n \times n$ positive definite matrix and let v be the vector that maximizes $g(v) = \prod_{i=1}^{n} |v_i|^2$, over all vectors satisfying $vAv^* = 1$. If we define the diagonal matrix $D = \sqrt{n} diag(v_1, ..., v_n)$, then DAD^* has all row and column sums equal to 1.*

We have a special name for such a matrix.

Definition Let A be a complex $n \times n$ matrix with all row and column sums equal to 1. Then A is said to be *doubly quasi-stochastic (DQS)*.

Theorem 1.3.1 requires that A be a positive definite matrix. It is clear that $G_{|\phi\rangle}$ is a positive definite matrix, and in fact is a special type of matrix (the following definition agrees with our above construction of $G_{|\phi\rangle}$).

Definition Let $\mathcal{H}$ be a complex Hilbert space and let $|\phi\rangle \in \mathcal{H}^{\otimes n}$ be the Slater permanent $|\phi\rangle = \frac{1}{n!} \sum_{\sigma \in S_n} \bigotimes_{k=1}^{n} |\phi_{\sigma(k)}\rangle$. We define the Gram matrix of $|\phi\rangle$ to be the $n \times n$ matrix $G_\phi = (g_{ij})$, where $g_{ij} = \langle \phi_i | \phi_j \rangle$.

Now, note that for any diagonal matrix $D = diag(v)$, $per(D) = \prod_{i=1}^{n} (v_i)$ and the permanent of $DG_{|\phi\rangle}D^* = \prod_{i=1}^{n} |v_i|^2 G_{|\phi\rangle}$. Using Theorem 1.3.1, this once again gives us a new expression for the GME of $|\phi\rangle$.

$$\begin{aligned} E(|\phi\rangle) &= 1 - \max_{\substack{v\in\mathbb{C}^n \\ v(G_{|\phi\rangle})^{-1}v^*=1}} \prod_{i=1}^{n} |v_i|^2 \\ &= 1 - \max_{\substack{D\in\mathcal{M}_n(\mathbb{C}), D \text{ diagonal} \\ \sqrt{n}D(G_{|\phi\rangle})^{-1}\sqrt{n}D^* \text{ is DQS}}} |per(D)|^2 \\ &= 1 - \max_{\substack{D\in\mathcal{M}_n(\mathbb{C}), D \text{ diagonal} \\ \sqrt{n}D(G_{|\phi\rangle})^{-1}\sqrt{n}D^* \text{ is DQS}}} \frac{per(D^*G_{|\phi\rangle}^{-1}D)}{per(G_{|\phi\rangle}^{-1})}. \end{aligned}$$

Unfortunately, we don't know what the permanent of $G_{|\phi\rangle}^{-1}$ is. However, it is relatively easy to see that $per(G_{|\phi\rangle}) = n!\langle\phi|\phi\rangle = n!$. For this reason, we would like to get our above expression in terms of $per(G_{|\phi\rangle})$. The following lemma will allow us to do this:

Lemma 1.3.2 *Let A be a positive definite matrix and suppose that D^*AD is DQS for some diagonal matrix D. Then $D^{-1}A^{-1}(D^*)^{-1}$ is also DQS.*

Proof As D^*AD is positive definite, D^*AD is DQS if and only if $D^*ADe = e$, where $e = (1, 1...1)^T$. Taking inverses completes the proof. □

We adopt the following notation:

$$sc(G_{|\phi\rangle}) = \{B \in \mathcal{M}_n(\mathbb{C}) : B \text{ is DQS and } B = D^*G_{|\phi\rangle}D \text{ for some diagonal matrix } D\}.$$

We can now state the theorem of Pereira and Boneng, which will serve as motivation for much of the remainder of this dissertation.

Theorem 1.3.3 ([2], Theorem 4.2) *Let $|\phi\rangle$ be a Slater permanent in $((\mathbb{C})^n)^{\otimes n}$ whose*

Gram matrix $G_{|\phi\rangle}$ is invertible. Let M be the matrix in $sc(G_{|\phi\rangle})$ that has smallest permanent. Then

$$E(|\phi\rangle) = 1 - \frac{n!}{(n^n)per(M)}.$$

Proof We have seen that

$$E(|\phi\rangle) = 1 - \max_{\substack{D\in\mathcal{M}_n(\mathbb{C}),D \text{ diagonal} \\ \sqrt{n}D(G_{|\phi\rangle})^{-1}\sqrt{n}D^* \text{ is DQS}}} \left(\frac{per(D^*G_{|\phi\rangle}^{-1}D)}{per(G_{|\phi\rangle})}\right).$$

By Lemma 1.3.2, this is equivalent to

$$\begin{aligned} E(|\phi\rangle) &= 1 - \max_{\substack{D\in\mathcal{M}_n(\mathbb{C}),D \text{ diagonal} \\ \frac{1}{\sqrt{n}}D(G_{|\phi\rangle})\frac{1}{\sqrt{n}}D^* \text{ is DQS}}} \left(\frac{per(D^*G_{|\phi\rangle}D)}{per(G_{|\phi\rangle})}\right)^{-1} \\ &= 1 - \max_{B\in sc(G_{|\phi\rangle})} \left(\frac{n^n per(B)}{per(G_{|\phi\rangle})}\right)^{-1} \\ &= 1 - \frac{n!}{n^n per(M)}. \end{aligned}$$

□

We have just seen that one can calculate the geometric measure of entanglement by finding all diagonal matrices D such that $D^*G_{|\phi\rangle}D$ is doubly quasi-stochastic. If this relation holds, we say that $D^*G_{|\phi\rangle}D$ is a *complex scaling* of $G_{|\phi\rangle}$, or (equivalently) that D *scales* $G_{|\phi\rangle}$. The search for such matrices will comprise the majority of the remainder of this thesis. It is important to note that for any positive definite matrix A, there is always at least one complex scaling of A. This was originally shown in [3]:

Theorem 1.3.4 ([3], Lemma 2.9) *Let A be a positive definite matrix, then there exists a diagonal matrix D such that D^*AD is a doubly quasi-stochastic matrix.*

We will not prove the above theorem at this moment, as it follows as a corollary to the main result of the next chapter (Theorem 2.4.1).

Chapter 2

Matrix Scalings

2.1 Introduction

In Theorem 1.3.3, we are minimizing over a very particular set of matrices. Given the matrix $G_{|\phi\rangle}$, we are restricting our optimization to doubly quasi-stochastic matrices that satisfy $B = D^* G_{|\phi\rangle} D$ for some diagonal matrix D. Theorem 1.3.4 guarantees that such a matrix always exists, as $G_{|\phi\rangle}$ is positive definite. The problem of searching for matrices of this type is not a particularly novel one. Indeed, many mathematicians have studied similar problems, albeit with different assumptions on the matrices in question. We begin this chapter by introducing two of the most famous results conerning these so-called *matrix scalings*. At the end of this chapter, we will introduce a new housekeeping result which supersedes this pair, combining them with Thereom 1.3.4 into one "unifying" theorem.

This short survey certainly does not contain every important scaling result from the last

50 years. Notable papers that are not discussed at length in this thesis include Johnson and Reams' investigation into symmetric matrix scalings ([26]) and Sinkhorn's study of non-negative row-stochastic scalings ([27]). For a more in-depth review of scalings, the interested reader is encouraged to see [28].

2.2 Historical results: Scaling non-negative matrices and copositive matrices

The first notable results on scalings were those of Richard Sinkhorn in [29]. Motivated by a connection to Markov processes, Sinkhorn was interested in manipulating the entries of a matrix to achieve double-stochasticity. In particular, Sinkhorn proved the following theorem:

Theorem 2.2.1 ([29], Theorem 1) *Given an $n \times n$ matrix with strictly positive entries A, there exists a pair of positive definite diagonal matrices, D_1, D_2 such that D_1AD_2 is a doubly-stochastic matrix. Further, D_1 and D_2 are unique up to a scalar factor.*

Remark (1) Authors often refer to the matrix D_1AD_2 as the *Sinkhorn scaling* of A.

(2) It is worth mentioning that in 1961 (three years before Sinkhorn's paper), Marcus and Newman [30] announced this result for symmetric matrices. Despite this, however, the paper by Sinkhorn is considered by most to be the first major result in the study of matrix scalings.

(3) Later, Sinkhorn and Knopp [31] proved by way of a constructive algorithm that we can

still find D_1 and D_2 if we relax our condition on A, requiring non-negativity of the entries and certain zero-nonzero patterns rather than strict positivity. Their algorithm consists of alternately scaling the rows and columns of our matrix, where an easy limiting argument proves convergence.

(4) While Sinkhorn proved his theorem by constructing a convergent algorithm, there have been many different proofs of Theorem 2.2.1 over the years. Perhaps the two most notable proofs are one by Menon [32] using fixed point theory, and a geometric proof by Borobia [33] using convex analysis on the column space of doubly-stochastic matrices.

When considering Sinkhorn's result (Theorem 2.2.1), it is not too difficult to see that when our non-negative matrix A is symmetric, we must have $D_1 = D_2$. In [4], the authors extended this idea to all symmetric copositive matrices.

Theorem 2.2.2 ([4], Theorem 1) *Let $A \in \mathbb{M}_n(\mathbb{R})$ be a symmetric, strictly copositive matrix. Then there exists a positive definite diagonal matrix D such that DAD is doubly quasi-stochastic.*

Remark (1) In [4], this theorem is actually stated in a bit more generality, where they show that we can scale A to have any desired positive row (column) sums. We are really only interested in the special case where all row sums are equal, and so we state the theorem as above.

(2) As all positive definite real matrices are copositive, the conclusion holds for all positive definite real matrices as well. Combining this with the fact that $D^* = D$ for real diagonal

matrices, we can see that Pereira's result from [3] (Theorem 1.3.4 from our previous chapter) reduces to a special case of Theorem 2.2.2 in the case where we assume that the matrix A in Theorem 1.3.4 is assumed to be real.

(3) Marshall and Olkin used a compactness argument to guarantee the existence of D in the above theorem. In [5], however, O' Leary gives a constructive proof of the existence of D for positive definite matrices. In fact, she gives an algorithm to construct D that is relatively easy to implement. We will expand upon this algorithm in greater detail in Chapter 6.

For the remainder of this chapter, we will work towards a "unifying theorem" that will allow us to prove Theorem 1.3.4, Theorem 2.2.1, and Theorem 2.2.2 as corollaries. Before doing so, however, we introduce a bit of necessary terminology from the theory of convex cones.

2.3 Convex Cones: A brief introduction

This will only be a brief introduction to the theory of convex cones, with enough information to serve our purposes in the next section. The interested reader is encouraged to look to [34] for more information on convex cones, or [35] for an excellent treatment of the more general principles of convexity.

We begin with a reminder of what is meant when we talk of convex cones:

Definition Let $\mathcal{K}$ be a subset of a (real or complex) vector space V. We say that $\mathcal{K}$ is

a *convex cone* if for any $\alpha, \beta \in \mathcal{K}$, and for any non-negative constants $\lambda_1, \lambda_2 \in \mathbb{R}_+$, it is necessarily the case that $\lambda_1\alpha + \lambda_2\beta \in \mathcal{K}$.

We will be dealing with convex cones of matrices in particular. We have already seen a few examples of these, in our introduction.

Example *It is easy to verify that the following sets each comprise a convex cone:*

(1) The set of $n \times n$ positive semi-definite matrices $\mathcal{PSD}_n$

(2) The set of $n \times n$ copositive matrices $\mathcal{COP}_n$.

(3) The set of $n \times n$ completely positive matrices $\mathcal{CP}_n$.

(4) The set of $n \times n$ matrices with all non-negative real entries $\mathcal{NN}_n$.

(5) The set of $n \times n$ doubly non-negative matrices $\mathcal{DNN}_n$.

We now introduce an entity that arises in areas such as linear programming [36], functional analysis [37], and matrix analysis (as we will see in our next section).

Definition Let S be a subset of the inner product space V. Then we define the *dual cone of* $\mathcal{S}$, denoted $\mathcal{S}^*$, to be the set of vectors $v \in V$ such that $\langle s, v\rangle \geq 0$ for all $s \in \mathcal{S}$.

We will be interested in the case where $\mathcal{S}$ is itself a cone of matrices. In this case, we will be using the so-called "Frobenius inner product" (or "trace inner product"), defined by $\langle A, B\rangle_F = tr(AB^*)$. The interested reader is encouraged to verify the following:

Example *The convex cones introduced above have the following dual cones:*

(1) $\mathcal{PSD}_n^* = \mathcal{PSD}_n$ (i.e. this cone is "self-dual").

(2) and (3): $\mathcal{COP}_n^* = \mathcal{CP}_n$ *and* $\mathcal{CP}_n^* = \mathcal{COP}_n$.

(4) $\mathcal{NN}_n$ *is self-dual.*

(5) $\mathcal{DNN}^* = \{A + B : A = A^*, A \in \mathcal{NN}_n, B \in \mathcal{PSD}_n\}$.

In our next section, we will need to construct convex cones that contain certain sets. The following observation will allow us to do this with ease. Recall that the convex hull of a set $\mathcal{S}$, denoted $conv(\mathcal{S})$, is defined to be the set of all convex combinations of elements of $\mathcal{S}$.

Proposition 2.3.1 *Let* $\mathcal{S}$ *be a subset of a vector space* V*, such that for any* $s \in S$ *and any non-negative constant* $k \in \mathbb{R}_+$*,* $ks \in S$*. Then* $conv(\mathcal{S})$ *is a convex cone.*

Remark When a set $\mathcal{S}$ has the property defined in Proposition 2.3.1, we often refer to $\mathcal{S}$ as a (not necessarily convex) *cone*.

In our next section, we will see how we can use the theory of convex cones to prove all of the scaling results that we have seen thus far.

2.4 A unifying scaling theorem

Before we can state the main result of this chapter, we must first introduce a bit of notation. Let $f : \mathbb{R}^2 \times \mathbb{T} \to \mathbb{C}^2$ be defined as follows:

$$f(r_1, r_2, t) = (r_1 e^{it}, r_2 e^{it}).$$

Let $\mathcal{RO}$ be the space of complex $n \times n$ rank-one matrices and define the homomorphism $\phi : (C^2)^n \to \mathcal{RO}$ by $\phi((z_{11}, z_{12}), (z_{21}, z_{22}), ..., (z_{n1}, z_{n2})) = vw^*$, where $v_i = z_{i1}$, and $w_i = z_{i2}$ for $i = 1...n$.

Fix $n \in \mathbb{N}$, and choose a subset $\mathcal{G} \subseteq \mathbb{R}^2 \times \mathbb{T}$. We define $\mathcal{H}_\mathcal{G}$ to be the set of rank-one matrices obtained as the image of the Cartesian product $(f(\mathcal{G}))^{\times n}$ under ϕ. Lastly, let $\mathcal{K}_\mathcal{G}$ denote the convex hull of $\mathcal{H}_\mathcal{G}$, that is:

$$\mathcal{K}_\mathcal{G} = conv\{\mathcal{H}_\mathcal{G}\}.$$

Example *Suppose that* $\mathcal{G} = \mathbb{R}_+^2 \times \mathbb{T}$. *Then* $f(\mathcal{G})$ *will consist of all pairs of complex numbers whose components have the same complex argument. Elements of* $\mathcal{H}_\mathcal{G} = \phi((f(\mathcal{G}))^n)$ *are matrices of the form:*

$$\begin{pmatrix} r_{11}e^{it_1} \\ r_{12}e^{it_2} \\ \vdots \\ r_{1n}e^{it_n} \end{pmatrix} \begin{pmatrix} r_{21}e^{it_1} \\ r_{22}e^{it_2} \\ \vdots \\ r_{n1}e^{it_n} \end{pmatrix}^* , \quad \textit{where } r_{ij} \in \mathbb{R}_+, t_j \in \mathbb{T}.$$

(i.e. $\mathcal{H}_\mathcal{G}$ *is the set of all rank one* $n \times n$ *complex matrices* A *with non-negative diagonals, and* $arg(a_{ij}) = -arg(a_{ji})$.*) By extension,* $\mathcal{K}_\mathcal{G}$ *is just the set of all* $n \times n$ *complex matrices with non-negative diagonals.*

We are now finally ready to state our theorem. In what follows, we will alternately consider cones of matrices in $\mathbb{M}_n(\mathbb{R})$ and cones of matrices in $\mathbb{M}_n(\mathbb{C})$. When considering

purely real cones, we take the dual cone to also contain only real matrices (that is, the dual is calculated using $\mathbb{M}_n(\mathbb{R})$ as the underlying vector space). We let $relint(\mathcal{A})$ denote the relative interior of a convex set $\mathcal{A}$, and $A \circ B$ denote the Hadamard product of two $m \times n$ matrices A and B (i.e. $(A \circ B)_{ij} = a_{ij}b_{ij}$).

Theorem 2.4.1 *Let $\mathcal{S}$ be any subspace of $\mathbb{R}^2$ such that neither the first nor the second coordinate is identically zero, and let $\mathcal{T}$ be one of $\{\{0\}, \mathbb{T}\}$. Define $\mathcal{G} = (\mathcal{S} \cap \mathbb{R}^2_+) \times \mathcal{T}$. Then for any $A \in relint(\mathcal{K}^*_{\mathcal{G}})$, there exists a rank-one element $B \in \mathcal{K}_{\mathcal{G}}$ such that $A \circ B \in \mathcal{K}^*_{\mathcal{G}}$ is doubly quasi-stochastic.*

Proof We begin by showing that for any $A \in {\mathcal{K}_{\mathcal{G}}}^*$ and rank one $B \in \mathcal{K}$ we must necessarily have $A \circ B \in {\mathcal{K}_{\mathcal{G}}}^*$. This amounts to showing that $tr((A \circ B)C^*) \geq 0$ for any $C \in \mathcal{K}_{\mathcal{G}}$. Indeed:

$$\begin{aligned} tr((A \circ B)C^*) &= e^T(A \circ B \circ \overline{C})e \\ &= e^T(A \circ (B \circ \overline{C}))e \\ &= tr(A(\overline{B} \circ C)^*). \end{aligned}$$

Thus we need only show that $\overline{B} \circ C \in \mathcal{K}_{\mathcal{G}}$ (as A is in ${\mathcal{K}_{\mathcal{G}}}^*$). Firstly note that $\overline{B} \in \mathcal{K}_{\mathcal{G}}$, as $\mathcal{G}$ is closed under the complex conjugation function. Let $\overline{B} = v_B w_B^*, C = \sum \lambda_i v_i w_i^*$, then $\overline{B} \circ C = \sum \lambda_i (v_B \circ v_i)(w_B \circ w_i)^*$, and it is easy to see by the definition of $\mathcal{G}$ that this vector product must be in $\mathcal{K}_{\mathcal{G}}$ (as $v_B \circ v_i$ and $w_B \circ w_i$ must be in $\phi((f(\mathcal{G}))^n)$ for all i).

The remainder of our proof follows similar to reasoning to Pereira's proof of Lemma 2.9

in [3], but with additional generality. We will need to minimize a particular quantity over the set $\Omega = \phi(\mathcal{Q})$, where $\mathcal{Q} \subset \mathbb{C}^{2n}$, defined by

$$\mathcal{Q} = \{w = ((z_{11}, z_{12}), (z_{21}, z_{22}), ..., (z_{n1}, z_{n2})) \in (f(\mathcal{G}))^n : \prod_{i=1}^{n} |z_{i1}| = \prod_{i=1}^{n} |z_{i2}| = 1\}.$$

In simpler terms, Ω is the subset of rank one matrices $vw^* \in \mathcal{H}_\mathcal{G}$ such that the componenets of v and w satisfy $\prod_{i=1}^{n} |v_i| = \prod_{i=1}^{n} |w_i| = 1$. Consider the following minimization:

$$k := \min_{B \in \Omega} tr(AB^*). \tag{1}$$

We claim that for the (possibly non-unique) matrix $B_* \in \Omega$ that minimizes (1), $A \circ \overline{B_*}$ has all row and column sums equal and nonzero. We break our proof of this fact into three parts, demonstrating the following:

(1): The minimum in (1) is always well-defined and is non-zero.

(2): The matrix $A \circ \overline{B_*}$ has all real row and column sums.

(3): The matrix $A \circ \overline{B_*}$ has all row and column sums equal.

Verification of (1): Firstly, note that for any $B \in \mathcal{H}_\mathcal{G}$, if $|b_{ij}| < 1$ for all i, j then $B \notin \Omega$. We consider the subset $\mathcal{N} \subset \mathcal{H}_\mathcal{G}$, consisting of all $B \in \mathcal{H}_\mathcal{G}$ that satisfy $\max_{i,j} |b_{ij}| = 0.9$. $\mathcal{N}$ is disjoint with Ω, and it is a compact set. By virtue of its compactness, we can obtain the minimum of $tr(AB^*)$ over $\mathcal{N}$. Let $k_2 = \min_{B \in \mathcal{N}} tr(AB^*)$. As A is in the relative interior of $\mathcal{K}_\mathcal{G}{}^*$, we know that $k_2 > 0$.

Now, let us define a constant $l = max\{n, \frac{0.81n^2}{k_2^2}(tr(AJ_n^*))^2\}$ (where J_n is the all ones matrix), and let

$$\mathcal{M} = \{B \in \Omega \mid \|B\|_F^2 \leq l\}.$$

Then this is compact and nonempty (for example, as $J_n \in \Omega$ and $\|J_n\|_F^2 = n$, it must be in $\mathcal{M}$). Thus we attain a minimum of $tr(AB^*)$ over $\mathcal{M}$. Let $\epsilon = \min_{B \in \mathcal{M}}(tr(AB^*))$. Again, $\epsilon > 0$ as A is in the relative interior of ${\mathcal{K}_\mathcal{G}}^*$.

We claim that ϵ is in fact a global minimum over all of Ω. Indeed, suppose that $C \in \Omega \backslash \mathcal{M}$. Then $\|C\|_F^2 > \frac{0.81n^2}{k_2^2}(tr(AJ_n^*))^2$. By the definition of the Frobenius norm, we know that there must exist indices i, j such that

$$|c_{ij}| > \frac{0.9}{k_2} tr(AJ_n^*). \tag{2}$$

Suppose that $\max_{i,j} |c_{ij}| = m$. Then $C = \frac{m}{0.9}N$, for some $N \in \mathcal{N}$. Thus

$$\begin{aligned} tr(AC^*) &= tr\left(\frac{m}{0.9}AN^*\right) \\ &= \frac{m}{0.9}tr(AN^*) \\ &\geq \frac{mk_2}{0.9} \\ &> tr(AJ_n^*) \\ &\geq \epsilon, \end{aligned}$$

where the last three inequalities follow from the definition of k_2, (2), and the definition of ϵ, respectively.

Thus, we have that ϵ is the minimum over all of Ω, as desired. For the remainder of the proof, let $M = A \circ \overline{B_*}$, where B_* is the (not necessarily unique) element of Ω satisfying $tr(AB_*^*) = \epsilon$. Note that this means that $\min_{B \in \Omega} tr(MB^*) = tr(MJ_n^*) = \epsilon > 0$.

Verification of (2): If $\mathcal{T} = \{0\}$, then every element of ${\mathcal{K}_\mathcal{G}}^*$ is real, and our result follows. We thus restrict ourselves to the case where $\mathcal{T} = \mathbb{T}$. Let us denote the k-th row (column) sum of M as r_k (c_k), and suppose that at least one row or column sum of M is not real. Without loss of generality, suppose that it is row 1. That is, suppose $r_1 = (1, 0, 0..., 0)M(1, ..., 1)^T = re^{i\theta} \notin \mathbb{R}$. Since $M \in {\mathcal{K}_\mathcal{G}}^*$ and $C = (e^{i\theta}, 1, 1, ..., 1)(e^{i\theta}, 1, ..., 1)^*$ must be in $\mathcal{K}_\mathcal{G}$, we must have $tr(MC^*) \geq 0$ and we conclude that $c_1 = se^{-i\theta}$. Then

$$tr(MC^*) = tr(MJ_n^*) - r_1 - c_1 + m_{11} - |r_1| - |c_1| - m_{11} < tr(MJ_n^*).$$

But $C \in \Omega$, and this inequality contradicts the definition of M. Thus, all row sums must be real.

Verification of (3): Now, suppose that 2 rows (columns) of M do not have equal row (column) sums. Without loss of generality, suppose that $r_1 > r_2$. We break our proof into 2 different cases for $\mathcal{S}$.

Suppose $\mathcal{S} = \mathbb{R}^2$, and let $\delta > 0$. Define $C_\delta = vw^* \in \Omega$, where $v = (1-\delta, (1-\delta)^{-1}, 1, ..., 1)^T$, and $w = (1, 1, ..., 1)^T$. Then

$$\begin{aligned} tr(MC_\delta^*) &= tr(MJ_n^*) - r_1 - r_2 + (1-\delta)r_1 + (1-\delta)^{-1}r_2 \\ &= tr(MJ_n^*) - r_2 - \delta r_1 + r_2(1 + \delta + \delta^2 + ...) \\ &= tr(MJ_n^*) + \delta(r_2 - r_1) + r_2(\mathcal{O}(\delta^2)), \end{aligned}$$

whence we see that for small enough δ (ensuring that the linear term dominates), the terms following $tr(AJ_n^*)$ must be negative (as $r_1 > r_2$), and thus $tr(MC_\delta^*) < tr(MJ_n^*)$, contradicting the definition of M.

Now suppose $\mathcal{S}$ is a one-dimensional subspace of $\mathbb{R}$, then $\mathcal{K}_\mathcal{G}$ consists of Hermitian matrices. Thus $K_\mathcal{G}^*$ must also only contain Hermitian matrices, which of course must have symmetric real part. Thus $r_i = c_i$ for all i (as we know from **(2)** that all row and column sums are real).

As $\mathcal{K}_\mathcal{G}$ contains all real, symmetric matrices, we consider $F_\delta = vv^* \in \Omega$, where $v = (1-\delta, (1-\delta)^{-1}, 1, ..., 1)^T$. Then (recalling that M must have symmetric real part and thus $Re(m_{12}) = Re(m_{21})$):

$$
\begin{aligned}
tr(MF_\delta^*) &= tr(MJ_n^*) - 2r_1 - 2(r_2 - Re(m_{21})) + m_{11} + m_{22} + m_{12} + m_{21} \\
&+ (1-\delta)^2 m_{11} + 2(1-\delta)(r_1 - m_{11} - Re(m_{12})) + (1-\delta)^{-2} m_{22} \\
&+ 2(1-\delta)^{-1}(r_2 - m_{22} - Re(m_{21})) \\
&= tr(MJ_n^*) - 2r_1 - 2(r_2 - Re(m_{21})) + m_{11} + m_{22} + 2Re(m_{12}) + m_{11} - 2\delta m_{11} \\
&+ \delta^2 m_{11} + 2r_1 - 2m_{11} - 2Re(m_{12}) - 2\delta(r_1 - m_{11} - Re(m_{12})) \\
&+ (1 + 2\delta + \delta^2(1 + \delta + ...) + ...)m_{22} + 2(1 + \delta + \delta^2 + ...)(r_2 - Re(m_{21}) - m_{22}) \\
&= tr(MJ_n^*) - 2\delta(r_1 - r_2) + \mathcal{O}(\delta^2),
\end{aligned}
$$

and again we see that for small enough δ, $tr(MF_\delta^*) < tr(MJ_n^*)$, contradicting the definition of M.

We have just seen that all rows have the same row sums ($r_i = r$, for all i) and by the same argument, all columns have the same columns sums ($c_i = c$ for all i). It follows that $r = c$, by the obvious fact that $\sum_i^n r_i = \sum_i^n c_i$. (While this observation is unnecessary in the case where $\mathcal{S}$ is one-dimensional, we do need this fact in the case where $\mathcal{S} = \mathbb{R}^2$.)

Through **(1)**,**(2)**, and **(3)**, we have shown that $M = A \circ \overline{B_*}$ has all row and columns sums equal to some positive real number, r (positivity follows from the fact that $tr(MJ_n^*) = \epsilon > 0$). Thus $A \circ \frac{\overline{B_*}}{r}$ has all row and column sums equal to 1, as desired. □

Remark Our assumptions on $\mathcal{S}$ amount to requiring that $\mathcal{S}$ is not the zero subspace, the x_1 axis, nor the x_2 axis. This assumption is necessary, as taking the Hadamard product with

any member of the resulting space of rank one matrices $\mathcal{H}_{\mathcal{G}}$ will always yield at least one zero row, preventing us from achieving a DQS matrix.

We now run down the specific results that arise from different choices of $\mathcal{S}$ and $\mathcal{T}$. Although most of these results were stated earlier as standalone theorems, we re-state them here as corollaries to emphasize the role of Theorem 2.4.1.

Corollary 2.4.2 ([29], Theorem 1) *Given an $n \times n$ matrix A with strictly positive entries, there exists a pair of positive definite diagonal matrices D_1, D_2 such that $D_1 A D_2$ is a doubly stochastic matrix.*

Proof In the language of Theorem 2.4.1, Let $\mathcal{S} = \mathbb{R}^2$, and let $\mathbb{T} = 0$. Then $\mathcal{K}_{\mathcal{G}}$ are the $n \times n$ matrices with non-negative entries, $relint(\mathcal{K}_{\mathcal{G}}{}^*)$ are the $n \times n$ matrices with positive entries. If $B = vw^*$ is the matrix guaranteed to exist by Theorem 2.4.1, then $D_1 = diag(v)$, $D_2 = diag(w)$ achieves the desired scaling. □

Corollary 2.4.3 ([4], Theorem 1) *Let A be a (real), symmetric, strictly copositive matrix. Then there exists a positive definite matrix D such that DAD has all row and column sums equal to 1.*

Proof In the language of Theorem 2.4.1, let $\mathcal{S}$ be any one-dimensional subspace of $\mathbb{R}^2$, and $\mathcal{T} = \{0\}$. Then $\mathcal{K}_{\mathcal{G}}$ is the set of all completely positive matrices. Thus, the dual $\mathcal{K}_{\mathcal{G}}^*$ is the set of copositive matrices. If $B = vv^*$ is the matrix guaranteed to exist by Theorem 2.4.1, (where $v \in \mathbb{R}^n$), then define $D = diag(v)$ and we achieve the desired scaling. □

Corollary 2.4.4 ([3], Lemma 2.9) *Let $A \in \mathbb{C}^{n\times n}$ be a positive definite matrix. Then there exists an $n \times n$ diagonal matrix D such that D^*AD is doubly quasi-stochastic.*

Proof In the language of Theorem 2.4.1, let $\mathcal{S}$ be any one-dimensional subspace of $\mathbb{R}^2$, and $\mathcal{T} = \mathbb{T}$. Then $\mathcal{K}_\mathcal{G}$ consists of all positive semi-definite matrices, a self-dual cone, and $relint(\mathcal{K}_\mathcal{G}^*)$ is the set of positive definite matrices. If $B = vv^*$ for some $v \in \mathbb{C}^n$, then letting $D = diag(v)$ achieves our desired scaling. □

Remark The fourth case, where $\mathcal{S} = \mathbb{R}^2$ and $\mathcal{T} = \mathbb{T}$, is a bit less interesting. In this case, $\mathcal{K}_\mathcal{G}^*$ is just the set of non-negative diagonal matrices and the relative interior of this set is the set of positive diagonal matrices. If we let $A \in relint(\mathcal{K}_\mathcal{G}^*)$, then we can just choose $B = vv^*$ to be any rank one element of $\mathcal{K}_\mathcal{G}$ satisfying $b_{ii} = a_{ii}^{-1}$. In the language of diagonal scaling this is essentially equivalent to saying $D^{-1/2}DD^{-1/2}$ has all row and column sums equal to 1, which is clearly true.

Chapter 3

Tensor Scalings

3.1 Introduction

This chapter will be focused on the scaling of entities known as *tensors*. While they are related, these are a different type of tensor than those that were discussed in Chapter 1. These tensors (also known as hypermatrices) are a natural generalization of matrices, and share many of the algebraic and analytic properties of matrices. Indeed, many authors have recently been able to extend "classical" results on square matrices to m-th order, n-dimensional tensors. Over the last twenty years, this has given rise to many interesting results on such entites as the (generalized) eigenvalues of tensors (as in [38], [39]), the (hyper)determinant (see [40]), and other properties of tensors that arise as natural extensions of those of matrices. These extensions have proven useful in such areas as the theory of homogeneous polynomials (we will see the connection later in this chapter) and linked object ranking [41],

as well as other fields. In the 1980s (20 years after the results of Sinkhorn and Marshall and Olkin), mathematicians became interested in how they could extend the scaling results of the previous chapter to general m-th order tensors.

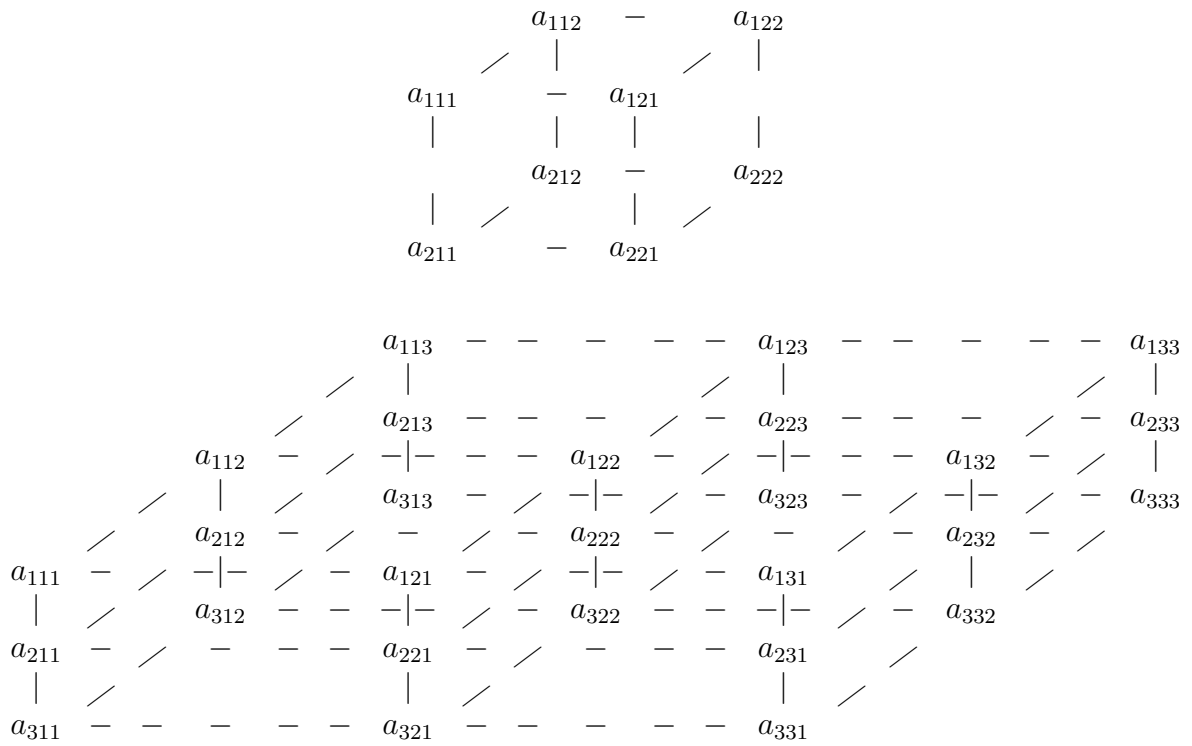
After defining some terminology, we will introduce some of the successful extensions of Sinkhorn's Theorem (Theorem 2.2.1) that arose from these researchers' efforts. After this, we will provide a natural extension of Marshall and Olkin's copositive scaling result (Theorem 2.2.2). While it seems reasonable to assume that some authors knew that such an extension was possible, Theorem 3.4.1 does not seem to be found anywhere in the literature.

3.2 General tensor definitions

Suppose we have an $m \times n$ matrix A with entries from $\mathbb{F}$. If we let I_n denote the set $\{1, 2, ..., n\}$, we can consider this as a function f_A taking elements from $I_m \times I_n$ into $\mathbb{F}$ – we simply define the function as $f_A(i, j) = a_{ij}$. A tensor (or multi-dimensional matrix) is merely the natural extension of this to higher-dimensional domains. An m-th order, n-dimensional tensor is simply a mapping from $I_n \times I_n \times ... \times I_n$ (where the product is taken m times) into $\mathbb{F}$. As with matrices, we refer to the image of $i_1, i_2, ..., i_m$ under this mapping as the $i_1, i_2, ...i_m$-th entry of the associated tensor. In what follows, we will only be concerning ourselves with the case where $\mathbb{F} = \mathbb{R}$ (i.e. all entries of our tensors are real).

For this chapter, we shall reserve the script roman letters $\mathcal{A}, \mathcal{B}, ...$ to denote tensors, and (as with matrices), we will denote the $i_1, i_2, ...i_m$-th entry of $\mathcal{A}$ by $(\mathcal{A})_{i_1,i_2,...i_m}$, or by the associated lower case letter with necessary subscript ($a_{i_1 i_2,...i_m}$, for example). Just as it

is often useful to view matrices as "rectangles" of numbers, it often helps to view tensors as hyper-rectangles. For example, the following images give a visualization of two generic 3rd order tensors. The first is a 3rd order, 2-dimensional tensor and the latter is a 3rd order, 3-dimensional tensor (otherwise known as a $2 \times 2 \times 2$ tensor and a $3 \times 3 \times 3$ tensor, respectively):



A 1st-order tensor is just a vector and a 2nd-order tensor is a matrix. We cannot provide a similar visualization for tensors of order $m > 3$, as we would need 4 or more spatial dimensions.

For every tensor, there is an associated homogeneous polynomial that arises in the following way. Given an m-th order, n-dimensional tensor $\mathcal{A}$, and a vector $x \in \mathbb{R}^n$, we can

define the tensor vector product, denoted $\mathcal{A}x^m$ as follows:

$$\mathcal{A}x^m = \sum_{i_1,...i_m=1}^{n} a_{i_1...i_m} x_{i_1}...x_{i_m}.$$

As we will be extending Theorem 2.2.2, we will need the following well-known definition (see, for example, [42]).

Definition Given an m-th order, n-dimensional tensor $\mathcal{A}$, we say that $\mathcal{A}$ is *copositive* if $Ax^m \geq 0$ for all $x \in \mathbb{R}^n_+$. If this inequality is strict whenever $x \neq 0$, then we say $\mathcal{A}$ is *strictly copositive*.

Remark If the concepts in this section are new to the reader, they should convince themselves that in the case where $m = 2$, this definition does, in fact, coincide with our definition of the set $\mathcal{COP}_n$ from Chapter 1.

We close our introduction with an generalization of matrix symmetry:

Definition Let $\mathcal{A}$ be an m-th order, n-dimensional tensor. $\mathcal{A}$ is said to be *(super)symmetric* if $a_{i_1...i_m} = a_{j_1...j_m}$ whenever $(i_1, i_2, ..., i_m) = \sigma(j_1, j_2, ..., j_m)$ for all permuations $\sigma \in S_m$.

3.3 Introduction to tensor scalings

As we seek to extend Marshall and Olkin's result on symmetric, copositive matrices, we must first decide on terminology. It is relatively straightforward to define the matrix tensor product DAD for symmetric tensors:

Definition Given an m-th order, symmetric, n-dimensional tensor A, and a diagonal matrix $D = \text{diag}(d_1, ..., d_n)$, we define the diagonal scaling DAD as the m-th order n-dimensional tensor

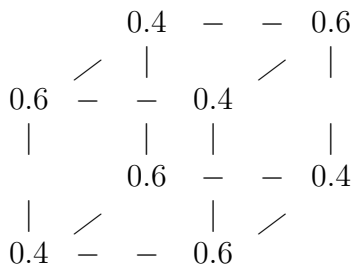
$$(DAD)_{i_1,...i_m} = a_{i_1...i_m} d_{i_1} ... d_{i_m}.$$

Less obvious, perhaps, is how we can take the notion of a matrix having equal row and column sums, and generalize this to tensors. Perhaps the following definition seems to be the obvious answer:

Definition Let $\mathcal{A}$ be an m-th order, n-dimensional tensor. We will call $\mathcal{A}$ *1D-quasi-stochastic* if the elements of every $n \times 1$ subtensor sum to 1. i.e. For all fixed $j_2, ...j_m$, we have:

$$\sum_{j_1=1}^{n} a_{j_1 j_2 ... j_m} = 1.$$

Example *Consider following 2-dimensional (super-symmetric), 3rd-order tensor:*

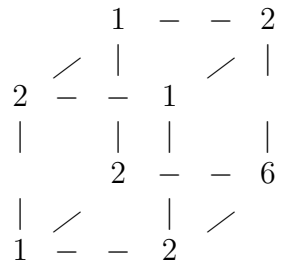


This is 1D-quasi-stochastic, as all 12 $n \times 1$ subtensor (all 4 "horizontal", all 4 "vertical"

and all 4 "depth" lines) sum to 1.

While this may seem the most natural extension, it is not the best choice. Our goal is to extend Theorem 2.2.2 to tensors, which means that we will want to be able to find a scaling for any symmetric, copositive tensor. It turns out that the concept of 1D-quasi-stochasticity is too restrictive.

Example *Let $\mathcal{A}$ be the following 2-dimensional, 3rd-order, symmetric, copositive tensor:*



Then $\mathcal{A}$ cannot be scaled to a 1D-quasi-stochastic tensor. To see this, suppose $D = diag(x_1, x_2)$ scales $\mathcal{A}$ to a 1D-quasi-stochastic tensor, and note that both x_1 and x_2 are nonzero (indeed, if one was zero, $D\mathcal{A}D$ would have many $n \times 1$ subtensors consisting of only zeroes). In particular, we must have:

$$\begin{aligned} 1 &= (D\mathcal{A}D)_{111} + (D\mathcal{A}D)_{112} \\ &= (D\mathcal{A}D)_{121} + (D\mathcal{A}D)_{122} \\ &= (D\mathcal{A}D)_{211} + (D\mathcal{A}D)_{212} \\ &= (D\mathcal{A}D)_{221} + (D\mathcal{A}D)_{222}. \end{aligned}$$

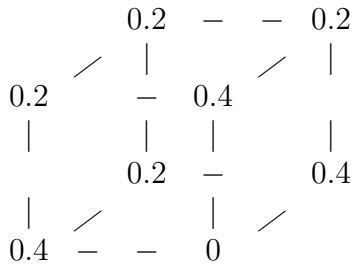
We can rewrite our second equality as $2x_1^3+x_1^2x_2 = x_1^2x_2+2x_1x_2^2$*, whence we obtain* $x_1^3 = x_1x_2^2$*, or* $x_1^2 = x_2^2$ *(as* x_1 *cannot be zero). Similarily, our last equality yields* $x_1^2x_2 = 6x_2^3$*. But we have already established that* $x_1^2 = x_2^2$*, implying* $x_2^3 = 6x_2^3$*. This requires* $x_2 = 0$*, which cannot be the case, and we have our contradiction.*

This means that we must seek another type of matrix to which we can scale $\mathcal{A}$. The following turns out to be a better choice:

Definition Given an m-th order, n-dimensional tensor $\mathcal{A}$, we say that $\mathcal{A}$ is *slice stochastic* if, for any fixed $(k, j) \in I_m \times I_n$, the (k, j)-th sum slice

$$S_j := \sum_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_m=1}^{n} a_{i_1\ldots i_{k-1},j,i_{k+1},\ldots,i_{m-1},i_m} = 1.$$

Example *The following 3rd order, 2-dimensional matrix is slice-stochastic:*



This is easily verified by checking that the sum of the entries in each one of the 6 slices, represented here by the faces of the cube, is equal to 1.

Remark (1) It is easy to see that in the case where $\mathcal{A}$ is symmetric, the (k, j)-th slice sum remains constant as we vary k, but keep j fixed. Thus in the symmetric case, we can fix a value of k, (say k_*) and only consider the (k_*, j)-th slice sums when checking for slice-stochasiticity.

(2) In the language of Qi [40], the slice-stochasticity of a super-symmetric tensor $\mathcal{A}$ is equivalent to $\mathcal{A}$ having an (H-eigenvalue, eigenvector)-pair $(\lambda, v) = \left(1, (1, 1, 1, 1..., 1)^T\right)$.

(3) A moment's thought should convince the reader that both 1D-stochasticity and slice-stochasticity reduce to doubly quasi-stochasticity when the order of the tensor is 2.

Slice-stochasticity has been considered by many authors over the years (see [43], [44], [45]). In fact, the following result of Bapat (or, more specifically, a result on homogeneous polynomials that follows from this theorem) was even used in Gurvits' famous original proof of the Van der Waerden inequality in [46].

Theorem 3.3.1 ([43], Theorem 6) *Let $\mathcal{A}$ be an mth order, n-dimensional non-negative tensor with positive diagonal (that is, $a_{ii...i} > 0$, for all i). Choose n positive numbers $r_1, ...r_n$. Then there exists a diagonal matrix D such that $D\mathcal{A}D$ has $(1, j)$-th slice sum equal to r_j for all $j = 1, ..., n$.*

Theorem 3.3.1 has an apparent application to our problem: In the case where we choose $r_1 = r_2 = ...r_n = 1$, and we ensure that $\mathcal{A}$ (and hence $D\mathcal{A}D$) is supersymmetric, then $D\mathcal{A}D$ is in fact slice-stochastic. In our next section, we extend this result from non-negative symmetric tensors to real, symmetric, copositive tensors.

3.4 Extending Marshall and Olkin: Scaling copositive tensors

Just as other authors have extended non-negative matrix scaling results to non-negative tensors, we will generalize Theorem 2.2.2 to copositive tensors. We begin by noting that for symmetric tensors, we can state our tensor-vector product in a more useful way.

Recall that if $\mathcal{A}$ is an m-th order, n-dimensional tensor, and x is an n-dimensional vector, we define the multiplication

$$\mathcal{A}x^m = \sum_{i_1,\dots i_m=1}^{n} a_{i_1\dots i_m} x_{i_1}\dots x_{i_m}.$$

It will benefit us to note that when $\mathcal{A}$ is super-symmetric, this can be written as:

$$\mathcal{A}x^m = \sum_{\substack{k_1,\dots k_n \geq 0 \\ k_1+\dots+k_n=m}} \binom{m}{k_1, k_2, \dots, k_n} a_{\underbrace{1\dots1}_{k_1}\underbrace{2\dots2}_{k_2}\dots\underbrace{n\dots n}_{k_n}} x_1^{k_1}\dots x_n^{k_n}, \tag{1}$$

where $\binom{m}{k_1,k_2,\dots,k_n}$ is the number of distinct permutations of $(\underbrace{1,\dots,1}_{k_1},\underbrace{2,\dots,2}_{k_2},\dots\underbrace{n,\dots,n}_{k_n})$, otherwise known as the multinomial coeffficient:

$$\binom{m}{k_1, k_2, \dots, k_n} = \frac{m!}{k_1!k_2!\dots k_n!}.$$

In Theorem 2.2.2, Marshall and Olkin showed that given any symmetric, strictly copositive matrix (i.e. 2-tensor) $\mathcal{A}$, there exists a diagonal matrix with positive entries such that

$D\mathcal{A}D$ is doubly quasi-stochastic. We will extend this result for strictly copositive tensors of any order.

Theorem 3.4.1 *Given a symmetric, strictly copositive tensor $\mathcal{A}$ of order m, there exists a positive diagonal matrix D s.t. $D\mathcal{A}D$ is slice-stochastic.*

We require the following lemmas:

Lemma 3.4.2 *For any diagonal matrix $D = diag(x_1, ...x_n)$ and supersymmetric m-th order, n-dimensional tensor $\mathcal{A}$, the jth slice sum of $D\mathcal{A}D$ is given by:*

$$S_j = \sum_{\substack{k_1,...k_n \geq 0, \\ k_1+...+k_n=m}} \frac{1}{m} \binom{m}{k_1, k_2, ..., k_n} k_j a_{\underbrace{1...1}_{k_1}\underbrace{2...2}_{k_2}...\underbrace{n...n}_{k_n}} x_1^{k_1}...x_n^{k_n}.$$

Proof Combining our definitions for the j-th sum slice and the diagonal scaling $D\mathcal{A}D$, we see that

$$S_j = \sum_{i_1,...i_{m-1}=1}^{n} a_{ji_1...i_{m-1}} x_j x_{i_1}...x_{i_{m-1}}.$$

Exploiting symmetry as in (1), we see that this is equivalent to the following:

$$S_j = \sum_{\substack{k_1,...k_n \geq 0 \\ k_j \geq 1 \\ k_1+...+k_n=m}} \binom{m}{k_1, k_2, ..., k_n}_{Fix} a_{\underbrace{1...1}_{k_1}\underbrace{2...2}_{k_2}...\underbrace{n...n}_{k_n}} x_1^{k_1}...x_n^{k_n},$$

where $\binom{m}{k_1,k_2,...,k_n}_{Fix}$ is the number of distinct permutations of $(j, \underbrace{1, ..., 1}_{k_1}, , ... \underbrace{j, ..., j}_{k_j-1}, ..., \underbrace{n, ..., n}_{k_n})$

that leave the first element (j) fixed. This will simply be

$$\binom{m}{k_1,k_2,...,k_n}_{Fix} = \binom{m}{k_1,...,k_j-1,...,k_n} = \frac{(m-1)!}{k_1!...(k_j-1)!...k_n!} = \frac{k_j}{m}\binom{m}{k_1,k_2,...,k_n}.$$

Hence, we have our desired expression

$$S_j = \sum_{\substack{k_1,...,k_n\geq 0\\ k_1+...+k_n=m}} \frac{1}{m}\binom{m}{k_1,k_2,...,k_n} k_j a_{\underbrace{1...1}_{k_1}\underbrace{2...2}_{k_2}...\underbrace{n...n}_{k_n}} x_1^{k_1}...x_n^{k_n},$$

where we no longer require $k_j \geq 1$, as the terms with $k_j = 0$ will vanish. □

The following lemma follows similar reasoning to part of our proof of Theorem 2.4.1.

Lemma 3.4.3 *Let $f(x) = \mathcal{A}x^k$, where $\mathcal{A}$ is a strictly copositive, supersymmetric tensor. Then f has a positive absolute minimum on*

$$\Omega = \{x \in \mathbb{R}^n_+ : \prod x_i = 1\}.$$

Proof Let $S = \{x \in \mathbb{R}^n : x_i \geq 0, \|x\|_\infty :=\max\ x_i = \delta\}$, for some δ small enough that $S \cap \Omega = \emptyset$. Then S is compact and, as f is continuous, it has a (positive) minimum on S, which we will denote ϵ. Then $\mathcal{A}x^k \geq \epsilon$ for $x \in S$ and hence for any $x \in \Omega$ we have $\mathcal{A}(\frac{x\delta}{\|x\|_\infty})^k \geq \epsilon$ (as $\frac{x\delta}{\|x\|_\infty} \in S$). i.e.

$$\mathcal{A}x^k \geq \frac{\|x\|_\infty^k \epsilon}{\delta^k} \geq \frac{\epsilon}{\delta^k}(\prod_{j=1}^n x_j^{1/n})^k. \tag{2}$$

Choose any $y \in \Omega$, and let $L = \frac{(\mathcal{A}y^k)\delta^k}{\epsilon} > 0$, and define

$$T = \{x \in \mathbb{R}^n : x \geq 0, \|x\|_\infty \leq L^{1/k}\}.$$

Then $T \cap \Omega$ is nonempty, as y is in it, and it is compact (as it is closed and bounded). Hence f attains a minimum on $T \cap \Omega$ at some point m. Further, this is an absolute minimum on Ω, as if $x_0 \in \Omega \backslash T$, then $\|x_0\|_\infty > L^{1/k}$, and (plugging this into (2)) we have $\mathcal{A}x_0^k \geq \frac{Ay^k \delta^k \epsilon}{\epsilon \delta^k}$, yielding $f(x_0) > f(y)$. □

Now we may prove our main result (this is a reasonably straightforward generalization of the proof of Lemma 1 in [4] concerning copositive matrices).

Proof of Theorem 3.4.1 From Lemma 3.4.3, we know that $\mathcal{A}x^k$ attains a minimum on Ω. Hence, we can use Lagrange multipliers to minimize this function, subject to the constraint $x \in \Omega$.

To this end, we set

$$\frac{d}{dx}(Ax^m - m\lambda \sum log(x_i)) = 0,$$

and solve:

$$\frac{d}{dx}\left(\sum_{\substack{k_1,\ldots k_n \geq 0 \\ k_1+\ldots+k_n=m}} \binom{m}{k_1, k_2, \ldots, k_n} a_{1\ldots12\ldots2\ldots n\ldots n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n} - m\lambda \sum log(x_i)\right) = 0,$$

whence we obtain:

$$\begin{pmatrix} \sum\limits_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \binom{m}{k_1,k_2,\dots,k_n} k_1 a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1-1}x_2^{k_2}\dots x_n^{k_n} \\ \sum\limits_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \binom{m}{k_1,k_2,\dots,k_n} k_2 a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1}x_2^{k_2-1}\dots x_n^{k_n} \\ \vdots \\ \sum\limits_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \binom{m}{k_1,k_2,\dots,k_n} k_n a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1}x_2^{k_2}\dots x_n^{k_n-1} \end{pmatrix} = m\lambda \begin{pmatrix} \frac{1}{x_1} \\ \frac{1}{x_2} \\ \vdots \\ \frac{1}{x_n} \end{pmatrix}.$$

This yields

$$\begin{pmatrix} \sum\limits_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \frac{\binom{m}{k_1,k_2,\dots,k_n} k_1 a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}}{m} \\ \sum\limits_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \frac{\binom{m}{k_1,k_2,\dots,k_n} k_2 a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}}{m} \\ \vdots \\ \sum\limits_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \frac{\binom{m}{k_1,k_2,\dots,k_n} k_n a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}}{m} \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{3}$$

Now, taking the column sum of both sides, we obtain:

$$n\lambda = \sum_{\substack{k_1,\dots k_n\geq 0\\ k_1+\dots+k_n=m}} \frac{\binom{m}{k_1,k_2,\dots,k_n} (k_1+k_2+k_3\dots+k_n) a_{1\dots 12\dots 2\dots n\dots n} x_1^{k_1}x_2^{k_2}\dots x_n^{k_n}}{m},$$

and (exploiting the fact that $k_1+\dots+k_n = m$), we obtain $n\lambda = \mathcal{A}x^k > 0$, by strict copositivity.

Hence $\lambda > 0$. Dividing (3) by λ yields:

$$\begin{pmatrix} \sum\limits_{\substack{k_1,\ldots k_n \geq 0 \\ k_1+\ldots+k_n=m}} \frac{\binom{m}{k_1,k_2,\ldots,k_n} k_1 a_{1\ldots12\ldots2\ldots n\ldots n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}}{m\lambda} \\ \sum\limits_{\substack{k_1,\ldots k_n \geq 0 \\ k_1+\ldots+k_n=m}} \frac{\binom{m}{k_1,k_2,\ldots,k_n} k_2 a_{1\ldots12\ldots2\ldots n\ldots n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}}{m\lambda} \\ \vdots \\ \sum\limits_{\substack{k_1,\ldots k_n \geq 0 \\ k_1+\ldots+k_n=m}} \frac{\binom{m}{k_1,k_2,\ldots,k_n} k_n a_{1\ldots12\ldots2\ldots n\ldots n} x_1^{k_1} x_2^{k_2} \ldots x_n^{k_n}}{m\lambda} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

and from Lemma 3.4.2 we see that the left hand side is simply the array of slice sums of $D\mathcal{A}D$, where $D = diag(\frac{x_1}{\sqrt[m]{\lambda}}, \frac{x_2}{\sqrt[m]{\lambda}}, \ldots \frac{x_n}{\sqrt[m]{\lambda}}) > 0$. □

After proving the existence of a scaling, a natural question that arises is: How many scalings exist? While the answer is unknown in the general case (and we will investigate this question for matrices in the next chapter), we conclude with an upper bound on the number of scalings of a 2-dimensional, m-th order tensor.

3.5 Ennumerating the scalings of a 2-Dimensional, m-th order tensor

In [2], the authors showed that any positive definite 2×2 matrix (i.e. 2nd order tensor) has at most 2 scalings. We will now extend this result for symmetric 2-dimensional, m-th order tensors.

Theorem 3.5.1 *Let A be a 2-dimensional, m-th order, real, supersymmetric tensor. Then there exists at most m real diagonal matrices D such that $D\mathcal{A}D$ is a unique slice-stochastic tensor.*

Proof Let $D =$diag(x_1, x_2). For $D\mathcal{A}D$ to be slice stochastic, then there are two slice sums that we need to consider (one with first index fixed as 1, and one with first index fixed as 2):

$$1 = a_{1\ldots1}x_1^m + c_1 a_{1,1,\ldots,1,2}x_1^{m-1}x_2 + \ldots + c_{m-1}a_{1,2,2,\ldots,2}x_1x_2^{m-1} \tag{1}$$

$$1 = a_{2\ldots2}x_2^m + c_1 a_{2,2,\ldots,2,1}x_2^{m-1}x_1 + \ldots + c_{m-1}a_{2,1,1,\ldots,1}x_2x_1^{m-1}, \tag{2}$$

where c_i is the appropriate multinomial coefficient. Taking the difference of these two equations, we obtain:

$$\begin{aligned} 0 &= a_{1,\ldots,1}x_1^m + (c_1 - c_{m-1})(a_{1,1,\ldots,2})x_1^{m-1}x_2 + (c_2 - c_{m-2})(a_{1,1,\ldots1,2,2}x_1^{m-2}x_2^2) + \ldots \\ &+ (c_{m-1} - c_1)a_{1,2,2\ldots,2}x_1x_2^{m-1} - a_{2,2,\ldots,2}x_2^m. \end{aligned}$$

Dividing both sides of this equation by the nonzero x_2^m, and performing a change of variables $x := \frac{x_1}{x_2}$, we are left with:

$$0 = a_{1,\ldots,1}x^m + (c_1 - c_{m-1})a_{1,1,\ldots,1,2}x^{m-1} + \ldots + (c_{m-1} - c_1)a_{1,2,\ldots,2}x - a_{2,\ldots,2}. \tag{3}$$

By the fundamental theorem of algebra, this has at most m solutions for $x = \frac{x_1}{x_2}$. We now

show that for each of these solutions, there is only one pair x_1, x_2 that can possibly satisfy our two equations (1) and (2).

Suppose $\frac{x_1}{x_2} = k$, so $x_1 = kx_2$. Then (1) and (2) become:

$$1 = a_{1\ldots1}k^m x_2^m + c_1 a_{1,1,\ldots,1,2}k^{m-1}x_2^m + \ldots + c_{m-1}a_{1,2,2,\ldots,2}kx_2^m \tag{1*}$$

$$1 = a_{2\ldots2}x_2^m + c_1 a_{2,2,\ldots,2,1}kx_2^m + \ldots + c_{m-1}a_{2,1,1,\ldots,1}k^{m-1}x_2^m, \tag{2*}$$

and we have exactly one solution for x_2^m. Thus, once we choose our solution to (3), we obtain one and only one value for x_2^m.

Further, if we have two different values of x_2 that agree on their m-th powers, the resulting scalings $D\mathcal{A}D$ must be the same. Indeed, let j represent the number of indices i_l that are equal to 1. Then a given entry of the tensor

$$(D\mathcal{A}D)_{i_1,i_2,\ldots i_m} = a_{i_1,i_2,\ldots i_m}x_1^j x_2^{m-j} = k^j a_{i_1,i_2,\ldots i_m}x_2^m,$$

depend only on j, x_2^m, k and the value of the original tensor $\mathcal{A}$. Thus, for each of the m solutions to (3), we have at most one pair (x_1, x_2), such that $D = diag(x_1, x_2)$ scales $\mathcal{A}$ in a unique way. Our result follows. □

This concludes our discussion on tensor scalings. Our focus for the remainder of the thesis will be on complex scalings of positive definite matrices (i.e. the D^*AD scalings found in Theorems 1.3.3 and 1.3.4). Our next chapter will investigate a similar question to the problem considered in this section. We ask: Given a positive definite matrix A, how many

ways can A be scaled to a doubly quasi-stochastic matrix?

Chapter 4

Enumerating Matrix Scalings

4.1 Introduction

Theorem 1.3.3 suggests an interesting way to study the geometric measure of entanglement of Slater permanents. If we can find all complex scalings of a given positive definite matrix A and then identify the scaling with smallest permanent, we can easily calculate the GME of the Slater permanent with Gram matrix A. This suggests an obvious question: How do we know when we have found all scalings? That is, how many scalings exist? Recall our notation from the end of Chapter 1. Given a positive definite matrix A, the set of scalings, denoted $sc(A)$, is given by:

$$sc(G_{|\phi\rangle}) = \{B \in \mathcal{M}_n(\mathbb{C}) : B \text{ is DQS and } B = D^* G_{|\phi\rangle} D \text{ for some diagonal matrix } D\}.$$

After introducing Theorem 1.3.3, the authors of [2] made the following conjecture on the

number of scalings of a given positive definite matrix A.

Conjecture 4.1.1 ([2], Conjecture 2.1) *Given a positive definite matrix $A \in \mathcal{M}_n(\mathbb{C})$, $|sc(A)| \leq 2^{n-1}$.*

Note that we are interested in the cardinality of the set of scalings $B = D^*AD$, rather than the set of diagonal matrices that scale A. This is because there are always infinitely many such D. Indeed, if D scales A, and ω is any complex number of modulus 1, then $E = \omega D$ will scale A to the same DQS matrix (that is, $D^*AD = E^*AE$). This motivates our next definition:

Definition Let D, E be diagonal matrices. We will say that D, E are *equivalent* if $D = \omega E$ for some $\omega \in \mathbb{T}$.

Remark As mentioned above, it is clear that if D and E are equivalent, then $D^*AD = E^*AE$. In fact, it is not too difficult to see that if a positive definite matrix A has no zero entries, then $D^*AD = E^*AE$ if and only if D and E are equivalent. An easy proof of this can be found in [2].

The focus of this chapter will be to investigate Conjecture 4.1.1, and the cardinality of $|sc(A)|$ for different types of matrices. We will begin by showing that this conjecture does not hold for general $n \times n$ matrices when $n \geq 3$. We will then derive the true upper bound for 3×3 real matrices, and give conditions for when this bound is attained. Lastly, we turn our attention to matrices of higher dimension, and prove the existence of matrices with infinitely

many scalings for all $n \geq 4$. Unless otherwise noted, all of the results in this section have been collected in [47].

4.2 Preliminary results

We begin by reviewing a bit of the evidence for Conjecture 4.1.1. The following two results are proven in [2]:

Proposition 4.2.1 ([2]) *Let A be a 2×2 positive definite matrix. Then $|sc(A)| \leq 2$. Further, $sc(A)$ consists entirely of real matrices, exactly one of which is doubly stochastic.*

Proposition 4.2.2 ([2], Proposition 3.1) *Let A be an $n \times n$ tridiagonal matrix. Then $|sc(A)| \leq 2^{n-1}$.*

The following result makes our search for scalings much easier.

Proposition 4.2.3 ([2], Proposition 2.4) *Let A be an $n \times n$ complex positive definite matrix. Then there is at most one positive definite diagonal matrix D that scales A.*

We have an easy consequence to Proposition 4.2.3. This was intrinsically proven in [2], although never explicitly stated. We include a proof here for completeness.

Proposition 4.2.4 *Given an $n \times n$ positive definite matrix A, there is (up to equivalence) at most 2^{n-1} real diagonal matrices D such that D^*AD is a doubly quasi-stochastic matrix.*

Proof We will show that for any $1 \times n$ sign pattern $(\pm, \pm, ..., \pm)$, there is at most one real

diagonal matrix $D = diag(d_1, d_2, ..., d_n)$ such that $(d_1, d_2, ...d_n)$ fits that sign pattern and D scales A uniquely.

To this end, suppose there are 2 diagonal matrices D, E with the same sign pattern, both of which scale A. Observing that $(ED^{-1})DAD(D^{-1}E) = EAE$ is DQS, we see that $D^{-1}E$ is a positive diagonal matrix that scales DAD. But clearly, the identity matrix I_n is a positive diagonal matrix that scales DAD. By Proposition 4.2.3, these must be the same matrix and we obtain $ED^{-1} = I$, or $E = D$, as desired. As there are 2^n unique sign patterns, this tells us that we have maximum 2^n real matrices that scale A. Lastly, any diagonal matrix D is equivalent to $-D$, and so $DAD = (-D)A(-D)$. Hence, we obtain at most $\frac{2^n}{2} = 2^{n-1}$ unique scalings, as desired. □

4.3 A counterexample to Conjecture 4.1.1

Before we introduce our counterexample, we remind the reader of the class of matrices that our counterexample belongs to, the *circulants*:

Definition An $n \times n$ matrix $A = (a_{ij})$ is called circulant if $a_{i_1 j_1} = a_{i_2 j_2}$ whenever $j_1 - i_1 = j_2 - i_2$, where subtraction is taken modulo n.

We often introduce a circulant matrix as $A = circ(a_{11}, ..., a_{1n})$, as these entries determine all other entries of the matrix. For a detailed treatment of circulant matrices, the reader is encouraged to see [48].

We now present our counterexample to Conjecture 4.1.1, a 3×3 circulant matrix with 6

scalings.

Example 4.3.1 *Let A be the positive definite matrix*

$$A = circ(0.5, 0.25, 0.25) = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}.$$

Then one can easily verify that the following diagonal matrices scale A to 6 unique doubly quasi-stochastic matrices:

$$D_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad D_2 = 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{2\pi/3} & 0 \\ 0 & 0 & e^{4\pi/3} \end{pmatrix} \qquad D_3 = 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{4\pi/3} & 0 \\ 0 & 0 & e^{2\pi/3} \end{pmatrix}$$

$$D_4 = \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & -\frac{2\sqrt{6}}{3} & 0 \\ 0 & 0 & -\frac{2\sqrt{6}}{3} \end{pmatrix} D_5 = \begin{pmatrix} -\frac{2\sqrt{6}}{3} & 0 & 0 \\ 0 & \sqrt{6} & 0 \\ 0 & 0 & -\frac{2\sqrt{6}}{3} \end{pmatrix} D_6 = \begin{pmatrix} -\frac{2\sqrt{6}}{3} & 0 & 0 \\ 0 & -\frac{2\sqrt{6}}{3} & 0 \\ 0 & 0 & \sqrt{6} \end{pmatrix}.$$

Remark It bears mentioning that in [2], (where Conjecture 4.1.1 first appears), there is an incorrect result that $|sc(M)| \leq 4$ for any 3×3 circulant matrix M. The problem arises when, in the proof, the authors erroneously argue that the entries of a certain circulant $C = circ(x, y, z)$ are all real. The error occurs when the authors assume that both x and

the vector $(\overline{x}y, x\overline{z}, y\overline{z})^T$ are real. While it is acceptable (in their proof) to assume that one of these is real, we cannot assume both.

4.4 A new upper bound for 3×3 real matrices

Now that we have a matrix that disproves Conjecture 4.1.1, one might wonder: Can we adjust the bound in Conjecture 4.1.1 so that the statement becomes true? We show that for 3×3 real matrices, the answer is yes.

Theorem 4.4.1 *Let A be a real, positive definite 3×3 matrix. Then $|sc(A)| \leq 6$.*

Remark The matrix in Example 4.3.1 shows that this bound is, in fact, attainable.

We will break the proof of Theorem 4.4.1 into three cases. We first prove that the inequality is strict when A has zero entries:

Proposition 4.4.2 *Let $A = (a_{ij})$ be a 3×3 real, positive definite matrix with zero entries $a_{\alpha\beta} = a_{\beta\alpha} = 0$, for some $1 \leq \alpha \neq \beta \leq 3$. Then $|sc(A)| \leq 4$, and every matrix in $sc(A)$ is real.*

Proof As A has (at least) 2 zero entries, it is permutationally similar to a tridiagonal matrix (that is, $P^{-1}AP = P^*AP$ is tridiagonal, for some permutation matrix P). Now, suppose that D scales the tridiagonal matrix P^*AP. Then D^*P^*APD is doubly stochastic, and $PD^*P^*APDP^*$ has the same rows and columns as D^*P^*APD (although they will be permuted). Thus, PDP^* scales A if and only if D scales the tridiagonal matrix P^*AP. This

tells us that A and P^*AP have the same number of scalings. By Theorem 4.2.2, a tridiagonal matrix has at most 2^{n-1} scalings. It is easy to see that a DQS tridiagonal matrix must be real. Our result follows. □

There are two more cases to consider as we prove Theorem 4.4.1. Proposition 4.4.3 and Corollary 4.4.4 will place an upper bound on scalings that do not possess a certain property, and Proposition 4.4.5 will provide a bound on the remaining scalings. We will need the following easy fact:

Observation: *Let z_1, z_2 be complex numbers satisfying $z_1 z_2 \in \mathbb{R}$. Then either $z_2 = k\overline{z_1}$ for some $k \in \mathbb{R}$, or $z_1 = 0$.*

Proposition 4.4.3 *Let $A = (a_{ij})$ be a 3×3 real positive definite matrix with no zero entries, and suppose $D = diag(d_1, d_2, d_3) \in \mathbb{M}_3(\mathbb{C})$ is an invertible diagonal matrix such that D^*AD is doubly quasi-stochastic. If $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} \neq 0$, then (up to equivalence) D is real.*

Proof Assume by a suitable rotation (which will not affect our scaling as the resultant diagonal matrix will be equivalent to our original) that d_1 is positive. Let $(D^*AD)_i$ denote the i-th column sum of D^*AD (that is, the i-th entry of eD^*AD, where e is the all ones row vector $(1, ..., 1)$). We consider the first column of the DQS matrix D^*AD:

$$(D^*AD)_1 = a_{11}|d_1|^2 + a_{12}d_1\bar{d}_2 + a_{13}d_1\bar{d}_3 = 1.$$

The first term is of course real, and this means that $d_1(a_{12}\bar{d}_2 + a_{13}\bar{d}_3)$ must be real as well. By the Observation above, this means that either $a_{12}d_2 + a_{13}d_3 = 0$ or $d_1 = k(a_{12}d_2 + a_{13}d_3)$

for some real k. We will consider each of these cases separately.

Case 1: Suppose $a_{12}d_2 + a_{13}d_3 = 0$. Then $d_2 = \frac{-a_{13}d_3}{a_{12}}$. Now consider the second column of our matrix:

$$\begin{aligned} 1 &= (D^*AD)_2 \\ &= a_{12}d_1d_2 + a_{22}|d_2|^2 + a_{23}d_2\bar{d_3} \\ &= a_{22}|d_2|^2 + d_2(a_{12}d_1 + a_{23}\bar{d_3}) \\ &= a_{22}|d_2|^2 - a_{13}d_1d_3 - \frac{a_{23}a_{13}|d_3|^2}{a_{12}} \\ &= a_{22}|d_2|^2 - \frac{a_{23}a_{13}|d_3|^2}{a_{12}} - a_{13}d_1d_3, \end{aligned}$$

whence we see that $a_{13}d_1d_3$ must be real. As a_{13} and d_1 are nonzero real numbers, we know that d_3 must be real as well. Lastly, as $d_2 = \frac{-a_{13}d_3}{a_{12}}$, d_2 must also be real. Hence D is a real matrix, as desired.

Case 2: Now suppose $a_{12}d_2 + a_{13}d_3 \neq 0$, so that

$$d_1 = k(a_{12}d_2 + a_{13}d_3) = k(a_{12}\bar{d_2} + a_{13}\bar{d_3}),$$

for some real (nonzero) k (where the second equality comes from the fact that d_1 is real). Again, we consider the second column:

$$\begin{aligned} 1 &= (D^*AD)_2 \\ &= a_{12}d_1\bar{d_2} + a_{22}|d_2|^2 + a_{23}d_2\bar{d_3} \\ &= a_{12}k(a_{12}\bar{d_2} + a_{13}\bar{d_3})d_2 + a_{22}|d_2|^2 + a_{23}d_2\bar{d_3} \\ &= a_{22}|d_2|^2 + (a_{12})^2k|d_2|^2 + (ka_{12}a_{13} + a_{23})d_2\bar{d_3}, \end{aligned}$$

from which we see that $(ka_{12}a_{13} + a_{23})d_2\bar{d_3} \in \mathbb{R}$. Note that this cannot be zero, as $ka_{13}a_{12} + a_{23} = 0$ would mean $k = -\frac{a_{23}}{a_{13}a_{12}}$, which (combined with our expression for d_1, above) would yield $d_2 = -\frac{a_{13}}{a_{23}}d_1 - \frac{a_{13}}{a_{12}}d_3$, contradicting our assumption on D that $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} \neq 0$.

Thus we have $d_2 = ld_3$, for some $l \in \mathbb{R}\backslash\{0\}$. Plugging this into our expression for d_1, we have $d_1 = k(a_{12}l + a_{13})d_3$. As d_1 cannot be zero if D is to scale A, $k(a_{12}l + a_{13}) \neq 0$, and d_3 must be real. We know that $d_2 = ld_3$, and so d_2 is also real. □

Proposition 4.4.3 combined with Proposition 4.2.4 gives us the following.

Corollary 4.4.4 *Let $A = (a_{ij})$ be a 3×3 real positive definite matrix with no zero entries. Up to equivalence, there are at most 4 complex diagonal matrices $D = diag(d_1, d_2, d_3)$ which satisfy $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} \neq 0$ and have the property that D^*AD are doubly quasi-stochastic. These diagonal matrices may all be taken to be real.*

We now consider scalings that do satisfy this condition.

Proposition 4.4.5 *Let $A = (a_{ij})$ be a 3×3 real positive definite matrix with no zero entries.*

Up to equivalence, there are at most 2 diagonal matrices $D = diag(d_1, d_2, d_3) \in \mathbb{M}_3(\mathbb{C})$ *that scale* A *and satisfy* $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} = 0$.

Proof Let us rewrite our condition in terms of d_2:

$$d_2 = -\frac{a_{13}}{a_{23}}d_1 - \frac{a_{13}}{a_{12}}d_3.$$

Again let us assume that d_1 is real and positive by multiplying D by an appropriate scalar, if necessary. We consider the first column of D^*AD:

$$\begin{aligned} 1 = (D^*AD)_1 &= a_{11}|d_1|^2 + a_{12}d_1\bar{d_2} + a_{13}d_1\bar{d_3} \\ &= a_{11}|d_1|^2 + a_{12}d_1(-\frac{a_{13}}{a_{23}}\bar{d_1} - \frac{a_{13}}{a_{12}}\bar{d_3}) + a_{13}\bar{d_3}d_1 \\ &= (a_{11} - \frac{a_{12}a_{13}}{a_{23}})|d_1|^2 - a_{13}d_1\bar{d_3} + a_{13}d_1\bar{d_3} \\ &= (a_{11} - \frac{a_{12}a_{13}}{a_{23}})|d_1|^2. \end{aligned}$$

Rearranging, we get

$$|d_1|^2 = \frac{1}{a_{11} - \frac{a_{12}a_{13}}{a_{23}}} = \frac{a_{23}}{a_{11}a_{23} - a_{12}a_{13}}. \quad (1)$$

Hence, $|d_1|^2$ is fixed. As d_1 is real and positive, this fixes d_1. Now consider the third column:

$$\begin{aligned} 1 = (D^*AD)_3 &= a_{13}\bar{d_1}d_3 + a_{23}\bar{d_2}d_3 + a_{33}|d_3|^2 \\ &= a_{13}\bar{d_1}d_3 + d_3a_{23}(-\frac{a_{13}}{a_{23}}\bar{d_1} - \frac{a_{13}}{a_{12}}\bar{d_3}) + a_{33}|d_3|^2 \\ &= a_{13}\bar{d_1}d_3 - a_{13}\bar{d_1}d_3 - \frac{a_{13}a_{23}}{a_{12}}|d_3|^2 + a_{33}|d_3|^2 \\ &= (a_{33} - \frac{a_{13}a_{23}}{a_{12}})|d_3|^2. \end{aligned}$$

Isolating $|d_3|^2$ yields

$$|d_3|^2 = \frac{1}{a_{33} - \frac{a_{13}a_{23}}{a_{12}}} = \frac{a_{12}}{a_{33}a_{12} - a_{13}a_{23}}, \tag{3}$$

and we have that $|d_3|^2$ is fixed as well.

Lastly, we look at the second column:

$$\begin{aligned} 1 = (D^*AD)_2 &= a_{12}\bar{d_1}d_2 + a_{22}|d_2|^2 + a_{23}\bar{d_3}d_2 \\ &= -\frac{a_{12}a_{13}}{a_{23}}|d_1|^2 - a_{13}d_1d_3 + a_{22}|d_2|^2 - a_{13}d_1\bar{d_3} - \frac{a_{23}a_{13}}{a_{12}}|d_3|^2 \\ &= 1 - a_{11}|d_1|^2 - 2a_{13}d_1Re(d_3) + a_{22}|\frac{a_{13}}{a_{23}}d_1 + \frac{a_{13}}{a_{12}}d_3|^2 - \frac{a_{23}a_{13}}{a_{12}}|d_3|^2, \end{aligned}$$

and hence

$$0 = -a_{11}|d_1|^2 - 2a_{13}d_1 Re(d_3) + a_{22}(\frac{a_{13}^2}{a_{23}^2}|d_1|^2 + 2\frac{a_{13}^2}{a_{23}a_{12}}d_1 Re(d_3) + \frac{a_{13}^2}{a_{12}^2}|d_3|^2) - \frac{a_{23}a_{13}}{a_{12}}|d_3|^2.$$

Rearranging, we obtain:

$$2a_{13}d_1(1 - \frac{a_{22}a_{13}}{a_{23}a_{12}})Re(d_3) = (\frac{a_{22}a_{13}^2}{a_{23}^2} - a_{11})|d_1|^2 + \frac{a_{13}}{a_{12}}(\frac{a_{22}a_{13}}{a_{12}} - a_{23})|d_3|^2. \quad (2)$$

Note that the coefficient on $Re(d_3)$ is not zero. To see this, suppose the left side is zero. This means that $a_{22}a_{13} = a_{23}a_{12}$. Substituting this into the right hand side, we obtain the following:

$$\begin{aligned} 0 &= (\frac{a_{23}a_{12}a_{13}}{a_{23}^2} - a_{11})|d_1|^2 + \frac{a_{13}}{a_{12}}(\frac{a_{23}a_{12}}{a_{12}} - a_{23})|d_3|^2 \\ &= (\frac{a_{12}a_{13}}{a_{23}} - a_{11})|d_1|^2 + 0. \end{aligned}$$

For this to be consistent, we need the coefficient on $|d_1|^2$ also to be zero. This means that $\frac{a_{12}a_{13}}{a_{23}} = a_{11}$. But looking back to (1), we see that this cannot be the case. Therefore the coefficient on $Re(d_3)$ is not zero.

Substituting the value obtained for d_1 and $|d_3|^2$ into (2), we fix the value for Re(d_3).

Given $|d_3|^2$ and $\mathrm{Re}(d_3)$, we have (maximum) two choices for d_3:

$$d_3 = Re(d_3) \pm \sqrt{(|d_3|^2 - (Re(d_3))^2)}i.$$

d_2 is of course fixed as soon as we have d_1 and d_3, by our original assumption on D.

Hence, we have maximum two (possibly complex) diagonal matrices D that scale A and satisfy $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} = 0$. □

We are now in a position where we can easily prove Theorem 4.4.1.

Proof of Theorem 4.4.1 Suppose that A has at least one zero entry. Then by Proposition 4.4.2, $|sc(A)| \leq 4$.

If A has no zero entries, then Corollary 4.4.4 gives us a maximum of 4 scalings that arise from diagonal matrices satisfying $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} \neq 0$ and Proposition 4.4.5 gives us a maximum of 2 scalings that arise from diagonal matrices satisfying $\frac{d_1}{a_{23}} + \frac{d_2}{a_{13}} + \frac{d_3}{a_{12}} = 0$. Hence, $|sc(A)| \leq 6$. □

Upon closer inspection of the proofs of Proposition 4.4.3 and Proposition 4.4.5, one can see that we have actually discovered a condition that guarantees that $sc(A)$ contains only real matrices.

Corollary 4.4.6 *A real* 3×3 *positive definite matrix* $A = (a_{ij})$ *will have all real scalings (and hence* $|sc(A)| \leq 4$*) if either of the following holds:*

(1^*) $C = A \circ A^{-1}$ *has at least one nonnegative off-diagonal entry (i.e.* $c_{ij} \geq 0$ *for some* $i \neq j$*).*

(2*)
$$1 \leq \left| \frac{\left(\frac{a_{22}a_{13}}{a_{23}^2} - \frac{a_{11}}{a_{13}}\right)\sqrt{\frac{a_{33}-\frac{a_{13}a_{23}}{a_{12}}}{a_{11}-\frac{a_{12}a_{13}}{a_{23}}}} + \left(\frac{a_{22}a_{13}}{a_{12}^2} - \frac{a_{23}}{a_{12}}\right)\sqrt{\frac{a_{11}-\frac{a_{12}a_{13}}{a_{23}}}{a_{33}-\frac{a_{13}a_{23}}{a_{12}}}}}{2\left(1-\frac{a_{22}a_{13}}{a_{23}a_{12}}\right)} \right|.$$

Proof Let us begin by proving sufficiency of (1*). Recall that the off-diagonal entries of A^{-1} can be expressed as

$$(A^{-1})_{ij} = \frac{1}{det(A)}(a_{ik}a_{kj} - a_{ij}a_{kk}), \qquad k \neq i, j.$$

As A is positive definite (and hence has positive determinant), we can thus re-write the condition in (1*) as:

$a_{ij}(a_{ik}a_{kj} - a_{ij}a_{kk}) \geq 0$, for all $1 \leq i, j, k \leq 3$, where i, j, k are mutually distinct.

If $a_{ij} = 0$, then Proposition 4.4.2 tells us that we have only real scalings. If $a_{ij} \neq 0$, we divide both sides by the positive number a_{ij}^2 to obtain

$$\frac{(a_{ik}a_{kj} - a_{ij}a_{kk})}{a_{ij}} \geq 0. \tag{4}$$

Now, suppose for the purposes of contradiction that (4) holds, but A has complex scalings. From the proof of Proposition 4.4.5, we know that (1), (2), and (3) must hold. If $(i, j, k) = (1, 2, 3)$ (or $(i, j, k) = (2, 1, 3)$, by symmetry), then (4) combines with (3) to yield $-\frac{1}{|d_3|^2} \geq 0$. Clearly this is impossible, and we have our contradiction. Similarly, if $(i, j, k) = (2, 3, 1)$ or $(i, j, k) = (3, 2, 1)$, (4) combines with (1) to yield $-\frac{1}{|d_1|^2} \geq 0$. It is easy to show that $(i, j, k) = (1, 3, 2)$ or $(i, j, k) = (3, 1, 2)$ will imply that $-\frac{1}{|d_2|^2} \geq 0$. (This can be seen

by reworking the proof of Proposition 4.4.5, where the condition is written in terms of d_3 instead of d_2, or simply by noting that any condition such as (1) or (3) which restricts the scalings of A must necessarily be invariant with respect to permutation-similarity.) Thus, if (1*) holds, we cannot have complex scalings.

Let us now consider (2*). Again, suppose A has complex scalings, and hence (1), (2), and (3) hold. Rearranging (2), we obtain:

$$\begin{aligned} Re(d_3) &= \frac{\left(\frac{a_{22}a_{13}^2}{a_{23}^2} - a_{11}\right)|d_1|^2 + \frac{a_{13}}{a_{12}}\left(\frac{a_{22}a_{13}}{a_{12}} - a_{23}\right)|d_3|^2}{2a_{13}d_1\left(1 - \frac{a_{22}a_{13}}{a_{23}a_{12}}\right)} \\ &= \frac{\left(\frac{a_{22}a_{13}}{a_{23}^2} - \frac{a_{11}}{a_{13}}\right)d_1 + \left(\frac{a_{22}a_{13}}{a_{12}^2} - \frac{a_{23}}{a_{12}}\right)\frac{|d_3|^2}{d_1}}{2\left(1 - \frac{a_{22}a_{13}}{a_{23}a_{12}}\right)}. \end{aligned} \tag{5}$$

Taking the absolute value of (5) and dividing by $|d_3|$ yields:

$$\left|\frac{Re(d_3)}{d_3}\right| = \left|\frac{\left(\frac{a_{22}a_{13}}{a_{23}^2} - \frac{a_{11}}{a_{13}}\right)\frac{d_1}{|d_3|} + \left(\frac{a_{22}a_{13}}{a_{12}^2} - \frac{a_{23}}{a_{12}}\right)\frac{|d_3|}{d_1}}{2\left(1 - \frac{a_{22}a_{13}}{a_{23}a_{12}}\right)}\right|.$$

Substituting our values in (1) and (3) in for d_1 and $|d_3|$ gives the right hand side of (2*). That is, the right hand side of (2*) is $\left|\frac{Re(d_3)}{d_3}\right|$. If d_3 is to exist and be complex, it must satisfy $\left|\frac{Re(d_3)}{d_3}\right| < 1$. Thus, our complex scalings do not exist if (2*) holds. (It is easy to see that if d_3 is real and d_1 is real, we must have that d_2 is real by examining the first row sum of D^*AD.)

If neither of our conditions hold, then we will necessarily have 2 complex scalings (ob-

tained by simply filling in the values of (a_{ij}) into (1), (2) and (3)). $\square$

Remark (1) Given the relative complexity of (2^*), one might hope that (2^*) implies (1^*), which would allow us to remove this condition from Corollary 4.4.6. Unfortunately, this is not the case. Indeed, consider the following DQS matrix:

$$B = \begin{pmatrix} 1.9 & -1 & 0.1 \\ -1 & 1.8 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}.$$

This matrix satisfies (2^*), and hence has only real scalings. However, it does not satisfy (1^*).

(2) Similarly, we can show that property (1^*) cannot be removed from Corollary 4.4.6 by examining the DQS matrix:

$$A = \begin{pmatrix} 0.422 & 0.16 & 0.418 \\ 0.16 & 0.673 & 0.167 \\ 0.418 & 0.167 & 0.415 \end{pmatrix}.$$

This matrix satisfies (1^*), as the (1,2)-entry of $A \circ A^{-1}$ is positive. However, it does not satisfy (2^*), as the value of the right hand side of (2^*) works out to be slightly less than 1 (~ 0.9985).

Let us take a moment to recapitulate the results that we have seen thus far this chapter.

We know that if A is a 2×2 positive definite matrix, $|sc(A)| \leq 2$. We also know that if A is a positive definite tridiagonal matrix, then $|sc(A)| \leq 2^{n-1}$. Lastly, if A is a 3×3 positive definite real matrix, then $|sc(A)| \leq 6$. With these results in mind, one might be tempted to modify Conjecture 4.1.1 and suggest a new upper bound on $|sc(A)|$ for general $n \times n$ positive definite matrices, or perhaps a bound for real $n \times n$ positive definite matrices. We will now show that such bounds do not exist for any dimension higher than $n = 3$.

4.5 Matrices with infinitely many scalings

In this section, we will construct an $n \times n$ matrix that has infinitely many scalings for all $n \geq 4$. Once again, our example will take the form of a real, circulant matrix:

Theorem 4.5.1 *Let $n \geq 4$ and suppose $C \in \mathbb{M}_n(\mathbb{R})$ is the following positive definite circulant matrix*

$$C = circ(a, b, b, b...b) = \begin{pmatrix} a & b & b & b & b & ... & b \\ b & a & b & b & b & ... & b \\ & & & \ddots & & & \\ b & b & b & b & ... & b & a \end{pmatrix},$$

where a and b satisfy the conditions $a > 0$, $\frac{-a}{n-1} < b < a$, and $b \neq 0$. Then $|sc(C)|$ is not

finite.

Proof For any complex number z of modulus 1, define E_z to be the diagonal matrix

$$E_z = \begin{pmatrix} z & 0 & 0 & 0 & 0 & ... & 0 \\ 0 & -z & 0 & 0 & 0 & ... & 0 \\ 0 & 0 & \omega & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & \omega^2 & 0 & ... & \\ 0 & 0 & 0 & 0 & \ddots & & 0 \\ 0 & ... & 0 & 0 & 0 & 0 & \omega^{n-2} \end{pmatrix},$$

where ω is the $(n-2)$th root of unity. (i.e. $\omega = e^{\frac{2\pi i}{n-2}}$). Next, let $D_z = \frac{1}{\sqrt{a-b}} E_z$ (this is of course well-defined, as C is positive definite and hence $a > b$). We claim that D_z scales C.

Indeed, $(D_z^* C D_z) = \frac{1}{a-b}(E_z^* C E_z)$, so we need only show that $E_z^* C E_z$ has row sums equal to $a - b$. We observe that

$$E_z^* C E_z = \begin{pmatrix} |z|^2 a & -|z|^2 b & \overline{z}\omega b & \overline{z}\omega^2 b & ... & \overline{z}\omega^{n-2} b \\ -|z|^2 b & |z|^2 a & -\overline{z}\omega b & \overline{z}\omega^2 b & ... & -\overline{z}\omega^{n-2} b \\ \overline{\omega} z b & -\overline{\omega} z b & |\omega|^2 a & \overline{\omega}\omega^2 b & ... & \overline{\omega}\omega^{n-2} b \\ & & & \ddots & & \\ & & & & & \\ & & & & & \\ \overline{\omega^{n-2}} z b & -\overline{\omega^{n-2}} z b & \overline{\omega^{n-2}}\omega b & ... & \overline{\omega^{n-2}}\omega^{n-3} b & |\omega^{n-2}|^2 a \end{pmatrix}.$$

Now it is simply a matter of investigating each row sum.

$$(E_z^* C E_z)_1 = |z|^2 a - |z|^2 b + \overline{z} b(\omega + \omega^2 + ... + \omega^{n-2}).$$

Of course, the sum of all $(n-2)$th roots of unity is zero, and $|z|^2 = 1$, by assumption on z. Hence this row sum is indeed $a - b$.

The terms in the second row are exactly the same as the first, so we inspect the third row next:

$$(E_z^* C E_z)_3 = \overline{\omega}(zb - zb) + |\omega|^2 a + \overline{\omega} b(\omega^2 + ... + \omega^{n-2}) = a + \overline{\omega} b(0 - \omega) = a - b,$$

where we have again used the fact that the roots of unity sum to 0, and $|\omega|^2 = 1$.

It is easy to see that the other rows (rows 4 through n) will all sum to $a - b$, by the exact same reasoning. i.e. Let $4 \leq r \leq n$. Then

$$(E_z^* C E_z)_r = \overline{\omega^{r-2}}(zb - zb) + |\omega^{r-2}|^2 a + b\overline{\omega^{r-2}}\Big(\sum_{\substack{m=1 \\ m \neq r-2}}^{n-2} \omega^m\Big) = 0 + a + b\overline{\omega^{r-2}}(0 - \omega^{r-2}) = a - b.$$

Hence, D_z scales C to a DQS matrix, for all complex z of modulus 1. Further, different choices of z will necessarily give rise to different scalings (for example, the $(n, 1)$ entry of $D_z^* C D_z$ is $\left(\frac{\overline{\omega^{n-2}} z b}{a-b}\right)$, and $b \neq 0$, by assumption). Thus $sc(C)$ contains infinitely many unique scalings, corresponding to each choice of z. $\square$

Remark As mentioned in Chapter 1, the interest in complex scalings arose in part due to their connection with the geometric measure of entanglement (Theorem 1.3.3). Once our scalings were obtained, one wanted to investigate the scalings' permanents. With this in mind, it is worth noting that while we have constructed infinitely many scalings of C, each of these scalings will have the same permanent (they will have all permanent $\frac{per(C)}{(a-b)^n}$).

As we've seen, it will not be possible to achieve an upper bound on $|sc(A)|$, as there are matrices with infinitely many scalings. This suggests that it might be difficult to find all elements of $sc(A)$ and then calculate the permanent of each one. However, Theorem 1.3.3 does not in fact require us to find all scalings. We need only the scaling with minimal permanent. Our next chapter will discuss how we can find such a scaling without finding all elements of $sc(A)$.

Chapter 5

Scalings of Extremal Permanent

5.1 Introduction

Theorem 1.3.3 demonstrates that we can calculate the geometric measure of entanglement of a Slater permanent $|\phi\rangle$ by taking the Gram matrix $G_{|\phi\rangle}$ and finding the element of $sc(G_{|\phi\rangle})$ that has smallest permanent. We introduce the following terminology:

Definition Given a positive definite doubly quasi-stochastic matrix M and associated equivalence class $sc(M)$, we say that M is *minimally (maximally) scaled* if $per(M) \leq per(C)$ ($per(M) \geq per(C)$) for all $C \in sc(M)$. We will denote the set of minimally (maximally) scaled $n \times n$ matrices as $MinSc_n$ ($MaxSc_n$).

Motivated by Theorem 1.3.3, the authors of [2] posed the following question:

Question 5.1.1 ([2], Question 4.1) *What is the structure of $MinSc_n$ and $MaxSc_n$? What are the element(s) in $MinSc_n$ which have the largest permanent? Which are the element(s)*

of $MaxSc_n$ which have smallest permanent?

It is clear why they were interested in $MinSc_n$. This set has a direct application to Theorem 1.3.3. It is not so obvious why we concern ourselves with $MaxSc_n$. The answer lies with the following observation:

Proposition 5.1.2 *Let A be a positive definite doubly quasi-stochastic matrix. Then A is maximally scaled if and only if A^{-1} is minimally scaled.*

Before we prove this result, we will need the following lemma (recall that we say that a diagonal matrix D *scales* A if D^*AD is DQS):

Lemma 5.1.3 *Let A be a positive definite doubly quasi-stochastic matrix. Then A is in $MaxSc_n$ ($MinSc_n$) if and only if for every D that scales A, $|det(D)| \leq 1$ ($|det(D)| \geq 1$).*

Proof Suppose A, B are in the same equivalence class $sc(A)$. Then there exists a diagonal matrix D such that $D^*AD = B$. One can easily verify that $b_{jk} = \overline{d_j} d_k a_{jk}$. Thus, for any $\sigma \in S_n$, the corresponding summand in the expression for the permanent of B can be written as:

$$\prod_{j=1}^{n} \overline{d_j} d_{\sigma(j)} a_{j\sigma(j)} = \prod_{j=1}^{n} |d_j|^2 a_{j\sigma(j)} = |det(D)|^2 \prod_{j=1}^{n} a_{j\sigma(j)}.$$

Thus, $per(B) = |det(D)|^2 per(A)$, and our result follows easily. □

Remark It is immediate from Lemma 5.1.3 that a maximally (minimally) scaled matrix M also has the largest (smallest) determinant in its equivalence class $sc(M)$.

The proof of Proposition 5.1.2 follows easily:

Proof of Proposition 5.1.2 Firstly, note that the positive definite matrix A^{-1} is certainly doubly quasi-stochastic, as $Ae = e$ if and only if $A^{-1}e = e$. Now suppose that $B = D^*A^{-1}D$ is a scaling of A^{-1}. Then taking inverses, $B^{-1} = D^{-1}A(D^*)^{-1}$ is a scaling of A. This shows us that D scales A^{-1} if and only if $(D^*)^{-1}$ scales A. Applying Lemma 5.1.3, A^{-1} is minimally scaled if and only if $|det(D)| \geq 1$ for all such D, if and only if $|det(D^*)^{-1}| \leq 1$ if and only if A is maximally scaled. □

We would like to achieve a bound on the permanent of elements from $MinSc_n$ (as this would, in turn, give us a bound on the GME of all Slater permanents in the corresponding tensor space). In conversation with the first author of [2], the following conjecture was made:

Conjecture 5.1.4 *Let A be a positive definite doubly quasi-stochastic matrix. If A is an element of $MinSc_n$, then $per(A) \leq 1$, with equality if and only if $A = I_n$, the identity matrix.*

In light of Proposition 5.1.2, one approach for determining the veracity of Conjecture 5.1.4 reveals itself. We note the following:

Proposition 5.1.5 *Let A be a positive definite matrix and suppose that $per(A) \leq 1$. Then $per(A^{-1}) \geq 1$.*

Proof As A is positive definite, there exists a positive definite matrix S such that $A = SS^*$ (and $A^{-1} = (S^{-1})^*S^{-1}$). Upon application of Theorem 1.2.4, it must be the case that

$$1 = |per(SS^{-1})|^2 \leq per(SS^*)per((S^{-1})^*S^{-1}) = per(A)per(A^{-1}),$$

and our result follows. □

Proposition 5.1.5 (with Proposition 5.1.2) shows us that we could falsify Conjecture 5.1.4 if we could find an element of $MaxSc_n$ with permanent less than one. This motivates the following conjecture (also originating from conversation with the first author of [2]):

Conjecture 5.1.6 *Let A be a positive definite doubly quasi-stochastic matrix. If A is an element of $MaxSc_n$, then $per(A) \geq 1$, with equality if and only if $A = I_n$.*

In [2], the authors fully answered Question 5.1.1 in the 2×2 case. In addition to Proposition 4.2.1 (which shows $|sc(A)| \leq 2$ in this case), they proved the following:

Proposition 5.1.7 ([2]) *Let A be a 2×2 positive definite matrix. Then $sc(A)$ consists entirely of real matrices, exactly one of which is doubly-stochastic. This doubly-stochastic matrix is the minimal scaling.*

Remark As the identity matrix is the positive definite doubly-stochastic matrix with maximal permanent, Proposition 5.1.7 shows that Conjecture 5.1.4 holds for 2×2 matrices. Conjecture 5.1.6 necessarily follows by Proposition 5.1.5.

In this chapter, we will further investigate Question 5.1.1. In particular, we will prove a number of results on the structures of $MaxSc_n$ and $MinSc_n$, and then investigate the two conjectures introduced above. We will prove Conjecture 5.1.6 and then provide evidence for Conjecture 5.1.4. Lastly, we will use Theorem 1.3.3 to arrive at entanglement bounds for certain Slater permanents. With the exception of Section 5.4, Section 5.7, Proposition 5.3.3,

Proposition 5.6.7, and Corollary 5.6.8, all of the results in this section have been collected in [49].

5.2 A characterization of $MaxSc_n$

Suppose we wish to know whether or not a DQS matrix is maximally or minimally scaled. As seen in the previous chapter, it often does not seem feasible to simply find all scalings and check the permanent of each one. For this reason, we must discover other methods for identifying elements of $MinSc_n$ and $MaxSc_n$. In this section, we will identify a characterization of the set of maximally scaled matrices ($MaxSc_n$) and use this characterization to show that for any given positive definite matrix A, $sc(A)$ must contain a maximal scaling and a minimal scaling.

We introduce the following notation:

$$\Sigma = \{y \in \mathbb{C}^n : \prod_{j=1}^{n} |y_j| = 1\}.$$

The following result can be found in the proof of [[3], Lemma 2.9], or by examining our proof of Thereom 2.4.1:

Lemma 5.2.1 *Let A be a positive definite matrix, and suppose $d \in \Sigma$ is the vector that minimizes y^*Ay over Σ. Let $\frac{d^*Ad}{n} = \lambda$. Then $D = \frac{1}{\sqrt{\lambda}}$ $(diag(d))$ scales A.*

Remark The vector d is always guaranteed to exist, as shown in [3] (and our proof of Theorem 2.4.1).

Before we introduce the main result of this section, we make the following observation which will be useful in the proof to follow:

Observation Suppose A is an $n \times n$ positive definite matrix, and suppose D is a positive diagonal matrix that scales A. Then

$$D^*AD = DAD = D(Re(A))D + iD(Im(A))D,$$

whence we see that D must be the unique positive diagonal matrix that scales $Re(A)$ (as guaranteed to exist by Theorem 2.2.2 and guaranteed to be unique by Proposition 4.2.3).

Theorem 5.2.2 *Let M be a positive definite DQS matrix. Then M is maximally scaled if and only if*

$$k := \min_{y \in \Sigma} \frac{y^*My}{n} = 1.$$

Proof Suppose $k = 1$. We wish to show that M is maximally scaled. By Lemma 5.1.3, it suffices to show that for all diagonal D that scale M, we must have $|det(D)| \leq 1$. To this end, suppose D scales M, and let $D = PU$ be the polar decomposition of D (so that P is a positive diagonal matrix, and U is a diagonal unitary matrix). We note that

$$|det(D)| = |det(UP)| = |det(U)||det(P)| = |det(P)| = det(P).$$

Thus it suffices to show that $det(P) \leq 1$.

As D scales M, we have $D^*MD = PU^*MUP$, and we see that P is a positive diagonal

matrix that scales U^*MU. By the Observation above, this means that P is the unique positive diagonal matrix that scales $Re(U^*MU)$. By our proof of Theorem 2.4.1 (or by the proof of Theorem 2.2.2 found in [4]), $det(P) \leq 1$ iff $\lambda = \min_{x \in \Psi} \frac{x^T Re(U^*MU)x}{n} \geq 1$, where $\Psi = \{x \in \mathbb{R}^n_+ : \prod x_i = 1\}$.

To this end, let $x_c \in \Psi$ and let $y_c = Ux_c \in \Sigma$. Then, as x_c is real,

$$\frac{x_c^T Re(U^*MU)x_c}{n} = \frac{x_c^T U^*MUx_c}{n} = \frac{y_c^* My_c}{n} \geq k = 1,$$

where the inequality arises from our definition of k. Thus we have $\frac{x^T Re(U^*MU)x}{n}$ for all $x \in \Psi$, meaning $\lambda \geq 1$ and M is maximally scaled, as desired.

Now suppose that M is maximally scaled. We wish to show that $k = 1$. By Lemma 5.1.3, we know that if D scales M, we must have $|det(D)| \leq 1$. Suppose there exists $y_c \in \Sigma$ such that $\frac{y_c My_c}{n} < 1$. Then, by Lemma 5.2.1, There exists a vector $d \in \Sigma$, such that $D = \frac{1}{\sqrt{\lambda}} diag(d)$ scales M, where $\lambda = \frac{d^*Md}{n} \leq \frac{y_c^* My_c}{n} < 1$. Thus $|det(D)| = \frac{1}{\lambda^{n/2}} > 1$, contradicting the maximality of M. Thus $k \geq 1$. We complete the proof by noting that for $e = (1, 1, 1..., 1)^T$, $\frac{e^*Me}{n} = 1$ by the double quasi-stochasticity of M. □

Remark (1) As noted above, we will always have $e^*Me = n$ for DQS A. Hence to show that a Hermitian DQS matrix M is maximally scaled, it suffices to show that $\min_{y \in \Sigma} \frac{y^*My}{n} \geq 1$ (where positive definiteness follows from the easy fact that $k = 0$ if M is singular).

(2) We will often write y^*My as $\langle My, y \rangle$, the standard Euclidean inner product.

As noted in Theorem 4.5.1, there are cases where there are infinitely many scalings

in an equivalence class $sc(A)$. It may not, then, be immediately clear that every such equivalence class actually has a maximal and minimal scaling. It may seem possible that the set $\{l \in \mathbb{R} : l = per(B) \text{ for some } B \in sc(A)\}$ has an infimum (or supremum), but that infimum (supremum) is never attained. Theorem 5.2.2 allows us to show that this cannot the case.

Corollary 5.2.3 *Let A be an $n \times n$ positive definite matrix. Then there exists an $n \times n$ matrix $M_1 \in sc(A)$ such that $per(M_1) \geq per(B)$ for all $B \in sc(A)$, and an $n \times n$ matrix $M_2 \in sc(A)$ such that $per(M_2) \leq per(B)$ for all $B \in sc(A)$.*

Proof Let us prove the existence of M_1 first. If A is already maximally scaled, then we are done. Suppose A is not maximally scaled. Then by Theorem 5.2.2, there exists a $d \in \Sigma$ such that $\frac{d^*Ad}{n} < 1$ and by the remark following Lemma 5.2.1, we can choose this d to be the vector that minimizes $\frac{y^*Ay}{n}$ over Σ. Now define $M_1 = D^*AD$ to be the scaling obtained in the method outlined in Lemma 5.2.1 (i.e. $D = \frac{diag(d)}{\sqrt{\lambda}}$). We claim that M_1 is maximally scaled. Suppose not. By Theorem 5.2.2, this means that there exists a vector $z \in \Sigma$ such that $\frac{z^*M_1z}{n} < 1$. But this means that $\frac{z^*M_1z}{n} = \frac{1}{\lambda}\frac{(z \circ d)^*A(z \circ d)}{n} < 1$, (where $z \circ d$ is the Hadamard product of z and d) and hence $\frac{(z \circ d)^*A(z \circ d)}{n} < \lambda$. As $z \circ d$ is clearly in Σ, this contradicts the definition of d. Thus, M_1 must be maximally scaled. The existence of M_2 follows by taking inverses. $\square$

We may now investigate Question 5.1.1. We begin by considering the structure of $MaxSc_n$.

5.3 Topological properties of $MaxSc_n$

The characterization of $MaxSc_n$ given in Theorem 5.2.2 yields a number of results on the structure of $MaxSc_n$. We begin with the more immediate consequences of Theorem 5.2.2.

Proposition 5.3.1 *$MaxSc_n$ is closed, for all $n \in \mathbb{N}$.*

Proof Let $\{M_j\}_{j=1}^{\infty}$ be a sequence of maximally scaled $n \times n$ matrices, with limit M. It is clear that M is DQS, as the limit of DQS matrices. We claim that M must also be maximally scaled. Indeed, suppose that v is the vector that minimizes $\frac{y^*My}{n}$ over Σ. Then $\frac{v^*Mv}{n} = \lim_{j\to\infty} \frac{v^*M_jv}{n}$. As M_j is maximally scaled, we know that $\frac{v^*M_jv}{n} \geq 1$ for all $j \in \mathbb{N}$. Thus the same is true of the limit, $\frac{v^*Mv}{n} \geq 1$. By the remark following Theorem 5.2.2, M is maximally scaled, whence we see that $MaxSc_n$ contains all its limit points and must be closed. □

Another immediate result that will be of use to us as we prove Conjecture 5.1.6 is the convexity of $MaxSc_n$:

Corollary 5.3.2 *Let $n \in \mathbb{N}$, then $MaxSc_n$ is convex.*

Proof Let $A, B \in MaxSc_n$, and let $\lambda \in (0, 1)$. It is immediate that $C = \lambda A + (1 - \lambda)B$ is also doubly quasi-stochastic. Now,

$$\begin{aligned}
\min_{y\in\Sigma} \frac{y^*(\lambda A + (1-\lambda)B)y}{n} &= \min_{y\in\Sigma} \frac{\lambda y^* A y}{n} + \frac{(1-\lambda)y^* B y}{n} \\
&\geq \lambda \min_{y\in\Sigma} \frac{y^* A y}{n} + (1-\lambda) \min_{x\in\Sigma} \frac{x^* B x}{n} \\
&= \lambda + (1-\lambda) \\
&= 1.
\end{aligned}$$

Thus, C is maximally scaled by (the remark following) Theorem 5.2.2, as desired. □

It is clear from Theorem 5.2.2 that $MaxSc_n$ will have elements of arbitrarily large (spectral) norm. Indeed, take any maximally scaled matrix M and fix the eigenvectors while increasing any of the eigenvalues not associated with eigenvector $(1, 1..., 1)^T$. The resultant matrix will always remain maximally scaled (as can be verified by expressing all vectors $y \in \Sigma$ as a linear combination of the eigenvectors of M in the statement of Theorem 5.2.2). It may be less obvious, however, that our characterization allows us to give a positive lower bound on the spectrum of elements in $MaxSc_n$.

Proposition 5.3.3 *Let $A \in MaxSc_n$, and let $\lambda \in \sigma(A)$. Then $\lambda \geq \frac{0.1n}{10^{2m}}$, where*

$$m = \left\lceil log_{10} \left(\sqrt{2n} \left(\frac{(1.1)^{\frac{n}{2}}}{\left(.99^{\lfloor \frac{n}{2} \rfloor}\right)^{\frac{n}{2}}} + 1 \right) + 3\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) \right\rceil.$$

Proof For the purposes of contradiction, suppose that there exists an element $A \in MaxSc_n$ with smallest eigenvalue $\lambda_{min} < \frac{0.1n}{10^{2m}}$ and let v denote the associated (normalized) eigenvec-

tor. Recall that A is doubly quasi-stochastic, and so v can be taken to be orthogonal to $u = (1, 1..., 1)^T$, another eigenvector of A. Now, choose $\theta \in [0, 2\pi)$ so that $e^{i\theta}v$ has at least $\frac{n}{2}$ components with positive real part (and hence at most $\lfloor \frac{n}{2} \rfloor$ components has negative real part). Let $\omega = e^{i\theta}$ and define $w_j = u + 10^j \omega v$, where $j \in \mathbb{N}$. Let us denote the components of w_j by $w_j^{(k)}$.

We will show that for at least one j between $\left\lceil log_{10}\left(\sqrt{2n}\left(\frac{(1.1)^{\frac{n}{2}}}{\left(.99^{\lfloor \frac{n}{2} \rfloor}\right)^{\frac{n}{2}}} + 1\right)\right)\right\rceil$ and m, we must have

$$\frac{\langle Aw_j, w_j \rangle}{n \prod_{k=1}^{n} \left| w_j^{(k)} \right|^{\frac{2}{n}}} < 1, \tag{1}$$

which would contradict the maximality of A, as $\frac{w_j}{\prod_{k=1}^{n} \left| w_j^{(k)} \right|^{\frac{1}{n}}} \in \Sigma$.

Let us investigate the numerator of (1) first. Let us assume that j is in the range specified above. Then

$$\begin{aligned} \langle Aw_j, w_j \rangle &= \langle A\left(u + 10^j \omega v\right), u + 10^j \omega v \rangle \\ &= \langle Au, u \rangle + 10^{2j} \langle Av, v \rangle \\ &= n + 10^{2j} \lambda_{min} \\ &< n + 0.1n, \end{aligned} \tag{2}$$

where the last inequality follows from our definition of λ_{min}, and our bounds on j.

Thus to prove (1), it suffices to show that $\frac{1.1}{\prod_{k=1}^{n}\left|w_j^{(k)}\right|^{\frac{2}{n}}} < 1$. We now inspect the denominator and show that indeed, $\prod_{k=1}^{n}\left|w_j^{(k)}\right|^{\frac{2}{n}} > 1.1$, for at least one j in our range.

Let $v^{(k)}$ denote the k-th component in the eigenvector v. Without loss of generality, assume $v^{(1)} \geq v^{(2)} \geq v^{(3)}... \geq v^{(n)}$. We begin by showing that for all j in our range, $\left|w_j^{(1)}\right| \geq \frac{(1.1)^{n/2}}{\left(.99^{\lfloor n/2 \rfloor}\right)^{\frac{n}{2}}}$.

As $\|v\| = 1, \left|v^{(1)}\right|^2 \geq \frac{1}{n}$, and hence either $\left|Re\left(\omega v^{(1)}\right)\right|^2 > \frac{1}{2n}$, or $\left|Im\left(\omega v^{(1)}\right)\right|^2 \geq \frac{1}{2n}$.

If $\left|Im\left(\omega v^{(1)}\right)\right|^2 \geq \frac{1}{2n}$, then $\left|w_j^{(1)}\right| = \left|1 + 10^j \omega v^{(1)}\right| > \left|10^j Im\left(\omega v^{(1)}\right)\right| \geq \left(\frac{(1.1)^{\frac{n}{2}}}{\left(.99^{\lfloor \frac{n}{2} \rfloor}\right)^{\frac{n}{2}}} + 1\right)$ and hence $\left|w_j^{(1)}\right| \geq \frac{(1.1)^{n/2}}{\left(.99^{\lfloor n/2 \rfloor}\right)^{\frac{n}{2}}}$, as desired.

If $\left|Im(v^{(1)})\right|^2 < \frac{1}{2n}$, then we have $\left|Re\left(\omega v^{(1)}\right)\right| \geq \frac{1}{\sqrt{2n}}$, and

$$\left|w_j^{(1)}\right| = \left|1 + \omega 10^j v^{(1)}\right| \geq \left|1 - \left|Re\left(\omega 10^j v^{(1)}\right)\right|\right| \geq \left|1 - \left(\frac{(1.1)^{\frac{n}{2}}}{\left(.99^{\lfloor \frac{n}{2} \rfloor}\right)^{\frac{n}{2}}} + 1\right)\right| = \frac{(1.1)^{n/2}}{\left(.99^{\lfloor n/2 \rfloor}\right)^{\frac{n}{2}}}.$$

Now, we show that for some j in our range, the product of the rest of the multiplicands in the denominator is at least $\left(.99^{\lfloor n/2 \rfloor}\right)^{\frac{n}{2}}$ (that is, $\prod_{k=2}^{n}\left|w_j^{(k)}\right| \geq \left(.99^{\lfloor n/2 \rfloor}\right)^{\frac{n}{2}}$).

Firstly, note that at least half of the other components of w_j have real part greater than (or equal) 1, by our assumption on ω, and hence they have modulus at least one.

For the remaining (at most) $\lfloor \frac{n}{2} \rfloor$ components, let us fix j and suppose that $\left|w_j^{(l)}\right| < .99$, for some l. Then $-.99 < 1 + Re\left(10^j \omega v^{(l)}\right) < .99$. Hence

$$-1.99 < Re\left(10^j \omega v^{(l)}\right) < -0.01. \tag{3}$$

At this point, it is clear that (3) can only hold for at most three different $j \in \mathbb{N}$, regardless of the value of $v^{(l)}$. Hence, as we have at most $\lfloor \frac{n}{2} \rfloor$ components that satisfy this inequality, and j is allowed to take on $3\lfloor \frac{n}{2} \rfloor + 1$ values, there must be at least one choice of j in our range such that $\left| w_j^{(l)} \right| \geq .99$ for all $l = 2...n$. Let us call this choice j_*, and note that this means $\prod_{k=2}^{n} \left| w_{j_*}^{(k)} \right| \geq \left(.99^{\lfloor n/2 \rfloor} \right) > \left(.99^{\lfloor n/2 \rfloor} \right)^{\frac{n}{2}}$. Putting this all together, we have

$$\prod_{k=1}^{n} \left| w_{j_*}^{(k)} \right|^{\frac{2}{n}} = \left| w_{j_*}^{(1)} \right|^{\frac{2}{n}} \prod_{k=2}^{n} \left| w_{j_*}^{(k)} \right|^{\frac{2}{n}} > \left(\frac{(1.1)^{n/2}}{\left(.99^{\lfloor n/2 \rfloor} \right)^{\frac{n}{2}}} \left(.99^{\lfloor n/2 \rfloor} \right)^{\frac{n}{2}} \right)^{\frac{2}{n}} \geq 1.1.$$

This inequality together with (2) yields $\frac{\langle Aw_{j_*}, w_{j_*} \rangle}{n \prod_{k=1}^{n} \left| w_{j_*}^{(k)} \right|^{\frac{2}{n}}} < 1$, and we see that A is not maximally scaled by Theorem 5.2.2. By contradiction, no such λ_{min} can exist. □

Remark (1) There are many ways to improve the bound above, but the above bound was chosen for readability, as it is the simplest to demonstrate. Further, any stronger bounds that have been attained for generic elements of $MaxSc_n$ do not seem to give rise to any new meaningful results (that is, the return is not worth the investment).

(2) In Section 5.6, we will show that for elements from a few specific subsets of $MaxSc_n$, we can find a much more useful lower bound on their spectrum.

(3) By taking inverses, Proposition 5.3.3 also yields an upper bound on the spectrum of elements in $MinSc_n$ (see Proposition 5.4.1).

We have seen that Theorem 5.2.2 can be used to derive a few topological properties of $MaxSc_n$. We will now demonstrate how this result can be used to prove similar properties of $MinSc_n$.

5.4 Topological properties of $MinSc_n$

While it may not have been clear that Theorem 5.2.2 could help us investigate $MinSc_n$, we will now show that it does in fact allow us to derive a few topological properties of this set. We first formalize our remark at the end of the last section.

Proposition 5.4.1 *$MinSc_n$ is relatively compact.*

Proof Combine Proposition 5.1.2 with Proposition 5.3.3 to arrive at an upper bound on the spectral norm. The closure, $\overline{MinSc_n}$ must then also be bounded. Identifying the $n \times n$ matrices with $\mathbb{C}^{n^2}$, we apply the Heine-Borel theorem and achieve our desired result. $\square$

Relative compactness is okay, but it would be nice if we could conclude that $MinSc_n$ is in fact closed (making $MinSc_n$ compact). Unfortunately, a moment's thought convinces us that the closure of $MinSc_n$ must contain some positive semi-definite matrices, which are not considered to be in $MinSc_n$ by our definition. The natural question is: Suppose we extend this definition to include DQS positive semi-definite matrices that cannot be scaled to DQS matrices of smaller permanent; will this allow us to conclude closedness? We can come very close:

Proposition 5.4.2 *Suppose that M is a limit point of $MinSc_n$. Then M is a DQS matrix that cannot be scaled to another DQS matrix of smaller permanent.*

Proof Suppose $M \in \mathcal{PSD}_n$ is the limit of a sequence of minimally scaled matrices $\{M_m\}$. For the purposes of contradiction, let us assume that M is not minimally scaled, so there exists a diagonal matrix D, with $|per(D)| < 1$, such that $D^*MDe = e$.

Choose $\delta > 0$ small enough that $\left(\frac{1+\delta}{|1-\delta|^2}\right)^{2/n} < \frac{1}{|per(D)|}$, and fix m large enough that $\|M_m - M\| < \frac{\delta}{\sqrt{n}\|D\|^2}$.

Then

$$\begin{aligned}\|D^*M_mDe - e\| = \|D^*M_mDe - D^*MDe\| &= \|D^*(M_m - M)De\| \\ &\leq \|D^*\|\|M_m - M\|\|D\|\|e\| \\ &< \frac{\|D\|^2\delta\sqrt{n}}{\sqrt{n}\|D\|^2} \\ &= \delta.\end{aligned}$$

Let $B = D^*M_mD$ (so $\|Be - e\| < \delta$), and denote $v := Be$. We now examine B^{-1} (guaranteed to exist, as M_m is a positive definite matrix, and D must be invertible as it scales M).

Recall from the proof of Corollary 5.2.3 that if $k = \min_{y\in\Sigma} \frac{\langle B^{-1}y,y\rangle}{n}$, then the diagonal matrix with largest permanent that scales B^{-1} has permanent $\frac{1}{k^{n/2}}$. We note that $\frac{v}{\prod_{j=1}^{n} |v_j|^{1/n}} \in \Sigma$ and so

$$k \leq \frac{\langle B^{-1}v, v\rangle}{n\prod_{j=1}^{n} |v_j|^{2/n}} = \frac{\langle e, v\rangle}{n\prod_{j=1}^{n} |v_j|^{2/n}} = \frac{\sum_{k=1}^{n} v_k}{n\prod_{j=1}^{n} |v_j|^{2/n}}.$$

Now, as $\|v - e\| < \delta$, we have that $\sum_{k=1}^{n} v_k \leq n + n\delta$, and $\prod_{j=1}^{n} |v_j| \geq |1 - \delta|^n$.

Thus,

$$\frac{\langle B^{-1}v, v\rangle}{n\prod_{j=1}^{n}|v_j|^{2/n}} \leq \frac{1+\delta}{|1-\delta|^2}.$$

We see that $k = \min_{y\in\Sigma} \frac{\langle B^{-1}y,y\rangle}{n} \leq \frac{1+\delta}{|1-\delta|^2}$, and thus the diagonal matrix with largest permanent that scales B^{-1} has permanent with modulus equal to $\frac{1}{k^{n/2}} > |per(D)|$, by our assumption on δ.

Let E be this diagonal matrix of largest permanent that scales B^{-1}. Then $(E^{-1})^*$ is the diagonal matrix of smallest permanent that scales B, and hence $D(E^{-1})^*$ scales M_m. But as $|per(E)| = \frac{1}{k^{n/2}} > |per(D)|$, $per(E^{-1}) < 1/per(D)$, and thus $per(D(E^{-1}))^* < 1$, contradicting the fact that M_m is minimally scaled. □

Remark In the case where M is positive definite, the above result amounts to showing that $M \in MinSc_n$. It may seem that if we simply extended the definition of minimally scaled matrices to include positive semi-definite matrices that cannot be scaled to a DQS matrix of smaller permanent, the above result gives closure of this set. This may not be the case however, as there may exist "minimally scaled" positive semi-definite matrices that are not the limit of minimally scaled positive-definite matrices.

We now turn our attention to Conjecture 5.1.6 and bounds on the permanent of elements of $MaxSc_n$.

5.5 Permanental bounds for $MaxSc_n$

In this section, we will use Theorem 5.2.2 to verify Conjecture 5.1.6. That is, we will prove the following:

Theorem 5.5.1 *Let $M \in MaxSc_n$. Then $per(M) \geq 1$. Further, this lower bound is only achieved when $M = I_n$.*

We first require the following lemma, which follows easily from Corollary 5.3.2.

Lemma 5.5.2 *Let $M \in MaxSc_n$. Then $Re(M)$ is also maximally scaled.*

Proof As M is maximally scaled, then $\overline{M} = (\overline{m_{ij}})$ must also be a maximally scaled DQS matrix. (This follows from the observation that D scales M if and only if D^* scales $\overline{M}$, and the fact that $|det(D)| = |det(D^*)|$.) By Corollary 5.3.2, $MaxSc_n$ is convex, and hence $\frac{1}{2}(M + \overline{M}) = Re(M)$ is maximally scaled, as desired.

Our main tool in proving Theorem 5.5.1 will be the following propostion, which reveals a necessary condition for a matrix to be maximally scaled.

Proposition 5.5.3 *Let $M \in MaxSc_n$. Then for any column, the sum of the off-diagonals in that column is a non-positive real number. (i.e. $\sum_{\substack{j=1 \\ j \neq k}}^{n} m_{jk} \leq 0$ for all $1 \leq k \leq n$).*

Proof As M is doubly quasi-stochastic and the diagonal elements of M are real (as M is positive definite), it is immediate that $\sum_{\substack{j=1 \\ j \neq k}}^{n} m_{jk}$ must be real. Indeed, letting $A := Re(M)$,

it is easy to see that $\sum_{\substack{j=1 \\ j\neq k}}^{n} m_{jk} = \sum_{\substack{j=1 \\ j\neq k}}^{n} a_{jk}$ Thus, it suffices to examine the columns of A, a symmetric DQS real matrix, which is maximally scaled by Lemma 5.5.2.

We proceed by way of contradiction. Let us fix k and suppose that the sum of the off-diagonal entries in the k-th column of A is positive. Let $y \in \Sigma$ be the vector with k-th entry -1, and 1 everywhere else, and consider y^*Ay.

$$\begin{aligned} y^*Ay &= \sum_{\substack{j,l=1 \\ j,l\neq k}}^{n} a_{jl} + a_{kk} - 2\sum_{\substack{j=1 \\ j\neq k}}^{n} a_{jk} \\ &= e^*Ae - 4\sum_{\substack{j=1 \\ j\neq k}}^{n} a_{jk} \\ &= n - 4\sum_{\substack{j=1 \\ j\neq k}}^{n} a_{jk} \\ &< n, \end{aligned}$$

where the inequality follows from our assumption on column k, and the last equality arises from the fact that A is DQS, and hence $e^*Ae = n$. This contradicts the maximality of A by Theorem 5.2.2. Thus, no such column of A (and, by extension, no such column of M) exists. □

Remark (1) This can be seen as an extension of Proposition 5.1.7, which demonstrates that in the 2×2 case, the maximally scaled matrix must have (equal) non-positive real numbers in both off-diagonal entries.

(2) As M is Hermitian, the statement of Proposition 5.5.3 still holds if we replace all in-

stances of "column" with "row".

Our result now immediately follows:

Proof of Theorem 5.5.1 By (the remark following) Proposition 5.5.3, the off-diagonal entries of any given row of M must sum to a non-positive real number, but (as M is DQS) the sum of the whole row must be 1. Thus, every diagonal entry of M must be greater than or equal to 1. By the Hadamard permanent inequality (Theorem 1.2.11), $per(M) \geq \prod_{j=1}^{n} m_{jj} \geq 1$, with equality if and only if A is diagonal or has a zero row. Clearly, M cannot have a zero row as it is DQS and therefore equality holds if and only if $M = I_n$. □

This proof used the fact that if A is maximally scaled, it necessarily has diagonal elements greater than 1. One might wonder if this condition is also sufficient. This is not the case, however, as evidenced by the following counterexample.

Example 5.5.4 *Let*

$$A = \begin{pmatrix} 2 & 0.1 & 0.1 & -1.2 \\ 0.1 & 3 & -2.2 & 0.1 \\ 0.1 & -2.2 & 3 & 0.1 \\ -1.2 & 0.1 & 0.1 & 2 \end{pmatrix}.$$

Then A is a doubly quasi-stochastic matrix with diagonal elements greater than 1, but it is not maximally scaled, as for $v = \{1, -1, -1, 1\}^T \in \Sigma$, $\frac{\langle Av, v \rangle}{4} = 0.6 < 1$.

It would be nice if we could use the fact that all elements of $MinSc_n$ are inverses of elements of $MaxSc_n$ to conclude from Theorem 5.5.1 that Conjecture 5.1.4 was true as well.

Unfortunately, the permanent is not as well-behaved as the determinant, and this is not possible. We can, however, prove Conjecture 5.1.4 for a large class of matrices in $MinSc_n$.

5.6 Permanental bounds for $MinSc_n$

In this section, we will provide evidence for Conjecture 5.1.4 by proving that the statement of the conjecture is valid for all real matrices, group matrices that arise from Abelian groups, and matrices of dimension less than 4. We begin with real matrices, and introduce our main tool: A spectral characterization of real, maximally (and minimally) scaled DQS matrices. Recall that we denote the spectrum of A as $\sigma(A)$.

Theorem 5.6.1 *Let A be a real $n \times n$ positive definite DQS matrix. Then*

(1) A is maximally scaled if and only if $\sigma(A) \subseteq [1, \infty)$.

(2) A is minimally scaled if and only if $\sigma(A) \subseteq (0, 1]$.

Proof (1): As the $n = 2$ case was implicitly shown in Proposition 5.1.7, we will assume $n \geq 3$. Suppose that our result does not hold for some A. i.e. Suppose that A is a real, maximally scaled DQS matrix with (at least) one eigenvalue strictly less than one. Call this eigenvalue λ_1. Further, let us take $v^{(1)}, v^{(2)}, ...v^{(n)}$ to be an orthonormal basis of $\mathbb{R}^n$, consisting of (real) eigenvectors of A. Let us reorder this basis so that $v^{(1)}$ is an eigenvector associated with λ_1, and $v^{(2)} = \frac{e}{\sqrt{n}} = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, ..., \frac{1}{\sqrt{n}}\right)^T$, which must necessarily be an eigenvector as A is doubly quasi-stochastic (and hence $Ae = e$). We now investigate the properties of $v^{(1)}$.

As $v^{(1)}$ and $v^{(2)}$ are orthogonal to each other, we know that the sum of the components

of $v^{(1)}$, $v_j^{(1)}$, must be 0 (as $0 = \langle v^{(1)}, v^{(2)} \rangle = \frac{1}{\sqrt{n}} \left(v_1^{(1)} + v_2^{(1)} + ... + v_n^{(1)} \right)$). Further, as our eigenvectors are normalized and each component $v_j^{(1)}$ is real, we know that

$$1 = \|v^{(1)}\|^2 = \left(v_1^{(1)}\right)^2 + \left(v_2^{(1)}\right)^2 + ... + \left(v_n^{(1)}\right)^2.$$

Combining these two properties, we obtain the following:

$$\begin{aligned} 0 &= \left(v_1^{(1)} + v_2^{(1)} + ... + v_n^{(1)} \right)^2 \\ &= \left(v_1^{(1)}\right)^2 + \left(v_2^{(1)}\right)^2 + ... + \left(v_n^{(1)}\right)^2 + 2 \sum_{\substack{j,l=1 \\ j<l}}^{n} v_j^{(1)} v_l^{(1)} \\ &= 1 + 2 \sum_{\substack{j,l=1 \\ j<l}}^{n} v_j^{(1)} v_l^{(1)}, \end{aligned}$$

whence we see that

$$\sum_{\substack{j,l=1 \\ j<l}}^{n} v_j^{(1)} v_l^{(1)} = -\frac{1}{2} \tag{4}$$

.

Now, let $r > 1$ be a positive real number. We consider the vector $w = \frac{iv^{(1)} + r\sqrt{n}v^{(2)}}{\prod_{j=1}^{n} \left| iv_j^{(1)} + r\sqrt{n}v_j^{(2)} \right|^{\frac{1}{n}}}$ (we will be choosing r large enough that there is no chance of dividing by zero here). It is easy to verify that $\prod_{j=1}^{n} |w_j| = 1$ and hence $w \in \Sigma$. We will now show that for large enough r, $\frac{\langle Aw, w \rangle}{n} < 1$.

$$\begin{aligned}
\frac{\langle Aw, w\rangle}{n} &= \frac{\langle A\left(iv^{(1)}\right), iv^{(1)}\rangle + \langle A\left(r(1,...,1)^T\right), r(1,...,1)^T\rangle}{n\prod_{j=1}^{n}\left|iv_j^{(1)} + r\right|^{\frac{2}{n}}} \\
&= \frac{\lambda_1 + r^2 n}{n\prod_{j=1}^{n}\left|r + iv_j^{(1)}\right|^{\frac{2}{n}}} \\
&= \frac{\frac{\lambda_1}{n} + r^2}{\left|r^n + r^{n-1}i\left(\sum_{j=1}^{n} v_j^{(1)}\right) - r^{n-2}\left(\sum_{\substack{j,l=1\\ j<l}}^{n} v_j^{(1)}v_l^{(1)}\right) + l.o.t\right|^{\frac{2}{n}}},
\end{aligned}$$

where we are using "l.o.t" to denote terms where the order of r is strictly less than r^{n-2}. Recall that the sum of the components of $v^{(1)} = 0$, and hence the second term in the denominator vanishes. Then, factoring $|r^{n-2}|^{\frac{2}{n}}$ out of the denominator, we are left with

$$\begin{aligned}
\frac{\langle Aw, w\rangle}{n} &= \frac{\frac{\lambda}{n} + r^2}{\left|r^{\frac{2(n-2)}{n}}\right|\left|r^2 - \left(\sum_{\substack{j,l=1\\ j<l}}^{n} v_j^{(1)}v_l^{(1)}\right) + \frac{l.o.t}{r^{n-2}}\right|^{\frac{2}{n}}} \\
&= \frac{\frac{\lambda}{n}\left(r^{\frac{2(n-2)}{n}}\right)^{-1} + r^{\frac{4}{n}}}{\left|r^2 - \left(-\frac{1}{2}\right) + \frac{l.o.t}{r^{n-2}}\right|^{\frac{2}{n}}},
\end{aligned}$$

where the last equality comes from (4). Lastly, let us expand the quadratic expression in

the denominator, and consider $\left(\frac{\langle Aw,w\rangle}{n}\right)^n$:

$$\begin{aligned}\left(\frac{\langle Aw,w\rangle}{n}\right)^n &= \frac{r^4 + n\left(\frac{\lambda}{n}\right) r^{\left(\frac{8}{n} - \frac{2(n-2)}{n}\right)} + l.o.t_2}{|r^4 + r^2 + l.o.t_3|} \\ &= \frac{r^4 + \lambda r^2 + l.o.t_2}{|r^4 + r^2 + l.o.t_3|} \\ &= \frac{r^2 + \lambda + \frac{l.o.t_2}{r^2}}{\left|r^2 + 1 + \frac{l.o.t_3}{r^2}\right|},\end{aligned}$$

where $l.o.t_2$ and $l.o.t_3$ are comprised of summands with the order of r strictly less than 2.

For large enough r, the denominator will eventually dominate the numerator (in particular, this must happen when $1 - \lambda > \frac{|l.o.t_2| + |l.o.t_3|}{r^2}$). Thus, $\left(\frac{\langle Aw,w\rangle}{n}\right)^n < 1$ for large enough r, meaning $\frac{\langle Aw,w\rangle}{n} < 1$, and A is not maximally scaled by Theorem 5.2.2. By contraposition, if A is maximally scaled, then $\sigma(A) \subseteq [1, \infty)$.

For the sufficiency condition, let us suppose that the smallest eigenvalue of A is $\lambda_{min} = 1$. Now note that for any vector $v \in \Sigma$ (i.e. any vector such that $\prod_{j=1}^{n} |v_j| = 1$), $\|v\|^2 \geq n$ by the arithmetic and geometric means inequality applied to the real numbers $|v_j|^2$. Hence, for any such vector, $\frac{\langle Av,v\rangle}{n} \geq \frac{\lambda_{min}\|v\|^2}{n} \geq 1$. Thus by Theorem 5.2.2, A is maximally scaled.

(2): This follows from the above result and Proposition 5.1.2. □

Remark The above proof relies only on the fact that A has a complete set of real eigenvectors. The spectral conditions in Theorem 5.6.1 will hold for any doubly quasi-stochastic matrix with such a set.

Note that we can immediately conclude the following, which was strongly suspected by the authors of [2]:

Corollary 5.6.2 *Let A be a positive definite doubly-stochastic matrix. Then A is minimally scaled.*

Proof All positive definite doubly-stochastic matrices have their spectra contained in $(0, 1]$. By Theorem 5.6.1, they must be minimally scaled. □

We can also use Theorem 5.6.1 to prove Conjecture 5.1.4 for real elements of $MinSc_n$.

Corollary 5.6.3 *Let A be a real element of $MinSc_n$. Then $per(A) \leq 1$, with equality if and only if $A = I_n$.*

Proof By Theorem 5.6.1 (and the fact that A is DQS, and hence has $Ae = e$), the largest eigenvalue of A is $\lambda_1 = 1$. Now suppose A is not the identity matrix. Then A is not diagonal (as the identity matrix is the only diagonal DQS matrix) and A cannot have a zero row, as it would not be DQS. Thus we can apply Theorem 1.2.6. As f (as defined in Theorem 1.2.6) is strictly increasing on $(0, \infty)$, and bounded by $\lambda_1^n = 1^n = 1$, we see that $f(1) = per(A) < 1$, as desired. Of course, if $A = I_n$, then $per(A) = 1$. □

We can also prove Conjecture 5.1.4 for a different set of matrices, by virtue of their eigenvectors:

Definition Let $A = (a_{jk})$ be an $n \times n$ matrix. Then A is said to be a *group matrix* if there exists a (finite) group $\mathcal{G}$ with exactly n elements, (so $\mathcal{G} = \{g_j\}_{j=1}^n$), and a mapping $f : \mathcal{G} \to \mathbb{C}$

such that $a_{jk} = f(g_j^{-1} g_k)$ for all $j, k \in (1, ..., n)$. If $\mathcal{G}$ is Abelian, then we will say that the matrix A is an *Abelian group matrix*.

Remark If, in the above definition, $\mathcal{G}$ is the cyclic group of order n, then A is a circulant matrix. Proving Conjecture 5.1.4 for circulant matrices is notable because, as shown in Theorem 4.5.1, these matrices can have infinitely many scalings. This may make a direct search for the maximal or minimal scaling of a circulant matrix quite challenging.

Before proceeding, we require the following result on the eigenvectors of Abelian group matrices:

Lemma 5.6.4 *Let A be an abelian group matrix, and let $v \in \mathbb{C}^n$ be an eigenvector of A. Then the components of v, v_j, all have the same modulus (that is, $|v_1| = |v_2| = |v_3| = ...|v_n|$).*

Proof It is well-known (see for example [50]), that the eigenvectors $v^{(k)}$ of a group matrix are given by $v_j^{(k)} = \mathcal{X}(g_j)$, where $\mathcal{X}$ is a character of the group and g_j runs through the elements of G. Further, the characters of an Abelian group are always unitary (see Lemma 3.7 in [51]), meaning that $\mathcal{X}(G) \subset \mathbb{T}$, where $\mathbb{T}$ is the complex unit circle. Thus the components of $v^{(k)}$ (prior to normalization) are all elements of the unit circle, and our result follows. □

Now we can prove our main result on positive definite group matrices. Our method of proof will be very similar to our results on real matrices, as we begin by showing that they must satisfy the same spectral properties as in Theorem 5.6.1:

Theorem 5.6.5 *Let A be a positive definite, DQS, abelian group matrix. Then:*

(1) A is maximally scaled if and only if $\sigma(A) \subseteq [1, \infty)$.

(2) A is minimally scaled if and only if $\sigma(A) \subseteq (0, 1]$.

Proof (1): Suppose that A is maximally scaled with (at least) one eigenvalue strictly less than one. Let λ_1 be this eigenvalue, and let v be the normalized eigenvector associated with λ. Then by Lemma 5.6.4, $|v_1| = |v_2| = ... = |v_n| = \frac{1}{\sqrt{n}}$. Let $w = \sqrt{n}v$. Then $w \in \Sigma$, and $\frac{\langle Aw,w\rangle}{n} = \frac{\lambda_1 \|w\|^2}{n} = \lambda_1 < 1$, thus contradicting the maximality of A, by Theorem 5.5.1.

(2): This follows by Proposition 5.1.2 and the fact that the inverse of a non-singular Abelian group matrix is itself an Abelian group matrix (Proposition 6 in [52]). □

Remark It is easy to see that in the 2×2 case, any positive definite doubly quasi-stochastic matrix must necessarily be a real, circulant matrix. For this reason, our above theorem can again be seen as yet another extension of Proposition 5.1.7.

Corollary 5.6.6 *Let A be a positive definite, DQS, abelian group matrix. If A is an element of* $MinSc_n$*, then* $per(A) \leq 1$*, with equality if and only if* $A = I_n$.

Proof This follows exactly the same reasoning as the proof of Corollary 5.6.3.

Lastly, we will show that, in fact, Conjecture 5.1.4 holds when $n = 3$. Again, this will follow immediately from the stronger result:

Theorem 5.6.7 *Let A be a positive definite, DQS,* 3×3 *matrix. Then:*

(1) A is maximally scaled if and only if $\sigma(A) \subseteq [1, \infty)$.

(2) A is minimally scaled if and only if $\sigma(A) \subseteq (0, 1]$. □

Proof (1) Let us suppose that $v_1, e,$ and v_3 are a set of orthonormal eigenvectors of A, corresponding to $\lambda_1, 1, \lambda_3$. For the purposes of contradiction, let us assume $\lambda_1 < 1$. Firstly, note that if A has all real eigenvectors, then our result follows from the remark succeeding Theorem 5.6.1. Suppose then that one of (and hence both) v_1 or v_3 is not a complex multiple of a real vector. Then we can write $v_3 = (1, r_1 e^{i\theta_1}, r_2 e^{i\theta_2})^T$, where at least one of $r_1 e^{i\theta_1}$ or $r_2 e^{i\theta_2}$ is non-real. (Note that know that all of the components of v_3 are non-zero, as any vector that is orthogonal to e and has a zero component is a complex multiple of a real vector – either $(1, 0, -1)^T$, $(0, 1, -1)^T$, or $(1, -1, 0)^T$.)

Denote $u = (1, e^{i2\theta_1}, e^{i2\theta_2})^T$ and let $U = diag(u)$. Then $u \in \Sigma$, and we observe that $\langle u, v_3 \rangle = \langle e, U^* v_3 \rangle = \langle e, \overline{v_3} \rangle = \langle v_3, e \rangle = 0$, where the last equality follows from the fact that e and v_3 are orthonormal eigenvectors.

Then we can see that $u \in \{span\{v_3\}\}^\perp$, and hence $u \in span\{v_1, e\}$. Further, u is clearly not a multiple of e, by construction, so we can write $u = kv_1 + le$, for constants k and l, where $k \neq 0$. We claim that $\langle Au, u \rangle < 1$. Indeed, $\langle Au, u \rangle = \lambda_1 |k|^2 + |l|^2$. Lastly, as u is unimodular, we know that $\langle u, u \rangle = |k|^2 + |l|^2 = 1$. Thus, $\lambda_1 |k|^2 + |l|^2 < 1$, as $\lambda_1 < 1$. Thus, $\langle Au, u \rangle < 1$, and A is not maximally scaled.

(2) This follows by Proposition 5.1.2. □

Our result follows:

Corollary 5.6.8 *Let M be a minimally scaled 3×3 matrix. Then per(M) ≤ 1, with equality iff $M = I_3$, the 3×3 identity matrix.*

We have shown that Conjecture 5.1.4 is true when $n \leq 3$ or when the matrix in question

is either a real matrix or a group matrix arising from an abelian group. In doing so, we demonstrated that minimally scaled matrices that satisfy any of these conditions necessarily have their spectrum contained in $(0, 1]$. One might wonder if this condition holds for all $n \times n$ minimally scaled matrices. It turns out that this is not the case, however, as can be shown by an application of Theorem 5.2.2. Indeed, there are many $n \times n$ matrices that have an eigenvalue smaller than 1, yet (as observed from running an optimization program) satisfy the condition in Theorem 5.2.2 and are thus maximally scaled. One such example is given by:

$$M = \begin{pmatrix} 5 & -1+2.49i & -2-1.66i & -1-0.83i \\ -1-2.49i & 4 & 1+0.83i & -3+1.66i \\ -2+1.66i & 1-0.83i & 5 & -3-0.83i \\ -1+0.83i & -3-1.66i & -3+0.83i & 8 \end{pmatrix}.$$

As M is a maximally scaled matrix with one eigenvalue less than 1 (~ 0.946), we see that M^{-1} must be a minimally scaled matrix with largest eigenvalue greater than 1 (~ 1.057).

We now turn our attention to the application of our results on minimally scaled matrices, as we examine what they can tell us about the geometric measure of entanglement of Slater permanents.

5.7 Application: Bounds on the geometric measure of entanglement

The purpose of this section is simply to combine Theorem 1.3.3 with the results of the previous section to arrive at a few bounds on the GME of particular Slater permanents.

Theorem 5.7.1 *Let Let $\mathcal{H}$ be an n-dimensional complex Hilbert space and let $|\phi\rangle \in (\mathcal{H})^{\otimes n}$ be the Slater permanent with $n \times n$ invertible Gram matrix G_ϕ. If the minimally scaled matrix in $sc(G_\phi)$ is real or a group matrix arising from an abelian group, then*

$$E(|\phi\rangle) \leq 1 - \frac{n!}{(n^n)},$$

with equality if and only if G_ϕ is diagonal (i.e. all the components of $|\phi\rangle$ are orthogonal to each other).

Proof Combine Theorem 1.3.3 with Corollaries 5.6.3 and 5.6.6. □

This leads us to the easy corollary:

Corollary 5.7.2 *Let $|\phi\rangle \in (\mathbb{C}^n)^{\otimes n}$ be the Slater permanent with $n \times n$ invertible Gram matrix G_ϕ. If there exists a diagonal matrix D such that $D^* G_\phi D$ is a non-negative matrix, then*

$$E(|\phi\rangle) \leq 1 - \frac{n!}{(n^n)},$$

with equality if and only if G_ϕ is diagonal.

Proof Suppose $D^*G_\phi D$ is a real, non-negative positive definite matrix. Then by Theorem 2.2.2, there exists a positive matrix D_2 such that $B = D_2^* D^* G_\phi D D_2$ is a doubly-stochastic matrix. By Corollary 5.6.2, B is minimally scaled. Thus by Theorem 5.7.1,

$$E(|\phi\rangle) \leq 1 - \frac{n!}{(n^n)}.$$

□

Remark It is easy to see which matrices satisfy the above condition. These are simply the positive definite matrices that can be written as $R \circ V$, where R has all non-negative entries and $V = uu^*$ for some unimodular $u \in \mathbb{T}^n$.

Lastly, we have our bound for all Slater permanents when $n = 3$:

Theorem 5.7.3 *Let $|\phi\rangle \in (\mathbb{C}^3)^{\otimes 3}$ be a Slater permanent with invertible Gram matrix G_ϕ. Then*

$$E(|\phi\rangle) \leq 1 - \frac{6}{27},$$

with equality if and only if G_ϕ is diagonal.

Proof Combine Theorem 1.3.3 and Theorem 5.6.8.

This shows that in the 3×3 case (and in the 2×2 case, as shown in [2]), we know

that the maximally entangled state is the state with orthogonal components. The following conjecture is equivalent to Conjecture 5.1.4.

Conjecture 5.7.4 *The (unique) maximally entangled Slater permanent in* $(\mathbb{C}^n)^{\otimes n}$ *is* $\frac{1}{n!}\sum_{\sigma\in S_n}\bigotimes_{k=1}^{n}|\phi_{\sigma(k)}\rangle$, *where* $\langle\phi_k|\phi_j\rangle = 0$ *for all* $j \neq k$.

This concludes our discussion of scalings with extremal permanent. A full proof of Conjecture 5.1.4 (and 5.7.4) still eludes us, but in our next chapter we will turn our attention to some other open problems on scalings worth investigating.

Chapter 6

Open problems

6.1 Introduction

We now turn our attention to a few ideas that seem worthy of future consideration. We will begin with a discussion of an algorithm that may be useful in constructing scalings of a given positive definite $n \times n$ matrix. While we have strong experimental evidence that this algorithm always converges, we have yet to prove this. After a discussion of this algorithm, we will introduce a few open problems found in the literature, and discuss how matrix scalings may yield useful new approaches for attacking these problems.

6.2 An algorithm of interest

Given a positive definite matrix A, we wish to create an algorithm that will construct many different scalings of A. Indeed, in the cases where $sc(A)$ is finite, we would like this algorithm

to construct all elements of $sc(A)$. To this end, we will construct an algorithm that seems to do just this, though with a few caveats. We begin by introducing a result that gives rise to an easily implementable program for the real case. In [5], the author provided a constructive proof of the positive definite case of Marshall and Olkin's result on symmetric copositive matrices (Theorem 2.2.2) that does not require minimizing over Σ. Her method of proof involved identifying the scaling as the solution to a certain differential equation. This allows us to employ a differential equation solver to arrive at our scaling. In this chapter, we will extend O'Leary's idea into an algorithm that will construct numerous complex scalings of a given positive definite matrix. The work in this chapter is still in progress, and it is included here merely to provide the reader with a useful tool for calculating multiple scalings.

6.3 O'Leary's method for scaling real, positive definite matrices

We begin by discussing the algorithm of O' Leary. We outline the method, and briefly discuss the reasoning for why it works. We will not give every detail of the proof, but these can be found in [5]. Let A be a real, positive definite $n \times n$ matrix and consider the following homotopy of equations:

$$H(t,d) = D(t)(tA + (1-t)I_n)D(t)e - e, \qquad t \in [0,1], \tag{1}$$

where $e = (1, 1...1)^T$, d is a (positive) vector valued function of t and $D(t)$ is the diagonal

matrix with entries $(d(t))_j$. Setting $H = 0$, our desire is to find $d(t)$, given t.

Suppose $t = 0$. Then we have $D(0)I_nD(0)e - e = 0$, and it follows that $d(0) = e$. If we can find $d(1)$ (i.e. $t = 1$), then we have the scaling that we seek (that is, $D(1) = diag(d(1))$ is the positive diagonal matrix that scales A to a doubly quasi-stochastic matrix). Following the argument in [5], we differentiate (1) to yield:

$$\partial_d H(t,d)d'(t) + \partial_t H(t,d) = 0, \quad d(0) = e, \tag{2}$$

where it is easy to calculate the above partial derivatives:

$$\partial_d H(t,d) = D(t)(tA + (1-t)I) + diag((tA + (1-t)I)d(t)),$$

$$\partial_t H(t,d) = D(t)(A - I)d(t).$$

Rearranging (2) allows us to solve for $d'(t)$. Indeed, at this point the author of [5] employs the Picard-Lindelhoff Theorem and a few analytic bounds on d(t) to show that $f(t,d) = -(\partial_d H(t,d)^{-1})\partial_t H(t,d)$ does indeed have a solution when $t = 1$, yielding $d(1)$ and, by extension, our positive matrix $D(1)$. We will not tarry on these details, but note that by running (2) through any decent differential equation solver, we arrive at an easy to implement scaling algorithm for any real positive definite matrix.

We now introduce an algorithm (Algorithm O) that will calculate a (D^*AD) scaling of a given *complex* positive definite matrix with high probability (but not with certainty). We will later modify this algorithm to arrive at a program that seems to calculate scalings with

100% efficacy.

6.4 Algorithm O

Our first algorithm is easy to state and very easy to implement.

Algorithm O: Input: C, a positive definite matrix, tolerance level *tol*

(0) $C_0 := C$, $i = 1$

(1) $G = DC_{i-1}D$, where D is the diagonal matrix that scales the real part of C (using O' Leary's method, for example).

(2) Define $\{w_j\}$ to be an orthonormal set of eigenvectors for G, and let $W_j = diag(w_j)$. Let $W_j = P_jU_j$ be the polar decomposition of W_j.

(3) $C_i = U^*GU$, where U is the matrix that realizes the following minimization problem: $\min_{j=1\ldots n} \|U_j - I_n\|$.

(4) If $\|C_ie - e\| < tol$, define $B = C_i$ and **stop**. Otherwise continue to step (5).

(5) $i = i + 1$, Repeat steps (1)-(4).

Output: The DQS matrix $B = E^*CE$, for some diagonal E.

end

Remark Step (3) amounts to finding the eigenvector that is closest to e, then using the ray pattern of this vector to scale the matrix. The resultant matrix of this step will have the same eigenvalues, and the moduli of the components of the eigenvectors also remain unchanged. However, it now has an eigenvector with all non-negative components.

This algorithm is not perfect and can stand to be improved. There are still rare cases where it gets caught in a 2-loop between step (1) and step (3). As well, in the case where the real part of G has repeated eigenvalue 1, we may converge to a non-DQS matrix (see the remark after Proposition 6.5.1) and our condition in (4) is never satisfied. That said, it does calculate a scaling for "most" matrices (around 98% of the time, according to our tests), as we will discuss in the next chapter. Further, by adding one additional step (see Section 6.6), it seems that we can circumvent these issues and obtain an algorithm that can calculate many different scalings for *any* given matrix. Before introducing this "modified algorithm", we discuss why Algorithm O seems to converge in most cases.

6.5 Discussion of Algorithm O

Algorithm O consists of two repeated steps: a positive scaling (step (1)) and a unitary scaling (step 3). In other words, our first step scales the real part of C_{i-1}, and our second step ensures that the resultant matrix has a non-negative eigenvector. Note that the set of matrices M that satisfy both of these properties is exactly the set of doubly quasi-stochastic matrices, provided $\lambda = 1$ (which is guaranteed to be an eigenvalue of $Re(M)$, by step (1)) is a *simple* eigenvalue of $Re(M)$.

Proposition 6.5.1 *Let M be a positive definite matrix with eigenvector v, where v has all non-negative components. Further, suppose that $A := Re(M)$ is doubly quasi-stochastic with $\lambda = 1$ a simple eigenvalue of A. Then M is doubly quasi-stochastic.*

Proof As v is a real eigenvector of M, it must also be an eigenvector of A. As A is DQS, A has $e = (1, 1, ..., 1)^T$ as an eigenvector with associated eigenvalue $\lambda = 1$. This is a simple eigenvalue, meaning that v must either be orthogonal to e or $v = e$ (up to multiplication by a constant). But v is non-negative and thus cannot be orthogonal to e. Thus $v = e$ and M is DQS (as $Me = Ae = e$). □

Remark If A has a repeated eigenvalue of 1, then A will have two distinct, non-orthogonal eigenvectors v and e, both non-negative. This explains why − in rare cases − Algorithm O converges to a non-DQS matrix: A matrix M will be a non-DQS fixed point of Algorithm O if M has a non-negative eigenvector $v \neq e$ and DQS real part A. (Again, these can only occur simultaneously if A has repeated eigenvalue $\lambda = 1$.)

Alternating step (1) and step (3) with the goal of arriving at a matrix in the intersection of these two sets may suggest some sort of "alternating projections" method, as introduced by von Neumann in [53], but this is not immediately applicable for numerous reasons. First of all, the classical alternating projections method requires convexity, and while the set of positive definite matrices with scaled real part is convex, the set of positive definite matrices with with (at least) one non-negative eigenvector is not. Further, these steps are not *projecting* onto their respective sets in the usual sense. One could try to apply a similar method that uses a more generalized notion of projections, such as the Alternating Bregman Projection method used in [54] (indeed, as outlined in [55], when $Re(C)$ is non-negative, we can view step (1) as one of these "Bregman projections") but step (3) does not seem to fit into any such framework at this time.

Generalizing the alternating projections method is not the only method that offers hope for proving convergence. In general terms, if step (3) does not perturb $Re(G)$ too much (so that it is still close to DQS), and if step (1) does not perturb the eigenvectors too much, then our algorithm should converge. To this end, we introduce a pair of results of Eisenstat and Ipsen [56]. This result came about as the authors sought to extend the famous results of Davis and Kahan [57] on the eigenvectors of a perturbed matrix. We do not state the results in their original form, but instead re-word them to suit our setting.

Proposition 6.5.2 ([56], Theorem 4.1) *Suppose $B = DAD$, where A, B are positive definite with simple eigenvalues, and D is a positive diagonal matrix. If $\lambda_B \in \sigma(B)$, and $\lambda_A \in \sigma(A)$ is the closest eigenvalue to λ_B, then*

$$|\lambda_B - \lambda_A| \leq \lambda_A \|I - D^{-2}\|.$$

Proposition 6.5.3 ([56], Theorem 4.3) *Suppose $B = DAD$, where A, B are positive definite with simple eigenvalues, and D is a positive diagonal matrix. Suppose $\lambda_B \in \sigma(B)$, and $\lambda_A \in \sigma(A)$ is the closest eigenvalue to λ_B. Let v_A, v_B be eigenvectors associated with λ_A, λ_B, respectively. Lastly, define $\theta \in [0, \pi/2]$ to be the angle between v_A and v_B. Then*

$$\sin\theta \leq \frac{\min\{\alpha_1, \alpha_2\}|\lambda_A|}{\min\limits_{\lambda \in \sigma(B)\setminus\{\lambda_B\}} |\lambda - \lambda_A|} + \|I - D\|,$$

where $\alpha_1 = \|D^{-1} - D\|$, and $\alpha_2 = \|I - D^{-2}\|$.

This result suggests that we can bound U from step (3), given D from step (1). Suppose

that D is close to I. Then α_1 and α_2 are small, meaning that $\sin\theta$ is close to 0 (assuming $|\lambda_A|$ is not too large, and the minimization in the denominator is not too small), and hence U is also close to I.

Our numerical results suggest that in "almost all" cases (our tests have the efficacy rate at about 98%) the other direction holds as well. That is, if $\|U - I\|$ is small for U in step (3), then D in step (1) of the next iteration will be close to I. An argument that yields conditions on when this necessarily occurs still eludes us, but it seems that Proposition 6.5.2 will likely be helpful, as step (1) essentially scales one of the eigenvalues to 1, and D is *usually* small if one of the eigenvalues of C_i is already close to 1 (indeed, it seems to be precisely when D is consistently large that our algorithm enters a 2-loop).

These ideas may be the key to proving when Algorithm O converges. The idea would be that at each subsequent iteration, D is closer to I, causing U to get closer to I, forcing the next D to be *even closer* to I, and so forth. We now introduce a modified algorithm that has successfully found (at least) one scaling in every test conducted.

6.6 Modifications to Algorithm O

As mentioned, Algorithm O usually works, but there are cases where it fails. We have already discussed the case where the matrix PU is singular, in which case we converge to a non-DQS matrix. The other failures seem to always take the form of 2-cycles, where subsequent iterations of the program alternate between two solutions, neither of which are doubly quasi-stochastic. It is not obvious what conditions on C cause this failure, but it is

easy to modify the algorithm to one that seems to always converge. Note that steps (1)-(4) from Algorithm 1 are exactly the same as steps (1)-(4) in Algorithm O.

Algorithm 1:

Input: Positive definite C, tolerance level tol, maximum iterations k_{max}

(0) $C_0 := C$, $k := 1$

(1) $G = DC_{k-1}D$, where D is the diagonal matrix that scales the real part of C, (using O' Leary's method, for example).

(2) Define $\{w_j\}$ to be an orthonormal set of eigenvectors for G, and let $W_j = diag(w_j)$. Let $W_j = P_jU_j$ be the polar decomposition of W_j.

(3) $C_k = U^*GU$, where U is the matrix that realizes the following minimization problem: $\min_{j=1...n} \|U_j - I_n\|$.

(4) If$\|C_ke - e\| < tol$, define $B = C_i$ and **stop**. Otherwise, continue to step (5).

(5) If $k < k_{max}$, set $k = k + 1$ and return to step (1). Otherwise continue to step (6).

(6) Construct a random complex diagonal vector D and define $C_1 = D^*C_kD$. Set $k = 1$ and return to step (1).

Output: The DQS matrix $B = E^*CE$, for some diagonal E.

end

Remark (1) This amounts to multiplying our matrix by a random diagonal matrix if we have been caught in an apparent loop (or a non-DQS fixed point). While this is not a particularly elegant solution, it has thus far shown to have 100% efficacy in trials. Whatever conditions cause our algorithm to loop or converge to a non-DQS fixed point,

they do not seem to be preserved by diagonal scaling. For these reasons, we usually only have to invoke step (6) once before convergence is achieved.

(2) A sensible choice of k_{max} seems to be $100n^2$, where n is the dimension of C.

Algorithm 1 gives us one scaling. This is nice, but one might wonder why it is better than using Lemma 5.2.1 to arrive at a different scaling algorithm, one that just employs an optimization program to minimize y^*Cy over Σ. The answer lies with one slight modification to Algorithm 1. The result is a program that calculates many difference scalings of the same matrix C. Note the change in Step (4):

Algorithm 2:

Input: Positive definite C, tolerance level tol, maximum iterations k_{max}

(0) $C_0 := C$, $k := 1$

(1) $G = DC_{k-1}D$, where D is the diagonal matrix that scales the real part of C, (using O' Leary's method, for example).

(2) Define $\{w_j\}$ to be an orthonormal set of eigenvectors for G, and let $W_j = diag(w_j)$. Let $W_j = P_jU_j$ be the polar decomposition of W_j.

(3) $C_k = U^*GU$, where U is the matrix that realizes the following minimization problem: $\min_{j=1...n} \|U_j - I_n\|$.

(4) If $\|C_ke - e\| < tol$, define $B = C_k$, print B and skip to step (6). Otherwise, continue to step (5).

(5) If $k < k_{max}$, set $k = k + 1$ return to step (1). Otherwise continue to step (6).

(6) Construct a random complex diagonal matrix D and define $C_1 = D^*C_kD$. Set $k = 1$

and return to step (1).

Output: A set of DQS matrices $\{B = E^*CE\}$ for some diagonal matrices E.

end

Remark Once this algorithm finds a scaling, it records the scaling (by printing it to screen) and then multiplies that scaling by a random diagonal matrix. It then begins a search for scalings of the resultant matrix. Of course, this often finds the same scaling multiple times, but in *all* test cases (excluding the case where C is diagonal and there is only one scaling), it has also resulted in the identification of many different scalings for the same matrix (with more scalings being found for higher-dimensional matrices). The minimization approach suggested above will always only calculate one scaling even if we introduce a step similar to step (6), as this minimization is unaffected by diagonal scaling.

Algorithm 2 has been tested on thousands of randomly constructed positive matrices C and in all cases (except where C is diagonal) it has found multiple scalings. This suggests that all non-diagonal matrices may necessarily have more than one scaling. As mentioned in [2], this is true in the 2×2 case, where all non-diagonal matrices have exactly 2 scalings. Further, in all tests where C has no zero entries, Algorithm 2 has found at least 2^{n-1} scalings. This motivates the following conjecture:

Conjecture 6.6.1 *Let C be an $n \times n$ positive definite matrix with no zero entries. Then $|sc(C)| \geq 2^{n-1}$.*

Remark As mentioned, this has been shown to be true when C is a 2×2 matrix. In the case

where C is real, this can be seen to be true by the following application of Theorem 2.2.2: If C is real, then it is scaled by a positive diagonal matrix (by Theorem 2.2.2). Further, for all 2^{n-1} non-equivalent diagonal matrices $D = (\pm 1, \pm 1, ..., \pm 1)$, D^*MD is a real, positive definite matrix with a different sign pattern than C. D^*MD can be scaled by a positive diagonal matrix, and this gives rise to 2^{n-1} real scalings which are necessarily different by virtue of their sign pattern.

This concludes our discussion of scaling algorithms. The remainder of this chapter will show how diagonal matrix scalings can be used to approach a variety of open problems from the literature.

6.7 The permanent conjectures of Chollet and Drury

As mentioned in Chapter 1, there are many open problems concerning the matrix permanent. In this section, we will look at two problems that specifically consider positive semi-definite matrices. We begin with the conjecture of Chollet, originally made in [58].

Conjecture 6.7.1 (The Chollet Conjecture) *Let A, B be $n \times n$ positive semi-definite matrices. Then*

$$per(A \circ B) \leq per(A)per(B),$$

where $A \circ B$ denotes the Hadamard product of A and B.

Remark While Gregorac and Hentzel [59] proved that this is true when $n \leq 3$, there has been very little progress for higher dimensions.

In an effort to prove Chollet's Conjecture, Stephen Drury made the following conjecture, as communicated by Fuzhen Zhang in [8]:

Conjecture 6.7.2 ([8], Drury permanent conjecture) *Let $A = (a_{ij})$ be an $n \times n$ positive semidefinite matrix and let B be the submatrix obtained by removing the first row and column of A. Further, let B_{kk} be the submatrix of B obtained by removing the k-th row and column of B $(k = 1, 2..., n-1)$. Then*

$$(a_{11}per(B))^2 + \left(\sum_{k=2}^{n} |a_{1k}|^2 per(B_{k-1k-1})\right)^2 \leq (per(A))^2.$$

If the Drury permanent conjecture could be proven true, then Chollet's conjecture would necessarily follow.

6.7.1 Matrix scalings and a counterexample to the Drury permanent conjecture

As we investigate these two conjectures, the utility of matrix scalings arises by virtue of the fact that the inequalities in both conjectures are invariant with respect to complex matrix scalings.

Proposition 6.7.3 *Let A, P be $n \times n$ positive semi-definite matrices, D, E be $n \times n$ invertible diagonal matrices with entries, and $C := D^*AD$, $G = E^*PE$. Then*

(1) $per(A \circ P) \leq per(A)per(P)$ if and only if $per(C \circ G) \leq per(C)per(G)$. (i.e. The Chollet Conjecture holds for (A, P) if and only if it holds for (C, G).)

(2) The Drury permanent conjecture holds for A if and only if it holds for C.

Proof (1): The left side of the second inequality becomes:

$$\begin{aligned} per(C \circ G) &= per(D^*AD \circ E^*PE) \\ &= per(D^*E^*(A \circ P)ED) \\ &= |per(D)|^2|per(E)|^2 per(A \circ P), \end{aligned}$$

while the right side becomes:

$$\begin{aligned} per(C)per(G) &= per(D^*AD)per(E^*PE) \\ &= |per(D)|^2 per(A)|per(E)|^2 per(P). \end{aligned}$$

The inequality clearly holds if and only if $per(A \circ P) \leq per(A)per(P)$.

(2): Define complex numbers $d_1, ..., d_n$ so that $D = diag(d_1, ..., d_n)$. We note that $(D^*AD)_{ij} = \overline{d_i} d_j a_{ij}$. We define $F = diag(d_2, ..., d_n)$, and as above, we will let F_{kk} be the submatrix of F obtained by removing the k-th row and column of F. The left hand side (LHS) of the inequality for matrix C becomes:

$$
\begin{aligned}
LHS &= \left((|d_1|^2 a_{11} per(F^*BF)\right)^2 + \left(\sum_{k=2}^{n} |\overline{d_1} d_k a_{1k}|^2 (per(F^*BF))_{k-1k-1}\right)^2 \\
&= |d_1|^4 a_{11}^2 \left(\prod_{i=2}^{n} |d_i|^4\right) (perB)^2 + \left(\sum_{k=2}^{n} |d_1|^2 |d_k|^2 |a_{1k}|^2 \prod_{j=2, j\neq k}^{n} |d_j|^2 per(B_{k-1k-1})\right)^2 \\
&= \left(\prod_{i=1}^{n} |d_i|^4\right) a_{11}^2 (perB)^2 + \left(\left(\prod_{j=1}^{n} |d_j|^2\right) \sum_{k=2}^{n} |a_{1k}|^2 per(B_{k-1k-1})\right)^2 \\
&= \left(\prod_{i=1}^{n} |d_i|^4\right) \left((a_{11} per(B))^2 + \left(\sum_{k=2}^{n} |a_{1k}|^2 per(B_{k-1k-1})\right)^2\right),
\end{aligned}
$$

and the right hand side bcomes

$$
\begin{aligned}
per(C)^2 &= (per(D^*AD))^2 \\
&= |per(D)|^4 (per(A))^2 \\
&= \prod_{i=1}^{n} |d_i|^4 (per(A))^2,
\end{aligned}
$$

whence we see that the inequality holds if and only if

$$
(a_{11} per(B))^2 + \left(\sum_{k=2}^{n} |a_{1k}|^2 per(B_{k-1k-1})\right)^2 \leq (per(A))^2.
$$

□

Proposition 6.7.3 simplifies our problem significantly, as it allows us to restrict our focus to particular classes of matrices.

Corollary 6.7.4 *The following are equivalent:*

(1) The Chollet conjecture is true for all $n \times n$ positive definite matrices.

(2) The statement of the Chollet conjecture holds for all $n \times n$ positive definite doubly quasi-stochastic matrices.

(3) The statement of the Chollet conjecture holds for all $n \times n$ matrices M that satisfy:

$$\min_{y \in \Sigma} \frac{y^* M y}{n} = 1.$$

(4) The statement of the Chollet conjecture holds for all $n \times n$ positive definite minimally scaled matrices.

(5) The statement of the Chollet conjecture holds for all $n \times n$ positive definite correlation matrices.

Proof By Proposition 6.7.3, the conjecture is true for positive definite A, B if and only if we can find diagonal D, E such that the conjecture holds for D^*AD and E^*BE. The proof is completed by the following observations:

(2) By Theorem 1.3.4, we can always scale a positive definite matrix to a positive definite DQS matrix.

(3) This is simply our characterization of maximally scaled matrices from Theorem 5.2.2. By Corollary 5.2.3, every positive definite matrix can be scaled to a maximally scaled matrix.

(4) Every positive definite matrix can be scaled to a minimally scaled matrix, by Corollary 5.2.3.

(5) Define $D = diag(\frac{1}{\sqrt{a_{11}}}, ..., \frac{1}{\sqrt{a_{nn}}})$, and $E = diag(\frac{1}{\sqrt{b_{11}}}, ..., \frac{1}{\sqrt{b_{nn}}})$. Then D^*AD and E^*BE are both correlation matrices. □

Similarily,

Corollary 6.7.5 *The following are equivalent:*

(1) The Drury permanent conjecture is true for all $n \times n$ positive definite matrices.

(2) The statement of the Drury permanent conjecture holds for all $n \times n$ positive definite doubly quasi-stochastic matrices.

(3) The statement of the Drury permanent conjecture holds for all $n \times n$ matrices that satisfy:

$$k := \min_{y \in \Sigma} \frac{y^* M y}{n} = 1.$$

(4) The statement of the Drury permanent conjecture holds for all $n \times n$ positive definite minimally scaled matrices.

(5) The statement of the Drury permanent conjecture holds for all $n \times n$ positive definite correlation matrices.

Corollary 6.7.5 indicates that we can restrict our thinking to correlation matrices, which simplifies things significantly and has led us to the following counterexample to the Drury permanent conjecture (originally appearing in [60]). In what follows, the matrix B is defined as in the statement of the conjecture.

Example *Let*

$$A = \begin{pmatrix} 1 & -0.0110 + 0.2225i & 0.2281 - 0.0237i & -0.1790 - 0.1755i \\ -0.0110 - 0.2225i & 1 & 0.9140 + 0.1634i & 0.8974 + 0.0541i \\ 0.2281 + 0.0237i & 0.9140 - 0.1634i & 1 & 0.8758 - 0.2433i \\ -0.1790 + 0.1755i & 0.8974 - 0.0541i & 0.8758 + 0.2433i & 1 \end{pmatrix}.$$

Then

$$(a_{11}per(B))^2 + \left(\sum_{k=2}^{n} |a_{1k}|^2 per(B_{k-1k-1})\right)^2 \approx 25.0522 > 25.0336 \approx (per(A))^2,$$

contradicting the Drury permanent conjecture.

It is still unknown whether the Drury permanent conjecture holds for all *real* positive semi-definite matrices. This is an open problem worth pursuing, as it would imply the real case of the Chollet conjecture.

6.8 Idel and Wolf - Unitary scalings

Up until this point, we have been considering problems involving D^*AD-scalings of positive definite matrices. We will close this chapter by considering a different type of matrix scaling. While we will focus on how these "unitary scalings" can be applied to a pair of open problems

from the literature, the very existence of these scalings is notable as it illustrates a surprising decomposition of unitary matrices that every matrix theorist ought to know.

We begin with some motivation. Let U be an $n \times n$ unitary matrix, and $v \in \mathbb{T}^n$. If it is also the case that $Uv \in \mathbb{T}^n$, then we call v a *biunimodular vector for* U. The search for biunimodular vectors for the $n \times n$ Fourier matrix has been the topic of study for many years (see [61], [62] and [63], for example) and in [64], the author extended this search to all unitary matrices.

The connection to matrix scaling is easy to see. As shown in [65], if v is a unimodular vector for U, and $Uv = w \in \mathbb{T}^n$, then $W^*UVe = e$, where $W = diag(w)$, and $V = diag(v)$. This shows that all row sums of W^*UV are 1. Further, one can multiply both sides by the adjoint of W^*UV to see that we do in fact have column sums equal to 1 as well. Further, if there exists such matrices V, W, then one can confirm that Ve is a biunimodular vector for U. Thus the search for biunimodular vectors of a unitary matrix U is equivalent to finding two diagonal unitary matrices V, W such that W^*UV is doubly quasi-stochastic.

Using the hunt for biunimodular vectors as motivation, Idel and Wolf [65] proved that such a scaling can always be found.

Theorem 6.8.1 ([65], Theorem 2) *Let U be an $n \times n$ unitary matrix. Then there exist two diagonal unitary matrices V, W, such that W^*UV is doubly quasi-stochastic.*

Remark Theorem 6.8.1 gives the following decomposition of any unitary matrix: Given a unitary matrix U, we can write $U = LSR$, where L, R are diagonal unitary matrices, and S is a doubly quasi-stochastic unitary matrix. Adapting the terminology in [65], we will call

this the *Sinkhorn-type Normal Form (SNF)* of U.

We will now see how these scalings may prove useful in solving a few open problems from the literature.

6.9 The $per(U^*D_\sigma U)$ problem

The following question is a famous open problem in matrix theory that was first introduced in [66]:

Question 6.9.1 *Suppose $\sigma = \{\lambda_1, ..., \lambda_n\}$ is a given set of positive real numbers, and let $\mathcal{A}_\sigma$ denote the set of postive definite matrices with spectrum σ. Which element(s) of $\mathcal{A}_\sigma$ have the largest permanent?*

It is easy to see that this question is equivalent to finding $\max_{U \in \mathcal{U}(n)} per(U^*D_\sigma U)$, where $\mathcal{U}(n)$ denotes the $n \times n$ unitary matrices and $D_\sigma = diag(\sigma)$. We will now use Theorem 6.8.1 to approach this problem in a new way (the following argument does not seem to have appeared anywhere in the literature at this time).

Let $U \in \mathcal{U}(n)$ with SNF $U = LSR$. Then

$$\begin{aligned} per(U^*D_\sigma U) &= per((LSR)^*D_\sigma(LSR) \\ &= per(R^*S^*L^*D_\sigma LSR). \end{aligned}$$

As R is a diagonal unitary matrix, it is immediate that

$$per(R^*S^*L^*D_\sigma LSR) = |per(R)|^2 per(S^*L^*D_\sigma LS) = per(S^*L^*D_\sigma LS).$$

Using this fact and observing that $L^*D_\sigma L = L^*LD_\sigma = D_\sigma$ (as diagonal matrices commute), we obtain

Proposition 6.9.2 *Let $\sigma = \{\lambda_1, ..., \lambda_n\}$ be a given set of positive real numbers, and define $D_\sigma = diag(\sigma)$. Then*

$$\max_{U \in \mathcal{U}(n)} per(U^*D_\sigma U) = \max_{S \in \mathcal{DQU}(n)} per(S^*D_\sigma S),$$

where $\mathcal{DQU}(n)$ consists of all $n \times n$ DQS unitary matrices.

Thus we have simplified our question a little bit. Rather than considering all unitary matrices, we need only look to the ones that are also doubly quasi-stochastic. One result follows easily, as we have a simple solution in the 2×2 case. (This is a new proof to an old result. Grone et. al solved the 2×2 case in [67], using a relatively complicated argument that made use of an entity known as the "permanental adjoint".)

6.9.1 The $n = 2$ case

We require the following well-known lemma (see, for example [68]).

Lemma 6.9.3 *Let A be an $n \times n$ matrix, with all row and column sums equal to k, and*

denote the $n \times n$ Fourier matrix by F_n. Then

$$F_n^* A F_n = \begin{pmatrix} k & 0 \\ 0 & B \end{pmatrix},$$

for some $(n-1) \times (n-1)$ matrix B.

Note that in the case where A is a unitary matrix, then $F_n^* A F_n$ is the product of unitary matrices and must itself be unitary.

Theorem 6.9.4 *Let A be a 2×2 positive definite matrix with spectrum $\sigma = \{\lambda_1, \lambda_2\}$. Then $per(A) \leq \frac{\lambda_1^2 + \lambda_2^2}{2}$, with equality if and only if*

$$\frac{1}{2} \begin{pmatrix} \lambda_1 + \lambda_2 & \pm(\lambda_1 - \lambda_2)i \\ \mp(\lambda_1 - \lambda_2)i & \lambda_1 + \lambda_2 \end{pmatrix}.$$

Proof By Proposition 6.9.2, we need only consider $\max_{S \in \mathcal{DQU}(n)} S^* D_\sigma S$, where $D_\sigma = diag(\lambda_1, \lambda_2)$. Let $\mathcal{G}$ denote the set of diagonal unitary matrices of dimension 2 with (1,1) entry equal to 1. Then applying Lemma 6.9.3, we obtain

$$\max_{S \in \mathcal{DQU}(n)} per(S^* D_\sigma S) = \max_{V \in \mathcal{G}} per(F_2 V^* F_2^* D_\sigma F_2 V F_2^*).$$

Recalling that $F_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and denoting $U = diag(1, e^{i\theta})$, this matrix product evaluates to

$$\max_{S\in\mathcal{DQU}(n)} per(S^*D_\sigma S) = \max_{\theta\in[0,2\pi]} \frac{1}{2}\begin{pmatrix} (\lambda_1+\lambda_2)+(\lambda_1-\lambda_2)Re(e^{i\theta}) & -(\lambda_1-\lambda_2)Im(e^{i\theta})i \\ (\lambda_1-\lambda_2)Im(e^{i\theta})i & (\lambda_1+\lambda_2)-(\lambda_1-\lambda_2)Re(e^{i\theta}) \end{pmatrix},$$

whence we obtain

$$\max_{S\in\mathcal{DQU}(n)} per(S^*D_\sigma S) = \max_{\theta\in[0,2\pi]} \frac{1}{4}\left((\lambda_1+\lambda_2)^2-(Re(e^{i\theta}))^2(\lambda_1-\lambda_2)^2+(Im(e^{i\theta}))^2(\lambda_1-\lambda_2)^2\right).$$

It is clear now that our permanent increases as $|Re(e^{i\theta})|$ decreases and $|Im(e^{i\theta})|$ increases. Thus, our permanent is maximized at $\theta = \frac{\pi}{2}$ (or $\theta = -\frac{\pi}{2}$). This yields the maximizing matrix in the statement of the theorem, and we have:

$$\max_{U\in\mathcal{U}(n)} U^*D_\sigma U = \max_{S\in\mathcal{DQU}(n)} per(S^*D_\sigma S) = \frac{1}{4}(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2) = \frac{1}{2}(\lambda_1^2+\lambda_2^2). \qquad \square$$

Remark This seems to be simpler than the proof given by Grone et. al in [67], and suggests that the simplified problem of maximizing over $\mathcal{DQU}(n)$ may indeed be easier to approach than the original problem as stated in Question 6.9.1.

We will close this chapter with a discussion of *mutually unbiased bases* of Hilbert spaces, and the utility of unitary scalings in the study of these entities.

6.10 Mutually unbiased bases

Originally considered by Schwinger in [69], two orthonormal bases $\mathcal{B}_1, \mathcal{B}_2$ of a Hilbert space V are unbiased if, given a state that is prepared in $\mathcal{B}_1$, all outcomes of a measurement made

in $\mathcal{B}_2$ are equally likely. Equivalently:

Definition Let $\mathcal{B} = \{b_1, ..., b_m\}$ and $\mathcal{C} = \{c_1, ..., c_m\}$ be two orthonormal bases of an m-dimensional Hilbert space V. $\mathcal{B}$ and $\mathcal{C}$ are said to be *mutually unbiased* if:

$$|\langle b_i, c_j \rangle|^2 = \frac{1}{m} \quad \forall i, j = 1, 2, ..., m.$$

A set of k orthonormal bases $\mathcal{S} = \{\mathcal{B}_i\}_{i=1}^k$ is said to be mutually unbiased if all pairs of bases in $\mathcal{S}$ are mutually unbiased.

Let $m \in \mathbb{N}$. The existence (or non-existence) of a set of $m+1$ mutually unbiased bases of $\mathbb{C}^m$ is a famous open problem in quantum information. Motivated by this problem, we ask the following question:

Question 6.10.1 *Let $m > 1$. Given $\mathcal{B}_1$ and $\mathcal{B}_2$, a pair of mutually unbiased (orthonormal) bases of $\mathbb{C}^m$, let $\mathcal{V} = \{v_i\} \subset \mathbb{C}^m$ be the set of all vectors which satisfy $|\langle v_i, b_i \rangle|^2 = \frac{1}{m}$ for all $b_i \in \mathcal{B}_1 \cup \mathcal{B}_2$.*

(1)Is it necessarily the case that $|\mathcal{V}| \geq m$ (up to scalar multiplication)?

(2)Under what conditions can we find a basis of $\mathbb{C}^m$ from this set?

Remark Our latter question is essentially asking what conditions we can place on $\mathcal{B}_1, \mathcal{B}_2$ that will guarantee the existence of a third basis, $\mathcal{B}_3$, such that $\{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}$ form a set of mutually unbiased bases.

Question 6.10.1 can be reformulated in terms of our unitary scalings. To illustrate how, we let the vectors in our first basis, $\mathcal{B}_1$, form the columns of a matrix B_1, and the vectors in

$\mathcal{B}_2$ form the columns of a matrix B_2. As B_1 and B_2 are unitary matrices, we can define the unitary matrix $U := B_2^* B_1$.

Proposition 6.10.2 *A vector $v \in \mathbb{C}^m$ is a biunimodular vector for U if and only if $\frac{B_1 v}{\sqrt{m}}$ is in the set $\mathcal{V}$.*

Proof Suppose that v is a biunimodular vector for U i.e. $v \in \mathbb{T}^m$ and $Uv \in \mathbb{T}^m$. Let b_i and c_j be the i-th and j-th elements of $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively, and let e_k denote the k-th element of the standard order basis. Then

$$\left| \langle \frac{B_1 v}{\sqrt{m}}, b_i \rangle \right|^2 = \frac{1}{m} |\langle B_1 v, B_1 e_i \rangle|^2 = \frac{1}{m} |\langle v, e_i \rangle|^2 = \frac{1}{m},$$

and

$$\left| \langle \frac{B_1 v}{\sqrt{m}}, c_j \rangle \right|^2 = \frac{1}{m} |\langle B_1 v, B_2 e_j \rangle|^2 = \frac{1}{m} \langle B_1 v, B_1 U^* e_j \rangle|^2 = \frac{1}{m} |\langle Uv, e_j \rangle|^2 = \frac{1}{m},$$

and thus $\frac{B_1 v}{\sqrt{m}} \in \mathcal{V}$, as desired.

The reverse direction follows essentially the same argument. $\frac{B_1 v}{\sqrt{m}} \in \mathcal{V}$ guarantees that $|\langle v, e_j \rangle|^2 = 1$ and $|\langle Uv, e_j \rangle|^2 = 1$ for all $j = 1...m$. Thus, v is a biunimodular vector for U. $\square$

Proposition 6.10.2 shows that biunimodular vectors for U are in one-to-one correspondence with mutually unbiased vectors for the set $\{\mathcal{B}_1, \mathcal{B}_2\}$. By our remarks at the beginning of Section 6.8, we know that biunimodular vectors are in one-to-one correspondence with the number of unitary scalings of U. Thus we can approach the problem of finding mutually unbiased bases by finding unitary scalings of U. Indeed, Theorem 6.8.1 guarantees that there

is always at least one element of $\mathcal{V}$, but the method of proof used by Idel and Wolf does not give any more information as to how many elements of $\mathcal{V}$ there might be.

Most of the progress on the cardinality of $\mathcal{V}$ is found in [64]. Therein, we find the following low-dimensional result:

Proposition 6.10.3 ([64], Proposition 5.1) *Let U be a 2×2 unitary matrix with no zero entries. Then U has precisely two biunimodular vectors. Moreover, these vectors are orthogonal to each other.*

Remark In the case where U does have zero entries, it must be one of the two 2×2 permutation matrices, and thus every vector in $\mathbb{T}^2$ is a biunimodular vector for U.

This allows us to conclude the following, answering Question 6.10.1 in the 2-dimensional case:

Corollary 6.10.4 *Given 2 mutually unbiased orthonormal bases of $\mathbb{C}^2$, $\mathcal{B}_1, \mathcal{B}_2$. Then there exists a third basis, $\mathcal{B}_3$, such that $\{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\}$ is mutually unbiased.*

In higher dimensions, we see that this is not always the case:

Proposition 6.10.5 ([64], Example 6.1) *Let U be a 3×3 unitary matrix with exactly one zero entry. Then U has precisely 4 biunimodular vectors v_j. No two of these vectors are orthogonal to each other.*

Proposition 6.10.5 tells us something important about Question 6.10.1.2. As the vectors v_j are not orthogonal to each other, we know that $\frac{B_1 v_j}{\sqrt{m}}$ (as defined above Proposition 6.10.2)

are not orthogonal to each other either. Thus there is no orthonormal basis of vectors in $\mathcal{V}$ and there exists a pair of mutually unbiased bases $\{\mathcal{B}_1, \mathcal{B}_2\}$ of $\mathbb{C}^3$ that cannot be extended to a list of three mutually unbiased bases (this occurs when $B_2^* B_1$ has exactly one zero entry). In the context of Question 6.10.1.2, this tells us that we do indeed need to impose restrictions on our original two bases if we are to guarantee the existence of a third mutually unbiased basis. It is worth noting that we do still have that $|\mathcal{V}| \geq m$ in all observed cases (as suggested in Question 6.10.1.1).

Mutually unbiased bases and the $per(U^* D_\sigma U)$ problem are just a few examples of the utility of Idel and Wolf's unitary scalings. The proof that these scalings exist is relatively new ([65]), and it seems likely that these scalings and the associated Sinkhorn Normal Form of unitary matrices will continue to prove useful in a wide variety of applications.

Chapter 7

Conclusion

We conclude this dissertation with a summary of the novel results introduced herein.

The first original result in the thesis is Theorem 2.4.1. After introducing the scaling results of Sinkhorn, Marshall and Olkin, and Pereira, we combined these into one "housekeeping" result (Theorem 2.4.1) which encompasses all of these theorems. While this is a new proof of old results, the use of minimization and description of $\mathcal{G}$ in the theorem suggests the possibility of a generalization to more matrix Lie groups.

In Theorem 3.4.1, we were able to extend the result of Marshall and Olkin to real, super-symmetric, copositive tensors. This continues a tradition of tensor scalings which was started in the 1980s by mathematicians such as Bapat and Raghavan. We then derived an upper bound for the number of scalings of an m-th order, 2-dimensional tensor in Theorem 3.5.1. This can be seen as an extension of Proposition 4.2.1, originally proven by Pereira and Boneng.

We then directed our attention to positive definite "D^*AD" scalings. Via Example 4.3.1, we disproved a conjecture of Pereira and Boneng on the upper bound of $|sc(A)|$, and then proved that the true upper bound for 3×3 real positive definite matrices is 6 (Theorem 4.4.1). In proving this bound, we were able to arrive at necessary and sufficient conditions for when $sc(A)$ consists entirely of real matrices (Corollary 4.4.6). We ended Chapter 4 by showing that when $n \geq 4$, there exists $n \times n$ real, positive definite matrices that have infinitely many scalings (Theorem 4.5.1).

Theorem 4.5.1 suggests that finding all scalings of a given matrix might be too difficult. Motivated by the connection to the geometric measure of entanglement (Theorem 1.3.3), we shifted our attention to matrices M that have maximal or minimal permanent in their respective equivalence classes, $sc(M)$. We first gave a characterization of the set $MaxSc_n$ in Theorem 5.2.2, and then we used that characterization to investigate the topological properties of $MaxSc_n$ and $MinSc_n$. We showed that $MaxSc_n$ is closed and convex and also arrived at a lower bound on the spectrum of elements from $MaxSc_n$. After discussing a few topological properties of $MinSc_n$, we proved that the identity matrix I_n is the unique element of $MaxSc_n$ that has smallest permanent, and hence that $per(A) \geq 1$ for all $A \in MaxSc_n$ (Theorem 5.5.1).

The techniques from Chapter 5 then allowed us to prove the opposite permanental inequality for certain elements of $MinSc_n$. In particular, Theorem 5.6.1 gives spectral bounds on the real elements of $MinSc_n$ (and $MaxSc_n$), allowing us to show that I_n has the largest permanent of all real elements of $MinSc_n$ (Corollary 5.6.3). We then reached the same

conclusion for abelian group matrices in $MinSc_n$ (Corollary 5.6.6) and for the set $MinSc_3$ (Corollary 5.6.8). We finished the chapter by examining what the permanental bounds attained could tell us about the geometric measure of entanglement of Slater permanents. In Theorems 5.7.1 and 5.7.3, respectively, we gave a lower bound for the GME of Slater permanents with Gram matrices whose minimal scaling is real or an abelian group matrix, and a lower bound for Slater permanents from $(\mathbb{C}^3)^{\otimes 3}$. In these cases, we identified the maximally entangled Slater permanent as the one with pairwise-orthogonal components.

Chapter 6 provided an introduction to a few algorithms that find scalings for us. The most useful algorithm that we presented was Algorithm 2, which seems to always calculate many scalings for any positive definite M. As this is work in progress, we were not able to prove convergence of our algorithms, but we did briefly discuss some results that might allow us to prove convergence in the future. We then discussed a few open problems and how matrix scalings (both D^*AD scalings and unitary scalings) may help us approach these problems in new ways. Most notably, we disproved the Drury permanent conjecture and provided a new proof of the 2×2 $per(U^*D_\sigma U)$ problem (Theorem 6.9.4).

Bibliography

[1] Abraham Berman and Naomi Shaked-Monderer. *Completely positive matrices*. World Scientific, 2003.

[2] Rajesh Pereira and Joanna Boneng. The theory and applications of complex matrix scalings. *Special Matrices*, 2:68–77, 2014.

[3] Rajesh Pereira. Differentiators and the geometry of polynomials. *Journal of Mathematical Analysis and Applications*, 1:336–348, 2003.

[4] Albert W. Marshall and Ingram Olkin. Scaling of matrices to achieve specified row and column sums. *Numerische Mathematik*, 12(1):83–90, 1968.

[5] Dianne P. O'Leary. Scaling symmetric positive definite matrices to prescribed row sums. *Linear Algebra and Its Applications*, pages 185–191, 2003.

[6] Marvin Marcus and Henryk Minc. Permanents. *The American Mathematical Monthly*, 72(6):577–591, 1965.

[7] Leslie G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189–201, 1979.

[8] Fuzhen Zhang. An update on a few permanent conjectures. *Special Matrices*, 4:305–316, 2016.

[9] Marvin Marcus and Morris Newman. Inequalities for the permanent function. *Annals of Mathematics*, 75:47–62, 1962.

[10] Marvin Marcus. On two classical results of I. schur. *Bulletin of the American Mathematical Society*, 70(5):685–688, 1964.

[11] Herbert John Ryser. Compound and induced matrices in combinatorial analysis. In *Proceedings of Symposia in Applied Mathematics, Vol. 10*, pages 149–167, 1960.

[12] Dmitry I. Falikman. A proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix. *Mathematical Notes*, 29:475–479, 1981.

[13] Georgy P. Egorychev. A solution of van der Waerden's permanent problem. *Soviet Mathematics Doklady*, 23:619–622, 1981.

[14] Bartel Leendert van der Waerden. Aufgabe 45. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 35(117):23, 1926.

[15] Jacques Hadamard. Résolution dune question relative aux déterminants. *Bulletin des Sciences Mathématiques*, 17(1):240–246, 1893.

[16] Marvin Marcus. The permanent analogue of the Hadamard determinant theorem. *Bulletin of the American Mathematical Society*, 69(4):494–496, 1963.

[17] Ravindra Bapat. A note on the permanental Hadamard inequality. *Linear and Multilinear Algebra*, 30(3):205–208, 1991.

[18] Henryk Minc. *Permanents*. Addison–Wesley Publishing Co., 1978.

[19] Leon Armenovich Takhtajan. *Quantum mechanics for mathematicians*, volume 95. American Mathematical Society, 2008.

[20] Abner Shimony. Degree of entanglement. *Annals of the New York Academy of Sciences*, 755(1):675–679, 1995.

[21] Tzu-Chieh Wei and Paul M. Goldbart. Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Physical Review A*, 68(4):042307, 2003.

[22] Otfried Gühne, Michael Reimpell, and Reinhard F. Werner. Estimating entanglement measures in experiments. *Physical Review Letters*, 98(11):110502, 2007.

[23] Masahito Hayashi, Damian Markham, Mio Murao, Masaki Owari, and Shashank Virmani. Bounds on multipartite entangled orthogonal state discrimination using local operations and classical communication. *Physical Review Letters*, 96(4):040501, 2006.

[24] Tzu-Chieh Wei. Entanglement under the renormalization-group transformations on quantum states and in quantum phase transitions. *Physical Review A*, 81(6):062313, 2010.

[25] Robert Hübener, Matthias Kleinmann, Tzu-Chieh Wei, Carlos González-Guillén, and Otfried Gühne. Geometric measure of entanglement for symmetric states. *Physical Review A*, 80(3):032324, 2009.

[26] Charles R. Johnson and Robert Reams. Scaling of symmetric matrices by positive diagonal congruence. *Linear and Multilinear Algebra*, 57:123–140, 2009.

[27] Richard Sinkhorn. A relationship between arbitrary positive matrices and stochastic matrices. *Canadian Journal of Mathematics*, 18:303–306, 1966.

[28] Martin Idel. A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps. *arXiv preprint arXiv:1609.06349*, 2016.

[29] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annals of Mathematical Statistics*, 35(2):876–879, 1964.

[30] Marvin Marcus and Morris Newman. The permanent of a symmetric matrix. *Notices of the American Mathematical Society*, 8:595, 1961.

[31] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[32] M.V. Menon. Reduction of a matrix with positive elements to a doubly stochastic matrix. *Proceedings of the American Mathematical Society*, 18(2):244–247, 1967.

[33] Alberto Borobia and Rafael Cantó. Matrix scaling: A geometric proof of Sinkhorn's theorem. *Linear Algebra and Its Applications*, 268:1–8, 1998.

[34] Benno Fuchssteiner and Wolfgang Lusky. *Convex cones*, volume 56. Elsevier, 2011.

[35] Alexander Barvinok. *A Course in Convexity*, volume 54. American Mathematical Society Providence, 2002.

[36] Farid Alizadeh and Donald Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[37] Andrew J Kurdila and Michael Zabarankin. *Convex functional analysis*. Springer Science & Business Media, 2006.

[38] Liqun Qi. Eigenvalues of a real supersymmetric tensor. *Journal of Symbolic Computation*, 40(6):1302–1324, 2005.

[39] Dustin Cartwright and Bernd Sturmfels. The number of eigenvalues of a tensor. *Linear Algebra and Its Applications*, 438(2):942–952, 2013.

[40] L Qi. Eigenvalues of a supersymmetric tensor and positive definiteness of an even degree multivariate form. *Department of Applied Mathematics, The Hong Kong Polytechnic University*, 2004.

[41] Lek-Heng Lim. Singular values and eigenvalues of tensors: a variational approach. In *Computational Advances in Multi-Sensor Adaptive Processing, 2005 1st IEEE International Workshop on*, pages 129–132. IEEE, 2005.

[42] Liqun Qi. Hankel tensors: Associated Hankel matrices and Vandermonde decomposition. *arXiv preprint arXiv:1310.5470*, 2013.

[43] Ravindra Bapat. D1AD2 theorems for multidimensional matrices. *Linear Algebra and Its Applications*, 48:437–442, 1982.

[44] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and Its applications*, 114:717–735, 1989.

[45] T.E.S. Raghavan. On pairs of multidimensional matrices. *Linear Algebra and Its Applications*, 62:263–268, 1984.

[46] Leonid Gurvits. Van der waerden/Schrijver-Valiant like conjectures and stable (aka hyperbolic) homogeneous polynomials: one theorem for all. *Electron. J. Combin*, 15(1), 2008.

[47] George Hutchinson. On the cardinality of complex matrix scalings. *Special Matrices*, 4:141–150, 2016.

[48] Philip J. Davis. *Circulant Matrices*. John Wiley & Sons, 1979.

[49] George Hutchinson. On complex matrix scalings of extremal permanent. *Linear Algebra and Its Applications*, 522:111–126, 2017.

[50] Shigeru Kanemitsu and Michel Waldschmidt. Matrices of finite abelian groups, finite Fourier transform and codes. *Number Theory: Arithmetic in Shangri–La*, 8:90–106, 2013.

[51] Keith Conrad. Characters of finite abelian groups. Lecture Notes. Available at http://www.math.uconn.edu/∼kconrad/blurbs/grouptheory/charthy.pdf, 2010.

[52] R. Chalkley. Matrices derived from finite abelian groups. *Mathematics Magazine*, 49(3):121–129, 1976.

[53] John von Neumann. The geometry of orthogonal spaces, functional operators-vol. ii. *Annals of Mathematics Studies*, 22, 1950.

[54] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[55] Inderjit S Dhillon and Joel A Tropp. Matrix nearness problems with Bregman divergences. *SIAM Journal on Matrix Analysis and Applications*, 29(4):1120–1146, 2007.

[56] Stanley C Eisenstat and Ilse CF Ipsen. Relative perturbation results for eigenvalues and eigenvectors of diagonalisable matrices. *BIT Numerical Mathematics*, 38(3):502–509, 1998.

[57] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[58] John Chollet. Is there a permanental analogue to Oppenheim's inequality? *The American Mathematical Monthly*, 89(1):57–58, 1982.

[59] Robert John Gregorac and Irvin Roy Hentzel. A note on the analogue of Oppenheim's inequality for permanents. *Linear Algebra and Its Applications*, 94:109–112, 1987.

[60] George Hutchinson. A counterexample to the Drury permanent conjecture. *Special Matrices*, 5(1):301–302, 2017.

[61] Göran Björck. Functions of modulus 1 on z n whose Fourier transforms have constant modulus, and cyclic n-roots. In *Recent Advances in Fourier Analysis and Its Applications*, pages 131–140. Springer, 1990.

[62] Uffe Haagerup. Orthogonal maximal abelian *-subalgebras of the nxn matrices and cyclic n-roots. In S. Doplicher et. al, editor, *Operator Algebras and Quantum Field Theory*, pages 296–322. Accademia Nazionale dei Lincei/Intenrational Press, Roma, Italy/Cambridge, MA, 1997.

[63] John Gilbert and Ziemowit Rzeszotnik. The norm of the Fourier transform on finite abelian groups (la norme de la transformée de Fourier sur les groupes abéliens finis). In *Annales de l'institut Fourier*, volume 60, pages 1317–1346, 2010.

[64] Hartmut Führ and Ziemowit Rzeszotnik. On biunimodular vectors for unitary matrices. *Linear Algebra and Its Applications*, 484:86–129, 2015.

[65] Martin Idel and Michael M Wolf. Sinkhorn normal form for unitary matrices. *Linear Algebra and Its Applications*, 471:76–84, 2015.

[66] Marvin Marcus and Henryk Minc. Permanents. *The American Mathematical Monthly*, 72(6):577–591, 1965.

[67] Robert Grone, Charles R Johnson, SA Eduardo, and Henry Wolkowicz. A note on maximizing the permanent of a positive definite hermitian matrix, given the eigenvalues. *Linear and Multilinear Algebra*, 19(4):389–393, 1986.

[68] Philip J Davis and Igor Najfeld. Equisum matrices and their permanence. *Quarterly of Applied Mathematics*, 58(1):151–169, 2000.

[69] Julian Schwinger. Unitary operator bases. *Proceedings of the National Academy of Sciences*, 46(4):570–579, 1960.