An introduction to majorization and its applications to quantum mechanics

Michael A. Nielsen

Department of Physics, Univesity of Queensland,
Queensland 4072, Australia

October 18, 2002

Contents

Preface

Majorization is a powerful, easy-to-use and flexible mathematical tool which can be applied to a wide variety of problems in quantum mechanics. This book surveys the basic results of the theory of majorization, emphasizing the connections to quantum mechanics, and discussing a number of open problems.

The book is based on lectures given at the University of Queensland in 2002, and at the California Institute of Technology in 1998. Thus, the book is organized into chapters, each of which corresponds to roughly a one-hour lecture. However, the chapters contain supplementary material that could not typically be covered in a one-hour lecture. Much of this supplementary material is in the form of exercises scattered through the text, which I encourage readers to attempt as they read. More difficult problems may be found at the end of each chapter, along with some hints for the exercises and problems contained in the chapter.

Structure of the book

The wide applicability of majorization to quantum mechanics arises as a result of two simple but deep theorems connecting majorization to unitary matrices: *Horn's lemma*, and *Uhlmann's theorem*. These connections, and the ubiquity of unitary matrices in quantum mechanics, make majorization a powerful tool for the mathematical arsenal of a quantum theorist.

In view of these connections, I begin the book with an overview chapter, explaining the basic definitions of majorization, the statement of Horn's lemma and Uhlmann's theorem, and giving some examples illustrating the power of these theorems when applied to quantum mechanics.

The strategy of the remainder of the book is to build up the mathematical theory of majorization, while interspersing applications to quantum mechanics. Chapters 2 and 3 begin our development of the mathematical theory of

majorization, building up to a powerful result of Hardy, Littlewood and Polya connecting majorization to the *doubly stochastic* matrices. This connection enables us in Chapter 5 to prove Horn's lemma, connecting majorization to unitary matrices and thus to quantum mechanics.

The end matter of the book contains three appendices, a bibliography, and an index.

Points of special interest

I have tried to make the book accessible to anybody with a background in elementary quantum mechanics. The recent developments connecting majorization to quantum mechanics have largely occurred within the new subfield of *quantum information science*, and the point of view adopted here inevitably reflects that connection. Nonetheless, I have tried to make the results of the text accessible to readers without a background in quantum information science. To this end, a guide to nomenclature and notation may be found after this preface, should you get lost in unfamiliar notation.

One of the major goals of this book is to bring the reader into the thick of the exciting recent progress in applying majorization to problems in quantum mechanics. To that end, throughout the book I have emphasized unsolved problems whose solution can be expected to be related to the ideas and techniques introduced in the text.

Throughout the text I have given references to recent papers connecting majorization and quantum mechanics, in the hope that these will serve as useful leads to the reader interested in pursuing further work in this area. Of course, the usual caveats apply: despite the best of intentions, my knowledge is limited, and my apologies to any researcher whose work I have inadvertently omitted from citation. For older work on the basic theory of majorization, the historical coverage is much less complete, due in part to my own ignorance, and in part due to the existence of standard texts on majorization that present the history in a much more comprehensive way than is possible in a more specialized text such as this.

In this vein, let me mention Marshall and Olkin's classic text[38], which gives a very nearly comprehensive coverage of both the theory and history of majorization, up until 1979. Any reader seriously interested in majorization will eventually need to take a good look at Marshall and Olkin. It contains a wealth of additional material I have not covered, including many applications of majorization outside of physics. My own introduction to majorization was

through Chapters 2 and 3 of the book by Bhatia[8], and my presentation bears the mark of this influence. Alberti and Uhlmann[1] have also written a more specialized monograph on majorization that may be of interest to many readers. Finally, Ando has written an excellent pair[2, 3] of survey articles on majorization that provide a brief introduction to the subject, and cover much of the field's development since the publication of Marshall and Olkin.

Acknowledgments

This book is descended from notes written for a lecture series on majorization and its applications to quantum mechanics, which I gave at the California Institute of Technology during June of 1999. This book was developed during a more extended lecture series given at the University of Queensland during the second half of 2002. Many thanks to all participants in both lecture series, whose suggestions, questions, and comments have greatly improved my understanding, and this presentation.

My understanding of majorization has especially benefited from discussion with Howard Barnum, Dave Beckman, John Cortese, Sumit Daftuar, Chris Fuchs, Bob Gingrich, Lucien Hardy, Daniel Jonathan, Julia Kempe, Matt Klimesh, Andrew Landahl, John Preskill, Beth Ruskai, Ben Schumacher, Federico Spedalieri, Armin Uhlmann and Guifre Vidal, and my thanks goes to all of them.

Notation and nomenclature

Vectors will sometimes be written in the column format,

$$\left[\begin{array}{c} 0 \\ 1 \end{array} \right], \tag{1}$$

and sometimes for readability in the format $(0, 1)$. The latter should be understood as shorthand for the column vector, and not as a row vector. We use the notation r and s generically to denote real vectors in a d-dimensional real vector space. It is often convenient to re-order the components of such a vector $r = (r_1, \ldots, r_d)$ into decreasing order, $r^{\downarrow} = (r_1^{\downarrow}, \ldots, r_d^{\downarrow})$, where $r_1^{\downarrow} \geq r_2^{\downarrow} \geq \ldots \geq r_d^{\downarrow}$. Note that we say *decreasing* where some readers might prefer *non-increasing*; any possible ambiguity in our presentation is removed by saying that a sequence $r > s > \ldots$ is *strictly* decreasing. In a similar vein, we may re-order the components of r into increasing order, $r^{\uparrow} = (r_1^{\uparrow}, \ldots, r_d^{\uparrow})$, where $r_1^{\downarrow} \leq r_2^{\downarrow} \leq \ldots \leq r_d^{\uparrow}$.

Part I

Introduction to majorization

Chapter 1

Overview

Majorization is a mathematical relation that allows us to determine which of two probability distributions is more disordered. The present lecture provides an overview of majorization and its connections to quantum mechanics. In particular, we introduce the basic definitions of the theory of majorization in Section 1.1, explain two fundamental results (Horn's lemma and Uhlmann's theorem) connecting majorization and quantum mechanics in Section 1.2, and give some examples applications of majorization to quantum mechanics in Section 1.3. Finally, Section 1.4 outlines the reasons why the concept of majorization is needed, even though other concepts such as the Shannon entropy can be used for a similar purpose, namely, to quantify the notion that one probability distribution is more mixed than another.

1.1 What is majorization?

The intuition motivating majorization may be understood from the following definition: we say the d-dimensional real vector r is *majorized* by the d-dimensional real vector s, written $r \prec s$, if there exist d-dimensional permutation matrices P_j and a probability distribution $\{p_j\}$ such that

$$r = \sum_j p_j P_j s. \tag{1.1}$$

That is, r is majorized by s precisely when r can be obtained from s by randomly permuting the components of s, and then averaging over the permutations. At least naively this is a natural and appealing approach to

defining the notion that one vector is more disordered than another, and we will see that this naive appeal is more than justified by the rich mathematical structure arising from this definition.

As a simple example of majorization, suppose the vector s is a probability distribution on d outcomes, that is, the components are non-negative and sum to one. Then it is easy to see that

$$\left(\frac{1}{d}, \ldots, \frac{1}{d}\right) \prec s, \tag{1.2}$$

since the uniform distribution $(1/d, \ldots, 1/d)$ may be obtained by averaging over permutations $P_\pi s$ of s, where π is chosen uniformly at random from the symmetric group on n elements. This simple example agrees with our intuition that the uniform distribution on d elements is as or more disordered than any other probability distribution over d elements. It also illustrates a point that will be generically true throughout these notes: most often we use majorization to compare two *probability distributions*, that is, real vectors whose components are non-negative and sum to one.

Our definition for the majorization relation $r \prec s$ in terms of random permutations is satisfying from an intuitive point of view, and is often useful when proving theoretical results, but is perhaps not so useful for actual calcuations. Given two vectors of numbers r and s is there an effective procedure to determine whether $r \prec s$? Remarkably, such a procedure exists. First, we re-order the components of r and s into non-increasing order, writing, for example, $r^\downarrow = (r_1^\downarrow, \ldots, r_d^\downarrow)$ for the vector whose components are the same as those of r, but ordered so that

$$r_1^\downarrow \geq r_2^\downarrow \geq \ldots \geq r_d^\downarrow. \tag{1.3}$$

We prove in Chapter 3 that $r \prec s$ if and only if

$$\begin{aligned}
r_1^\downarrow &\leq s_1^\downarrow && (1.4)\\
r_1^\downarrow + r_2^\downarrow &\leq s_1^\downarrow + s_2^\downarrow && (1.5)\\
r_1^\downarrow + r_2^\downarrow + r_3^\downarrow &\leq s_1^\downarrow + s_2^\downarrow + s_3^\downarrow && (1.6)\\
&\vdots \\
r_1^\downarrow + \ldots + r_{d-1}^\downarrow &\leq s_1^\downarrow + \ldots + s_{d-1}^\downarrow && (1.7)\\
r_1^\downarrow + \ldots + r_d^\downarrow &= s_1^\downarrow + \ldots + s_d^\downarrow. && (1.8)
\end{aligned}$$

Note the equality appearing in the last expression. One way of looking at this equality is that any permutation of s leaves the sum of the components invariant, as does averaging over a set of permutations, and thus r and s must have the same sum. In any case, in the applications we shall be most concerned with, r and s are probability distributions, and thus the equality is automatically satisfied.

Exercise 1.1.1: (**Majorization and triangles**) Let $\theta_1, \theta_2, \theta_3$ be the angles of a triangle, expressed in radians. Show that

$$\left(\frac{\pi}{3}, \frac{\pi}{3}, \frac{\pi}{3}\right) \prec (\theta_1, \theta_2, \theta_3) \prec (\pi, 0, 0). \tag{1.9}$$

Find vectors a and b such that the conditions $a \prec (\theta_1, \theta_2, \theta_3) \prec b$ characterize (a) the acute triangles and (b) the obtuse triangles. A wide variety of geometric inequalities follow from the techniques of majorization. See Chapter 8 of Marshall and Olkin[38].

1.2 What connects majorization and quantum mechanics?

What connections are there between majorization and quantum mechanics? The quantum mechanical analogue of a probability distribution is the density matrix, so a natural beginning is to define a matrix notion of majorization. Supposing R and S are d-dimensional Hermitian matrices, we define $R \prec S$ (read "R is majorized by S") if $\lambda(R) \prec \lambda(S)$, where $\lambda(R)$ denotes the vector whose components are the eigenvalues of R, arranged in non-increasing order. Just as for vectors we are mostly concerned with the majorization relation for probability distributions, so for matrices we will mostly be concerned with the majorization relation for density matrices.

As an example of matrix majorization, suppose ρ is any density matrix for a quantum system with a d-dimensional state space. Then $I/d \prec \rho$, where I/d is the completely mixed state. This follows immediately from the observation made in the previous section that the uniform probability distribution on d outcomes is majorized by any other probability distribution on d outcomes.

Exercise 1.2.1: (**Majorization on the Bloch sphere**) Let $\vec{a}$ and $\vec{b}$ be real three-dimensional vectors of length at most one, and let ρ_a and ρ_b be the qubit states with Bloch vectors $\vec{a}$ and $\vec{b}$, that is,

$$\rho_a \equiv (I + \vec{a} \cdot \vec{\sigma})/2; \quad \rho_b \equiv (I + \vec{b} \cdot \vec{\sigma})/2. \tag{1.10}$$

Show that $\rho_a \prec \rho_b$ if and only if $\|a\| \leq \|b\|$.

The essential reason for the close connection between majorization and quantum mechanics may be appreciated by inspection of two elegant (and closely related) results: *Horn's lemma* and *Uhlmann's theorem*. Horn's lemma states that for vectors r and s, $r \prec s$ if and only if $r_i = \sum_j |u_{ij}|^2 s_j$ for some unitary matrix $u = (u_{ij})$ of complex numbers. Uhlmann's theorem states that $R \prec S$ for Hermitian matrices R and S if and only if there exist unitary matrices U_j and a probability distribution $\{p_j\}$ such that

$$R = \sum_j p_j U_j S U_j^\dagger. \tag{1.11}$$

The fundamental place of unitarity in quantum mechanics ensures that relations of the type featuring in Horn's lemma and Uhlmann's theorem arise frequently, and it is this which accounts for many of the applications of majorization to quantum mechanics.

1.3 Applications of majorization to quantum mechanics

In this section we take a quick peek at some specific applications of majorization to quantum mechanics, leaving the proofs until later.

On the probability of measurement outcomes

Suppose a quantum system with d-dimensional state space is in a state described by a density matrix ρ with vector of eigenvalues $\lambda(\rho) = (\lambda_1(\rho), \ldots, \lambda_d(\rho))$, so that $\rho = \sum_j \lambda_j(\rho) |j\rangle\langle j|$, where $|j\rangle$ are orthonormal eigenvectors of ρ . Suppose we measure the system in an orthonormal basis $|e_k\rangle$. Then result k occurs with probability

$$\begin{aligned} p(k) &= \langle e_k|\rho|e_k\rangle & (1.12) \\ &= \sum_j \lambda(j) |u_{jk}|^2, & (1.13) \end{aligned}$$

where $u_{jk} \equiv \langle e_k|j\rangle$ is a unitary matrix. From Horn's lemma it follows that the probability distribution $(p(k))$ is majorized by $\lambda(\rho)$. Conversely, given a probability distribution $(p(k))$ majorized by $\lambda(\rho)$, Horn's lemma implies that there exists an orthonormal basis $|e_k\rangle$ such that measuring ρ in that basis will give the outcome k with probability $p(k)$. We have proved the following theorem:

> *Theorem 1.3.1*: Let ρ be a density matrix. Then there exists an orthonormal basis $|e_k\rangle$ such that a measurement in the basis $|e_k\rangle$ yields probabilities $p(k)$ if and only if $(p(k)) \prec \lambda(\rho)$.

Exercise 1.3.1: Let ρ be an arbitrary state of a d-dimensional quantum system. Prove that there always exists an orthonormal basis $|e_k\rangle$ such that the probabilities for a measurement in that basis are uniformly distributed. Given ρ can you explicitly construct a basis $|e_k\rangle$ such that this is true?

Quantum measurement without post-selection

Suppose a quantum system is in the state ρ, and a von Neumann measurement described by a complete set of orthonormal projectors $P_1, \ldots, P_n$ occurs. To an observer who does not learn the result of the measurement the state of the system after the measurement is described by the posterior density matrix:

$$\rho' = \sum_j P_j \rho P_j. \tag{1.14}$$

We will use Uhlmann's theorem to show that $\rho' \prec \rho$, which makes precise the intuitive notion that ρ' is "more mixed" than ρ. Define matrices $U_1, \ldots, U_n$ by

$$U_k \equiv \sum_j \omega^{jk} P_j, \tag{1.15}$$

where $\omega \equiv \exp(2\pi i/n)$ is an nth root of unity. Note that the matrices U_k are unitary matrices. Furthermore, for any density matrix ρ,

$$\frac{\sum_k U_k \rho U_k^\dagger}{n} = \frac{\sum_{j_1 j_2 k} \omega^{j_1 k - j_2 k} P_{j_1} \rho P_{j_2}}{n}. \tag{1.16}$$

Substituting $\sum_k \omega^{j_1 k - j_2 k} = n\delta_{j_1 j_2}$ gives

$$\frac{\sum_k U_k \rho U_k^\dagger}{n} = \sum_j P_j \rho P_j = \rho', \tag{1.17}$$

and thus, by Uhlmann's theorem, $\rho' \prec \rho$. We have proved the following theorem:

> *Theorem 1.3.2*: Let ρ be a density matrix, and P_j a complete set of orthonormal projectors. Then the posterior density matrix $\rho' = \sum_j P_j \rho P_j$ satisifies $\rho' \prec \rho$.

Exercise 1.3.2: Suppose ρ is a single-qubit state subjected to a von Neumann measurement, with posterior state ρ'. Show that the Bloch vector of ρ' is never longer than the Bloch vector of ρ. Thus, the process of measurement moves the density matrix toward the middle of the Bloch sphere.

Characterizing the probabilities that appear in a decomposition of a density matrix

The next result is a beautiful constraint on the static properties of the density matrix. A given density matrix ρ may be represented in many different ways as an ensemble $\{r_j, |\psi_j\rangle\}$ of pure states,

$$\rho = \sum_j r_j |\psi_j\rangle\langle\psi_j|. \tag{1.18}$$

We show in Appendix C that for a fixed vector $r = (r_j)$ of probabilities there exists a set of pure states $|\psi_j\rangle$ such that (1.18) is true if and only if $r \prec \lambda(\rho)$. (Note that either $\lambda(\rho)$ or r may need to be "padded" with extra zeroes in order that they have the same dimension, and thus be comparable using the majorization relation.) Thus there is a fundamental connection between the static properties of the density matrix and majorization, a connection which we will see has implications for other fundamental quantities such as measurement probabilities.

Quantum measurements acquire information about the system being measured

Our final example of an application of majorization to quantum mechanics is a set of *dynamical* constraints on quantum measurement. Intuitively, we know that quantum measurements acquire (rather than lose) information about the system being measured. We'll see that this intuition can be made mathematically precise: if ρ is the initial state of a quantum system being measured and ρ_m are the post-measurement states conditioned on the measurement result m occurring, then we show in Chapter 6 that

$$\lambda(\rho) \prec \sum_m p_m \lambda(\rho_m), \tag{1.19}$$

where p_m is the probability for measurement outcome m. Thus, the eigenvalues of the initial state are more disordered than the average eigenvalues of the post-measurement state, in accord with our intuition that quantum measurements acquire information. A type of converse to this result also holds: provided an equation similar to (1.19) holds (with some additional technical restrictions), it is possible to find a quantum measurement which gives the post-measurement state ρ_m with probability p_m when performed with ρ as the initial state. Thus majorization provides a natural language to express sharp fundamental constraints on the ability of quantum measurements to acquire information about a quantum system.

1.4 Why do we need the concept of majorization?

Despite the applications to quantum mechanics that we've seen, you might ask why we need the notion of majorization when measures of disorder such as the Shannon and von Neumann entropies are already available? Couldn't these other measures be put to the same use in applications as we put majorization? These are good questions.

It turns out that the entropic measures arise naturally out of the theory of majorization in a sort of "law of large numbers limit" where we are considering a large number of identical systems. We'll examine how this works in detail in Chapter 8, but the essential point is that measures such as the entropy are essentially *weaker* than the notion of majorization, and as such

do not give as much detailed information as is provided by majorization. In particular, we will show that the tools of majorization can be applied to problems for which the concept of entropy is *insufficient* for the analysis.

Exercise 1.4.1: Let r and s be probability distributions such that $r \prec s$. Assume, as was stated earlier, that there exist probabilities p_j and permutations P_j such that $r = \sum_j p_j P_j s$. Use this fact to prove that $H(r) \geq H(s)$, where $H(r) \equiv -\sum_j r_j \log(r_j)$ is the Shannon entropy of a probability distribution. Note that we use the information-theoretic convention that logarithms are always taken to base two, unless otherwise noted.

Problems for Lecture 1

Problem 1.4.1: (**Alternate proof of Theorem 1.3.2**) Show that $\sum_j P_j \rho P_j \prec \rho$ for a complete orthonormal set of projectors by proving the result for a set of two orthonormal projectors, and then using induction.

Hints for Lecture 1

Hint for Exercise 1.3: The basis $|e_k\rangle$ to measure in is the Fourier transform of the eigenbasis of ρ.

Chapter 2

Elementary properties of majorization

This lecture begins our in-depth investigation of majorization by exploring some elementary properties of the majorization relation. In the last lecture we discussed several equivalent characterizations of majorization, each of which could potentially be used as a fundamental definition. Section 2.1 fixes a single fundamental definition for the majorization relation, which we use as the basis for all our later development, and explores the most basic properties of that relation. An occasionally irritating aspect of our fundamental definition is that in order to use it to compare two vectors, r and s, it is necessary to order the components of the vectors into non-increasing order. Section 2.2 gives a useful alternative characterization of majorization that does not require such an ordering, and which we therefore refer to as the *order-free* characterization of majorization.

2.1 Definition and most basic properties

We begin by fixing our fundamental definition of the majorization relation: $r = (r_1, \ldots, r_d) \prec s = (s_1, \ldots, s_d)$ if

$$\begin{aligned} r_1^{\downarrow} &\leq s_1^{\downarrow} && (2.1)\\ r_1^{\downarrow} + r_2^{\downarrow} &\leq s_1^{\downarrow} + s_2^{\downarrow} && (2.2)\\ r_1^{\downarrow} + r_2^{\downarrow} + r_3^{\downarrow} &\leq s_1^{\downarrow} + s_2^{\downarrow} + s_3^{\downarrow} && (2.3)\\ &\vdots \end{aligned}$$

$$r_1^{\downarrow} + \ldots + r_{d-1}^{\downarrow} \leq s_1^{\downarrow} + \ldots + s_{d-1}^{\downarrow} \tag{2.4}$$

$$r_1^{\downarrow} + \ldots + r_d^{\downarrow} = s_1^{\downarrow} + \ldots + s_d^{\downarrow}. \tag{2.5}$$

It will be convenient to write the last equation in one of two different forms; either $\sum_{j=1}^d r_j^{\downarrow} = \sum_{j=1}^d s_j^{\downarrow}$ or $\sum_{j=1}^d r_j = \sum_{j=1}^d s_j$. Since the sum is over all components of the vectors, the ordering does not matter. The definition of Equations (2.1)-(2.5) has the substantial advantage that it provides an easy and obvious computational procedure for comparing two vectors. Working from this definition we will establish a wide variety of elementary properties of majorization, eventually arriving in the next lecture at the equivalent characterization described in the last lecture, namely, that $r \prec s$ if and only if $r = \sum_j p_j P_j s$ for some probabilites $\{p_j\}$ and permutation matrices P_j.

In Equations (2.1) through (2.5) we have defined majorization in terms of a non-increasing order on the elements of the vectors being compared. Not surprisingly, there is an equivalent definition in terms of a non-decreasing order on the elements of the vectors being compared. Let $r^{\uparrow} = (r_1^{\uparrow}, \ldots, r_d^{\uparrow})$ denote the vector whose elements are the elements of r re-ordered into non-decreasing order,

$$r_1^{\uparrow} \leq r_2^{\uparrow} \leq \ldots \leq r_d^{\uparrow}. \tag{2.6}$$

Then we have the following characterization of majorization:

Theorem 2.1.1: $r \prec s$ if and and only if

$$\sum_{j=1}^k r_j^{\uparrow} \geq \sum_{j=1}^k s_j^{\uparrow}, \tag{2.7}$$

for $k = 1, \ldots, d$, with equality when $k = d$.

Proof: It is clear that the condition $\sum_{j=1}^d r_j^{\uparrow} = \sum_{j=1}^d s_j^{\uparrow}$ is equivalent to the condition $\sum_{j=1}^d r_j^{\downarrow} = \sum_{j=1}^d s_j^{\downarrow}$, so we need only check that the inequality (2.7) is equivalent to the inequality in the standard definition of majorization. To see this, define $T \equiv \sum_{j=1}^d r_j = \sum_{j=1}^d s_j$ and note that for $k = 1, \ldots, d-1$,

$$\sum_{j=1}^k r_j^{\uparrow} = T - \sum_{j=1}^{d-k} r_j^{\downarrow}; \quad \sum_{j=1}^k s_j^{\uparrow} = T - \sum_{j=1}^{d-k} s_j^{\downarrow}. \tag{2.8}$$

By substitution of these equations, it follows that

$$\sum_{j=1}^{k} r_j^{\uparrow} \geq \sum_{j=1}^{k} s_j^{\uparrow} \tag{2.9}$$

if and only if

$$\sum_{j=1}^{d-k} r_j^{\downarrow} \leq \sum_{j=1}^{d-k} s_j^{\downarrow}, \tag{2.10}$$

which establishes the desired equivalence. ■

Two variants of majorization are sometimes also useful. Let r and s be d-dimensional real vectors. Then r is *sub-majorized* by s, written $r \prec_w s$, if

$$\sum_{j=1}^{k} r_j^{\downarrow} \leq \sum_{j=1}^{k} s_j^{\downarrow} \tag{2.11}$$

for each k in the range 1 through d. The only difference between sub-majorization and majorization is the omission of the equality requirement at $k = d$ in the definition of sub-majorization. Similarly, we say that r is *super-majorized* by s, written $r \prec^w s$, if

$$\sum_{j=1}^{k} r_j^{\uparrow} \geq \sum_{j=1}^{k} s_j^{\uparrow} \tag{2.12}$$

for each k in the range 1 through d.

Theorem 2.1.2:

1. $r \prec s$ if and only if $r \prec_w s$ and $r \prec^w s$.
2. $r \prec s$ if and only if $-r \prec -s$.
3. Let α be any real number. Then $r \prec s$ implies $\alpha r \prec \alpha s$.
4. The relations $\prec, \prec_w$ and $\prec^w$ are reflexive and transitive.
5. $r \prec s$ and $s \prec r$ if and only if r is a permutation of the elements of s, that is, $r^{\downarrow} = s^{\downarrow}$.

Proof:

1. The forward implication follows immediately from the definition of $r \prec s$ and Theorem 2.1.1.

   To see the reverse implication, note that since $r \prec_w s$ and $r \prec^w s$ we have, respectively:

   $$\sum_{j=1}^{d} r_j = \sum_{j=1}^{d} r_j^{\downarrow} \leq \sum_{j=1}^{d} s_j^{\downarrow} \sum_{j=1}^{d} s_j; \quad \sum_{j=1}^{d} r_j = \sum_{j=1}^{d} r_j^{\uparrow} \geq \sum_{j=1}^{d} s_j^{\uparrow} = \sum_{j=1}^{d} s_j, \tag{2.13}$$

   and thus $\sum_{j=1}^{d} r_j = \sum_{j=1}^{d} s_j$. Together with the conditions for $r \prec_w s$ this implies that $r \prec s$.

2. Suppose $r \prec s$. Then some simple algebra and Theorem 2.1.1 yield

   $$\sum_{j=1}^{k} (-r)_j^{\downarrow} \;\; = \;\; -\sum_{j=1}^{k} r_j^{\uparrow} \leq -\sum_{j=1}^{k} s_j^{\uparrow} = \sum_{j=1}^{k} (-s)_j^{\downarrow}, \tag{2.14}$$

   with equality when $k = d$, so $-r \prec -s$. If $-r \prec -s$ then the previous argument shows that $-(-r) \prec -(-s)$, so $r \prec s$, and thus $r \prec s$ if and only if $-r \prec -s$.

3. Suppose $r \prec s$ and $\alpha \geq 0$. It is clear that $\alpha r \prec \alpha s$. Combining this observation with the previous result completes the proof.

4. All these results follow immediately from the transitivity and reflexivity of the relation $\leq$ on the real numbers.

5. A straightforward induction shows that $r_j^{\downarrow} = s_j^{\downarrow}$ for $j = 1, \ldots, d$.

■

Since the relation $\prec$ is transitive it defines a partial pseudo-order on real vectors. If we restrict ourselves to the comparison of *ordered vectors* r and s, that is, vectors such that $r = r^{\downarrow}$ and $s = s^{\downarrow}$, then $\prec$ forms a true partial order, and that is the terminology we will mostly use. Majorization gives only a *partial* rather than a *total* order on ordered vectors, since there are vectors r and s which are *incomparable* in the sense that $r \not\prec s$ and $s \not\prec r$, where $r \not\prec s$ means that r is *not* majorized by s. An obvious way in which this may occur is if $\sum_{j=1}^{d} r_j \neq \sum_{j=1}^{d} s_j$. However, even when the total sums of the elements of the vector are the same, as for the comparison of

two probability distributions, the phenomenon of incomparability may still occur. An example is the two probability distributions:

$$r = (0.5, 0.25, 0.25); \quad s = (0.4, 0.4, 0.2). \tag{2.15}$$

Note that $s_1^{\downarrow} = 0.4 < 0.5 = r_1^d a$, so $r \not\prec s$. Similarly, $r_1^{\downarrow} + r_2^{\downarrow} = 0.75 < 0.8 = s_1^{\downarrow} + s_2^{\downarrow}$, so $s \not\prec r$. Note that it is often useful to have examples of vectors incomparable by the majorization relation, and we will use the example of Equation (2.15) several times in later lectures.

Exercise 2.1.1: Suppose $r = (p, 1-p)$ and $s = (q, 1-q)$ are two two-element probability distributions. Show that r and s are always comparable. That is, either $r \prec s$, or $s \prec r$.

2.2 An order-free characterization of majorization

The definition of majorization in Equations (2.1)-(2.5) requires that the elements of the vectors to be compared are ordered in decreasing order. There is a useful alternate characterization of majorization that does not require such an ordering:

Theorem 2.2.1: (**Order-free characterization**)
$r \prec s$ if and only if $\sum_{j=1}^d r_j = \sum_{j=1}^d s_j$, and for all real t,

$$\sum_{j=1}^{d} (r_j - t)^+ \leq \sum_{j=1}^{d} (s_j - t)^+, \tag{2.16}$$

where $z^+ \equiv \max(z, 0)$ is the positive part of any real number.

Proof: We show the forward implication first. Suppose $r \prec s$ and let t be any real number. We analyse three cases.
Case 1: $t \leq r_d^{\downarrow}$. Then $(r_j^{\downarrow} - t) \geq 0$ for all j, and thus

$$\sum_{j=1}^{d} (r_j - t)^+ = \sum_{j=1}^{d} (r_j - t) \leq \sum_{j=1}^{d} (s_j - t) \leq \sum_{j=1}^{d} (s_j - t)^+. \tag{2.17}$$

Case 2: $r_1^{\downarrow} \leq t$. Then $(r_j - t)^+ = 0$ for all j and thus

$$\sum_{j=1}^{d}(r_j - t)^+ = 0 \leq \sum_{j=1}^{d}(s_j - t)^+. \tag{2.18}$$

Case 3: For some k, $r_{k+1}^{\downarrow} \leq t \leq r_k^{\downarrow}$. Then

$$\sum_{j=1}^{d}(r_j - t)^+ \;=\; \sum_{j=1}^{k}(r_j^{\downarrow} - t) \leq \sum_{j=1}^{k}(s_j^{\downarrow} - t) \leq \sum_{j=1}^{d}(s_j - t)^+. \tag{2.19}$$

Conversely, suppose Equation (2.16) holds for all t. Set $t \equiv s_{k+1}^{\downarrow}$. Then

$$\begin{aligned}
\sum_{j=1}^{k}(r_j^{\downarrow} - t) &\leq \sum_{j=1}^{k}(r_j^{\downarrow} - t)^+ && (2.20)\\
&\leq \sum_{j=1}^{d}(r_j - t)^+ && (2.21)\\
&\leq \sum_{j=1}^{d}(s_j - t)^+ && (2.22)\\
&= \sum_{j=1}^{k}(s_j^{\downarrow} - t). && (2.23)
\end{aligned}$$

Adding kt to both sides gives $\sum_{j=1}^{k} r_j^{\downarrow} \leq \sum_{j=1}^{k} s_j^{\downarrow}$, and by the assumptions of the theorem we already know $\sum_{j=1}^{k} r_j^{\downarrow} = \sum_{j=1}^{k} s_j^{\downarrow}$, so $r \prec s$, as we set out to show. ■

Exercise 2.2.1: (**Extension property of majorization**) Show that $r \prec s$ implies that $(r, u) \prec (s, u)$ for any vector u.

Exercise 2.2.2: Let e be the d-dimensional vector containing all 1s. Show that if $r \prec s$ then $r + \alpha e \prec s + \alpha e$, for all real α.

Exercise 2.2.3: Find an example of vectors r, s, t such that $r \prec s$ but $r + t \not\prec s + t$.

Exercise 2.2.4: Show that for any t which is not a multiple of the vector e of all 1s, there exist r and s such that $r \prec s$, but $r + t \not\prec s + t$.

Hints for Lecture 2

Hints for Exercises 2.2 and 2.2: Use the order-free characterization of majorization, Theorem 2.2.1.

Chapter 3

Double stochasticity and majorization

Thus far we have two equivalent definitions for majorization, one in terms of a set of inequalities, (2.1)-(2.5), the other the order-free characterization of majorization, Theorem 2.2.1. In this chapter we will use the order-free characterization to develop a truly powerful characterization of majorization in terms of *doubly stochastic matrices*.

3.1 Introduction

A real d by d matrix $D = (D_{ij})$ is *doubly stochastic* if the entries of D are non-negative, and each row and column of D sums to 1. A simple example of a doubly stochastic matrix is

$$D = \begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix}, \tag{3.1}$$

where t is a parameter in the range 0 to 1. This is the most general 2 by 2 doubly stochastic matrix; once the parameter in the top left corner is chosen, it determines the entries in the bottom left and top right, since columns and rows sum to 1, and these entries in turn determine the entry in the bottom right.

Doubly stochastic matrices can be interpreted as a type of noisy communications channel. Imagine a channel has as input some probability distribution (p_j), and an output distribution (q_k). The output probabilities are

linearly related to the input probabilities by a set of *transition probabilities* $p(k|j)$,

$$q_k = \sum_j p(k|j)p_j. \tag{3.2}$$

Generically, these transition probabilities are non-negative and satisfy the relation $\sum_k p(k|j) = 1$, since, given some fixed input j, the respective probabilities $p(k|j)$ for all the different possible outpts k must sum to one. Supposing we regard the entries D_{kj} of a doubly stochastic matrix D as transition probabilities for a noisy channel, then both these requirements are satisfied. The property of double stochasticity also gives rise to an additional constraint, if the entries of D are to be regarded as transition probabilities:

$$\sum_j p(k|j) = \sum_j D_{kj} = 1, \tag{3.3}$$

a requirement which is not generically true for all sets of transition probabilities $p(k|j)$. This extra constraint corresponds to the requirement that the uniform distribution $(1/d, 1/d, \ldots, 1/d)$ is a stationary state of the channel. This may be seen by observing that the requirement that the rows of D sum to one is equivalent to the condition that

$$D \begin{bmatrix} \frac{1}{d} \\ \frac{1}{d} \\ \vdots \\ \frac{1}{d} \end{bmatrix} = \begin{bmatrix} \frac{1}{d} \\ \frac{1}{d} \\ \vdots \\ \frac{1}{d} \end{bmatrix}. \tag{3.4}$$

Summarizing, a matrix D is doubly stochastic if and only if D can be thought of as a matrix of transition probabilities for a noisy channel with the uniform distribution as a stationary state.

Exercise 3.1.1: Show that the set of d by d doubly stochastic matrices is a convex set. That is, show that given doubly stochastic matrices D and E, the convex combination $pD + (1-p)E$, for $0 \leq p \leq 1$, is also doubly stochastic.

Exercise 3.1.2: Show that D is doubly stochastic if and only if the transpose D^T is doubly stochastic.

Exercise 3.1.3: Show that if D and E are doubly stochastic then the product DE is doubly stochastic.

Exercise 3.1.4: To what well-known noisy channel does the doubly stochastic matrix in (3.1) correspond?

The remainder of this section is devoted to proving two fundamental theorems linking majorization and double stochasticity.

> *Theorem 3.1.1*: D is doubly stochastic if and only if $Dr \prec r$ for all r.

Proof:

Suppose D is doubly stochastic. Then for any real t,

$$\sum_j ((Dr)_j - t)^+ = \sum_j \left(\sum_k D_{jk}(r_k - t) \right)^+ , \tag{3.5}$$

where we have applied the fact that $\sum_k D_{jk} = 1$. By the convexity of the function $(\cdot)^+$ and the double stochasticity of D, we deduce that

$$\begin{aligned} \sum_j ((Dr)_j - t)^+ &\leq \sum_{jk} D_{jk} (r_k - t)^+ & (3.6) \\ &= \sum_k (r_k - t)^+ . & (3.7) \end{aligned}$$

By the order-free characterization of majorization, Theorem 2.2.1 on page 14, it follows that $Dr \prec r$.

Conversely, suppose $Dr \prec r$ for all r. Let $e = (1, 1, \ldots, 1)$ be the vector containing all 1s. Then $De \prec e$. But $e \prec r$ for any vector r whose components sum to d. Thus $De \prec e$ and $e \prec De$, which implies that $De = e$, that is, the rows of D sum to one. Next, let e_i be the vector with a single 1 in the ith position, and zeroes elsewhere. Then $De_i \prec e_i$, which implies that all the entries of De_i must be non-negative, since by Theorem 2.1.1 on page 11 the smallest component of De_i must be at least as large as the smallest component of e_i. But De_i is the ith column of D, so we see that all the entries of D are non-negative. Finally, since $De_i \prec e_i$ the entries of De_i must have the same sum as the entries of e_i. That is, the columns of D all sum to 1, which concludes the proof that D is doubly stochastic. ■

The simplest non-trivial doubly stochastic matrix is the *T-transform*. A T-transform acts trivially on all but 2 dimensions, in which it has the form of a 2 by 2 doubly stochastic matrix,

$$T = \begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix}, \tag{3.8}$$

for some parameter t, $0 \leq t \leq 1$. As the following theorem shows, for the purposes of understanding majorization it is sufficient to consider doubly stochastic matrices which are products of T-transforms.

Theorem 3.1.2: The following statements are equivalent:

1. $r \prec s$.
2. $r = T_1 T_2 \ldots T_n s$, where n is finite and the matrices T_j are T-transforms.
3. $r = \sum_j p_j P_j s$, for some set of probabilities, p_j, and permutation matrices, P_j.
4. $r = Ds$ for some doubly stochastic matrix D.

Proof:

$1 \Rightarrow 2$: Suppose $r \prec s$. We will prove the result by induction on d, the dimension of the vector space r and s live in. For notational convenience we assume that the components of r and s have been ordered into decreasing order. The result is clear when $d = 1$, so let's assume the result is true for d, and try to prove it for $d + 1$-dimensional r and s.

Choose k such that $s_k \leq r_1 \leq s_{k-1}$. Such a k is guaranteed to exist because $r \prec s$ implies that $r_1 \leq s_1$ and $r_1 \geq r_d \geq s_d$. Choose t such that

$$r_1 = ts_1 + (1-t)s_k. \tag{3.9}$$

Now define z to be the result of applying a T-transform T with parameter t to the 1st and kth components of s, so that

$$\begin{aligned} z &= Ts & (3.10)\\ &= (ts_1 + (1-t)s_k, s_2, \ldots, s_{k-1}, (1-t)s_1 + ts_k, s_{k+1}, \ldots, s_{d+1}) & (3.11)\\ &= (r_1, s'), & (3.12) \end{aligned}$$

where

$$s' \equiv (s_2, \ldots, s_{k-1}, (1-t)s_1 + ts_k, s_{k+1}, \ldots, s_{d+1}). \tag{3.13}$$

Define $r' \equiv (r_2, r_3, \ldots, r_{d+1})$. We aim to show that $r' \prec s'$ and then apply the inductive hypothesis. Suppose $1 \leq m \leq k-2$. Then since $s_2 \geq \ldots \geq$

$s_{k-1} \geq r_1 \geq s_k \geq \ldots \geq s_{d+1}$ it follows that

$$\begin{aligned}
\sum_{j=1}^{m} r'_j &= \sum_{j=2}^{m+1} r_j && (3.14)\\
&\leq \sum_{j=2}^{m+1} s_j && (3.15)\\
&= \sum_{j=1}^{m} s'_j && (3.16)\\
&\leq \sum_{j=1}^{m} (s'_j)^{\downarrow}. && (3.17)
\end{aligned}$$

Next, consider the case when $k-1 \leq m \leq d$. Then

$$\begin{aligned}
\sum_{j=1}^{m} (s'_j)^{\downarrow} &\geq \sum_{j=1}^{m} s'_j && (3.18)\\
&= \sum_{j=2}^{k-1} s_j + [(1-t)s_1 + ts_k] + \sum_{j=k+1}^{m+1} s_j && (3.19)\\
&= \sum_{j=1}^{m+1} s_j - ts_1 + (t-1)s_k && (3.20)\\
&= \sum_{j=1}^{m+1} s_j - r_1 && (3.21)\\
&\geq \sum_{j=1}^{m+1} r_j - r_1 && (3.22)\\
&= \sum_{j=2}^{m} r_j && (3.23)\\
&= \sum_{j=1}^{m} r'_j. && (3.24)
\end{aligned}$$

Thus $r' \prec s'$. By the inductive hypothesis, $r' = T_1 \ldots T_n s'$ for some sequence of T-transforms on d dimensions. But the T-transforms can equally well be regarded as T-transforms on $d+1$ dimensions by acting trivially on the first dimension, and thus $r = T_1 \ldots T_n T s$, that is, r can be obtained from s by a finite sequence of T-transforms, as we set out to show.

$2 \Rightarrow 3$: Suppose $r = T_1 T_2 \dots T_n s$. Each T_j can be written as a convex combination of permutation matrices. Products of permutation matrices are permutation matrices. Thus $r = \sum_k p_k P_k s$ for some set of probabilities p_k and permutation matrices, P_k.

$3 \Rightarrow 4$: Suppose $r = \sum_j p_j P_j s$ for some probability distribution, p_j, and permutation matrices P_j. It is easy to check that $D \equiv \sum_j p_j P_j$ is doubly stochastic, and thus $r = Ds$ for some doubly stochastic D.

$4 \Rightarrow 1$: This has already been proved in Theorem 3.1.1.

■

Exercise 3.1.5: Find an example of vectors r and y and a doubly stochastic matrix D such that $r \prec y$ but $Dr \not\prec Dy$.

Exercise 3.1.6: Prove that not all doubly stochastic matrices are products of T-transforms.

Theorem 3.1.2 implies that $r = \sum_j p_j P_j s$ for probabilities p_j and permutation matrices P_j if and only if $r = Ds$ for doubly stochastic D. This suggests a natural representation theorem for doubly stochastic D as those matrices which can be written in the form $D = \sum_j p_j P_j$. In one direction this representation is easy to show. Namely, any matrix which can be written in the form $\sum_j p_j P_j$ is necessarily doubly stochastic. Surprisingly, the converse is more difficult to show; it does not seem to follow easily from 3.1.2, for example. However, the converse *is* true, and this representation theorem for doubly stochastic matrices is known as *Birkhoff's theorem*. Birkhoff's theorem may be stated as follows:

> *Theorem 3.1.3*: (**Birkhoff's theorem (Birkhoff 1946 [9])**)
> A d by d matrix is doubly stochastic if and only if it can be written in the form
>
> $$D = \sum_j p_j P_j \tag{3.25}$$
>
> for some set of probabilities p_j and permutation matrices P_j.

We prove Birkhoff's theorem in Appendix A. Note that the statement of the theorem in the appendix is slightly different, though equivalent, to that above. The statement in the appendix stressing the connections with

convex analysis, while the statement above is in language stripped of the terminology of convex analysis, while containing the same essential content.

For the purposes of our study of majorization, we will have little occasion to use Birkhoff's theorem, with most of our further development based on Theorem 3.1.2. Nonetheless, as Theorem 3.1.2 indicates, the concept of double stochasticity is critical in the study of majorization, and Birkhoff's theorem is therefore of interest as one of the deepest results about doubly stochastic matrices. Furthermore, the ideas used in the proof of Birkhoff's theorem are beautiful, useful, and stimulate many other interesting questions and connections, both within the theory of majorization, and in other areas of mathematics, making it worthwhile to spend time in the study of the proof given in the appendix.

Part II

Elementary applications of majorization to quantum mechanics

Chapter 4

Matrix majorization

So far we have been concerned with majorization between two vectors of real numbers, mostly vectors representing probability distributions. In quantum mechanics the object analogous to a probability distribution is the density matrix, so it is natural to consider matrix generalizations of majorization. If A and B are Hermitian matrices then we define $A \prec B$, A is majorized by B, if $\lambda(A) \prec \lambda(B)$, where $\lambda(A)$ is the vector of eigenvalues of A, arranged into non-increasing order.

The purpose of this chapter is to study the properties of matrix majorization. In the previous chapter, Theorem 3.1.2 showed that vector majorization could be characterized using doubly stochastic matrices. In Section 4.1 we will prove *Uhlmann's theorem*, which generalizes Theorem 3.1.2 to matrices, showing that matrix majorization can be characterized in terms of *quantum operations*, which are maps of matrices to matrices. Section 4.2 presents some applications of this theorem to understanding the properties of entropy under quantum dynamical quantum processes.

4.1 Characterizing matrix majorization using quantum operations

Matrix majorization is connected with the theory of *doubly stochastic quantum operations* in a way analogous to the connection between vector majorization and doubly stochastic matrices. A quantum operation is a map from density matrices to density matrices which represents the dynamical evolution in a quantum system. A review of the theory of quantum opera-

tions, and further references, is presented in Appendix B. Here we present only the most salient facts. In general, a quantum operation $\mathcal{E}$ may be written in the *operator-sum representation*

$$\mathcal{E}(\rho) = \sum_j E_j \rho E_j^\dagger, \tag{4.1}$$

where $\{E_j\}$ is a set of matrices known as *operation elements* for the quantum operation $\mathcal{E}$. The quantum operation represents a change in the state of a quantum mechanical system, with ρ being the initial state of the system, and $\mathcal{E}(\rho)$ being the final state of the system.

To represent a physical evolution, a quantum operation must satisfy the *trace-preserving* property, $\mathrm{tr}(\mathcal{E}(\rho)) = \mathrm{tr}(\rho) = 1$. With a little thought, this can be shown to be equivalent to the *completeness* relation,

$$\sum_j E_j^\dagger E_j = I. \tag{4.2}$$

A quantum operation with operation elements satisfying this property is a *trace-preserving* quantum operation.

A simple example of a quantum operation is the unitary evolution of a system according to a unitary matrix, U, $\mathcal{E}(\rho) = U\rho U^\dagger$. Note that this satisfies the completeness relation, since $U^\dagger U = I$. A less trivial example of a quantum operation is a qubit subjected to random rotations about the z axis of the Bloch sphere, according to some probability distribution $\Pr(\theta)$. This process can be described by the quantum operation

$$\rho \to \mathcal{E}(\rho) = \int d\theta \, p(\theta) R_z(\theta) \rho R_z(\theta)^\dagger, \tag{4.3}$$

where $R_z(\theta) = \exp(-iZ/2)$ is the unitary matrix describing a rotation of the Bloch sphere by an angle θ about the z axis, and Z is the Pauli sigma z matrix. Regarding the integral as a sum we can see that this process is already written in the operator-sum representation (Equation (4.1)), with operation elements $\sqrt{p(\theta)} R_z(\theta) \sqrt{d\theta}$.

There are many more examples of quantum operations. Indeed, essentially all quantum dynamical processes can be described by trace-preserving quantum operations, including complex processes such as decoherence and dissipation caused by interaction with the environment. The exception to this rule is processes where *post-selection* on measurement outcomes is allowed. For example, if we measure a qubit, then the state of the qubit

conditional on a particular measurement outcome will not be related to the initial state by a trace-preserving quantum operation. However, even in the case of measurements with post-selection, it turns out to be possible to use non trace-preserving quantum operations to describe the dynamics; see Appendix B for a description of how this is done.

Conversely, given any set of matrices satisfying the completeness relation, Equation 4.2, it turns out that there is a physical process that gives rise to that dynamical evolution. See Appendix B for details of how this is proved.

Exercise 4.1.1: A physically important example of a quantum operation is the process of *amplitude damping* on a single qubit. The amplitude damping operation has operation elements

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}; \quad E_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}, \tag{4.4}$$

where $0 \leq \gamma \leq 1$ is a parameter. Show that these operation elements satisfy the completeness relation, Equation (4.2), so amplitude damping is a trace-preserving quantum operation. Show that

$$\mathcal{E}\left(\begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}\right) = \begin{bmatrix} p+\gamma(1-p) & 0 \\ 0 & 1-p-\gamma(1-p) \end{bmatrix}. \tag{4.5}$$

The amplitude damping channel gets its name from the fact that it moves a fraction γ of the population in the state $|1\rangle$ into the state $|0\rangle$, as illustrated by the previous equation. Physically, such a process arises, for example, during spontaneous emission from a two-level atom with excited state $|1\rangle$ into the ground state $|0\rangle$.

Many physically important quantum operations $\mathcal{E}$ are *unital*, meaning that the identity matrix is a fixed point, $\mathcal{E}(I) = I$. In terms of the operation elements this may be expressed as

$$\sum_j E_j E_j^\dagger = I. \tag{4.6}$$

Dividing by the dimension, d, of state space, this condition may be expressed equivalently as $\mathcal{E}(I/d) = I/d$, so the physical significance of unitality is that the completely mixed state is a fixed point of the operation. We say a quantum operation is *doubly stochastic* if it is both trace-preserving and unital. The doubly stochastic quantum operations will be our main object of interest in this chapter.

Exercise 4.1.2: Show that a unitary quantum operation is doubly stochastic.

Exercise 4.1.3: Show that the quantum operation of Equation (4.3) is doubly stochastic.

Exercise 4.1.4: Not all physical quantum operations are doubly stochastic. Show that the quantum operation for amplitude damping is not doubly stochastic.

An interestin special class of doubly stochastic operations is the *random unitary* operations. Suppose p_j is a probability distribution and U_j are unitary matrices. Imagine that a quantum system undergoes evolution according to a unitary matrix U_j chosen at random with probability p_j. This corresponds to the quantum operation

$$\mathcal{E}(\rho) = \sum_j p_j U_j \rho U_j^\dagger, \tag{4.7}$$

with operation elements $\sqrt{p_j} U_j$. $\mathcal{E}$ is trace-preserving, since $\sum_j p_j U_j^\dagger U_j = I$, and unital since $\sum_j p_j U_j U_j^\dagger = I$. Thus random unitary operations are also doubly stochastic. We will see below that for the case of single qubits the doubly stochastic quantum operations correspond precisely to the random unitary operations, but in higher dimensions the random unitary operations form a strict subset of the doubly stochastic operations.

The following theorem characterizes majorization in terms of doubly stochastic and random unitary quantum operations. It is a matrix analogue of the result that $r \prec s$ if and only if there exists doubly stochastic D such that $r = Ds$.

Theorem 4.1.1: (**(Uhlmann's theorem)**)
Let A and B be Hermitian matrices. Then the following three conditions are equivalent:

1. $A \prec B$.
2. There exists a random unitary quantum operation $\mathcal{E}$ such that $A = \mathcal{E}(B)$.
3. There exists a doubly stcohastic quantum operation $\mathcal{E}$ such that $A = \mathcal{E}(B)$.

The connection between majorization and quantum mechanics, including this theorem, is due to Uhlmann[56, 57, 58, 59]. Some of Uhlmann's results were later generalized by Wehrl to the infinite dimensional case[64]. Many, though not, all of these results are collected in Wehrl's review paper[65].
Proof: We show that $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$. Suppose first that $A \prec B$. Let $\Lambda(M)$ denote the diagonal matrix whose entries are the eigenvalues of M, arranged into decreasing order. Because A and B are Hermitian, there exist unitary U and V such that

$$A = U\Lambda(A)U^\dagger; \quad B = V\Lambda(B)V^\dagger. \tag{4.8}$$

Since $A \prec B$ there exist permutation matrices P_j and probabilities p_j such that

$$\lambda(A) = \sum_j p_j P_j \lambda(B). \tag{4.9}$$

Straightforward matrix algebra can be used to check that this implies

$$\Lambda(A) = \sum_j p_j P_j \Lambda(B) P_j^\dagger. \tag{4.10}$$

Comparing with Equation (4.8) gives

$$\begin{aligned} A &= \sum_j p_j U P_j V^\dagger B V P_j^\dagger U^\dagger & (4.11) \\ &= \sum_j p_j U_j B U_j^\dagger, & (4.12) \end{aligned}$$

where $U_j \equiv U P_j V^\dagger$ is a product of three unitary matrices, and thus is unitary. This completes the proof that $1 \rightarrow 2$.

The proof that $2 \rightarrow 3$ is trivial, since every random unitary quantum operation is also doubly stochastic.

To prove $3 \rightarrow 1$, suppose $\mathcal{E}$ is a doubly stochastic quantum operation with operation elements $\{E_j\}$. We prove that $\mathcal{E}(B) \prec B$. Let U be the unitary matrix which diagonalizes $\mathcal{E}(B)$ and V the unitary matrix which diagonalizes B. Then $F_j \equiv U E_j V^\dagger$ defines a set of operation elements for a quantum operation $\mathcal{F}$. $\mathcal{F}$ is doubly stochastic, since $\sum_j F_j^\dagger F_j = V \sum_j E_j^\dagger E_j V^\dagger = I$ (the trace-preserving condition) and $\sum_j F_j F_j^\dagger = U E_j E_j^\dagger U^\dagger = I$ (the unitality condition). It follows that

$$\Lambda(\mathcal{E}(B)) = \mathcal{F}(\Lambda(B) = \sum_j F_j \Lambda(B) F_j^\dagger. \tag{4.13}$$

This equation may be rewritten in component form, where $F_{j,kl}$ represents the (k,l)th component of F_j,

$$\lambda(\mathcal{E}(B))_k = \sum_{jl} F_{j,kl}\lambda(B)_l F_{j,lk}^\dagger = \sum_{jl} |F_{j,kl}|^2 \lambda(B)_l. \tag{4.14}$$

Define the components of a matrix D by $D_{kl} \equiv \sum_j |F_{j,kl}|^2$, so the previous equation may be rewritten

$$\lambda(\mathcal{E}(B)) = D\lambda(B). \tag{4.15}$$

If we can show that D is doubly stochastic then it will follow that $\mathcal{E}(B) \prec B$. It is clear that the entries of D are non-negative, so all we need do is show that the row and column sums of D are 1. This follows from the trace-preserving and unitality properties of $\mathcal{F}$. For example, the unitality condition, $\sum_j F_j F_j^\dagger = I$, can be written out in component form on the diagonal to give

$$1 = \sum_{jl} F_{j,kl} F_{j,lk}^\dagger = \sum_l D_{kl}. \tag{4.16}$$

That is, the row sums of D are 1. Similarly the trace-preserving condition, $\sum_j F_j^\dagger F_j = I$, can be written out in component form on the diagonal to give

$$1 = \sum_{jl} F_{j,kl}^\dagger F_{j,lk} = \sum_l D_{lk}. \tag{4.17}$$

That is, the column sums of D are 1, which completes the proof. ■

Exercise 4.1.5: Show that if $\mathcal{E}$ is a trace-preserving quantum operation that is not doubly stochastic, then there exists Hermitian A such that $\mathcal{E}(A) \not\prec A$.

4.2 Which processes increase quantum entropy?

The von Neumann entropy is a measure of the disorder present in a quantum state, ρ. It is defined by $S(\rho) \equiv -\mathrm{tr}(\rho \log(\rho))$, where we take the logarithm to base 2. An introduction to the properties of the von Neumann entropy may be found in [43, 65, 44]. Suffice to say that the von Neumann entropy

is a quantity of great importance in both quantum statistical mechanics and quantum information science. We assume that the reader is already reasonably familiar with the properties of the von Neumann entropy, and motivated to consider the study of the von Neumann entropy an interesting topic. If this is not true, then the reader may safely skip to Section 6.0.1.

Majorization and the von Neumann entropy are similar, in that both offer approaches to the problem of quantifying what it means for one quantum state to be more disordered than another. We alluded to connections between majorization and entropy in Chapter 1, and will now work those connections out in more detail.

> *Proposition 4.2.1*: Suppose ρ and σ are density matrices such that $\rho \prec \sigma$. Then $S(\rho) \geq \sigma$.

We will give two proofs of this proposition, one in the main text, and the other in the exercises. The proof in the text relies on a well-known but somewhat nontrivial fact, the concavity of the von Neumann entropy, $S(\sum_k p_k \rho_k) \geq \sum_k p_k S(\rho_k)$. See any of [43, 65, 44] for a proof of this fact. A more elementary proof may be found below in Exercises 4.2, 4.2, and 4.2.
Proof: Suppose $\rho \prec \sigma$. By Theorem 4.1.1 it follows that there exist probabilities p_j and unitaries U_j such that $\rho = \sum_j p_j U_j \sigma U_j^\dagger$. From the concavity of the von Neumann entropy it follows that $S(\rho) \geq \sum_j p_j S(U_j \sigma U_j^\dagger)$. But $S(U_j \sigma U_j^\dagger) = S(\sigma)$, so $S(\rho) \geq S(\sigma)$, as required. ■

Exercise 4.2.1: The Shannon entropy of a probability distribution $\{p_j\}$ is defined by $H(\{p_j\}) \equiv -\sum_j p_j \log(p_j)$, where the logarithm is taken to base two. Show that the Shannon entropy is a concave function of the probability distribution.

Exercise 4.2.2: Show that the Shannon entropy and von Neumann entropy are related by the equation $S(\rho) = H(\lambda(\rho))$.

Exercise 4.2.3: Suppose $\rho \prec \sigma$. Show that there exist probabilities p_j and permutation matrices P_j such that $\lambda(\rho) = \sum_j p_j P_j \lambda(\sigma)$. Use the results of the previous two exercises to provide an alternative proof of Proposition 4.2.

An important physical question is to determine which trace-preserving quantum operations only ever increase, or, more precisely, never decrease,

the von Neumann entropy. *A priori* it is not obvious what the answer to this question is, but with the tools now at our disposal it is easy to prove that it is precisely the doubly stochastic quantum operations which never decrease the entropy.

> *Theorem 4.2.2*: Let $\mathcal{E}$ be a trace-preserving quantum operation acting on a d-dimensional state space. Then one of the following two possibilities holds:
>
> - $\mathcal{E}$ is doubly stochastic, in which case, $S(\mathcal{E}(\rho)) \geq S(\rho)$ for all density matrices ρ.
> - $\mathcal{E}$ is not doubly stochastic, in which case, there exists a density matrix ρ such that $S(\mathcal{E}(\rho)) < S(\rho)$.

Proof: Suppose $\mathcal{E}$ is doubly stochastic. Then it follows from Theorem 4.1.1 that $\mathcal{E}(\rho) \prec \rho$, and thus by Proposition 4.2, $S(\mathcal{E}(\rho)) \geq S(\rho)$.

Suppose $\mathcal{E}$ is not doubly stochastic. Since $\mathcal{E}$ is trace-preserving, by assumption, it follows that $\mathcal{E}(I) \neq I$, and thus $\mathcal{E}(I/d) \neq I/d$, that is, $\mathcal{E}(I/d)$ is not a maximally mixed state, and thus $S(\mathcal{E}(I/d)) < S(I/d) = \log(d)$. ■

Problems for Lecture 4

Problem 4.2.1: The following result was used to study the problem of simulating one Hamiltonian by another set of Hamiltonians, in the context of universality in quantum computation [42]. Let A and B be traceless Hermitian matrices, and assume that $B \neq 0$. Prove that there exists a positive constant c such that $A \prec cB$, and thus there exist probabilities p_j and unitary matrices U_j such that

$$A = c \sum_j p_j U_j B U_j^\dagger. \tag{4.18}$$

Hints for Lecture 4

Hint for Exercise 1.3: The basis $|e_k\rangle$ to measure in is the Fourier transform of the eigenbasis of ρ.

Chapter 5

Decomposing density matrices into pure states

Uhlmann's theorem provides a fundamental connection between majorization and quantum mechanics. In this chapter we apply Uhlmann's theorem to the problem of characterizing how a density matrix ρ can be decomposed into ensembles of pure states, $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. The solution of this problem will allow us to easily prove another important theorem connecting majorization to quantum mechanics, known as *Horn's lemma*. Horn's lemma states that $r \prec s$ if and only if there exists a unitary matrix u such that $r_j = \sum_k |u_{jk}|^2 s_k$. It arises widely in quantum mechanics, owing to the ubiquity of unitary matrices in that theory.

Section 5.1 considers the characterization of ensemble decompositions for a density matrix ρ, $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. Section 5.2 applies these results to prove Horn's lemma, and provides some simple applications of Horn's lemma to quantum mechanics; the next chapter contains more complex — and interesting — applications of both Horn's lemma and Uhlmann's theorem to a problem of great fundamental interest in quantum mechanics: characterizing the acquisition of information during a quantum measurement. Finally, in Section 5.3 we revisit the problem of characterizing the ensemble decomposition of a density matrix

5.1 What probabilities can appear in an ensemble for a density matrix?

The density matrix is an important tool used in quantum mechanics to deal with situations where we have incomplete knowledge of a quantum state. Recall that if a quantum system is in state $|\psi_j\rangle$ with probability p_j, then the density matrix describing that situation is defined to be

$$\rho \equiv \sum_j p_j |\psi_j\rangle\langle\psi_j|. \tag{5.1}$$

As discussed in detail in Appendix C, a given density matrix, ρ, may be given many different decompositions into pure state ensembles,

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j = \sum_k q_k |\phi_k\rangle\langle\phi_k|. \tag{5.2}$$

The reader not thorougly comfortable with this fact should pause to review the material in Appendix C before proceeding. As proved in the Appendix, the following theorem characterizes, for a given density matrix, the different decompositions into pure state ensembles possible for a given density matrix.

> *Theorem 5.1.1*: (**(Ensemble theorem)**)
> Let ρ be a rank-l density matrix, with $\rho = \sum_{j=1}^{l} \lambda_j(\rho)|j\rangle\langle j|$, where $\lambda_j(\rho)$ are the nonzero eigenvalues of ρ, and $|j\rangle$ is a corresponding set of orthonormal eigenvectors. Then a set of probabilities p_j $(j = 1, \ldots, m)$ and corresponding normalized state vectors $|\psi_j\rangle$ generate ρ if and only if there exists an $l \times m$ matrix u with orthonormal rows, and such that
>
> $$\sqrt{p_k}|\psi_k\rangle = \sum_j u_{jk}\sqrt{\lambda_j(\rho)}|j\rangle. \tag{5.3}$$

A discussion of the history and proof of this theorem may be found in Appendix C.

Theorem 5.1.1 characterizes the possible ensemble decompositions of a density matrix. It suggests an interesting related problem: for a given density matrix, ρ, for what class of probabilitiy distributions (r_j) does there exist a set of pure states $|\psi_j\rangle$ such that $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$? We will refer to

this problem as the *ensemble probability problem*. The solution to the ensemble probability problem involves majorization, and has many interesting implications for quantum mechanics.

The approach we take to the ensemble probability problem is based on two results interesting in their own right, the Schmidt decomposition and Schur's lemma, which we now review. The *Schmidt decomposition* for a pure state, $|\psi\rangle$, of a composite quantum system with components A and B, is a decomposition of the form

$$|\psi\rangle = \sum_j \lambda_j |j\rangle_A |j\rangle_B, \tag{5.4}$$

where $\lambda_j \geq 0$ are real coefficients, and $|j\rangle_A$ and $|j\rangle_B$ are orthonormal bases for system A and B, respectively. Such a Schmidt decomposition can be shown to exist for any pure state of a two-component quantum system. The proof is a simple application of a powerful result from linear algebra known as the *singular value decomposition*; see, for example, [43, 45] for proofs of the Schmidt decomposition.

A surprising consequence of the Schmidt decomposition is that the eigenvalues of the reduced density matrices $\rho_A \equiv \mathrm{tr}_B(|\psi\rangle\langle\psi|), \rho_B \equiv \mathrm{tr}_A(|\psi\rangle\langle\psi|)$ are closely related. From Equation 5.4 we see that

$$\rho_A = \sum_j \lambda_j^2 |j\rangle_A \langle j|_A; \quad \rho_B = \sum_j \lambda_j^2 |j\rangle_B \langle j|_B, \tag{5.5}$$

so the nonzero eigenvalues are in fact equal. When system A and system B have the same dimension, this implies that $\lambda(\rho_A) = \lambda(\rho_B)$. When the dimensions of A and B are not equal, the non-zero components of these vectors will be the same, but the dimensionality of the two vectors will be different. However, even in this case we will write $\lambda(\rho_A) = \lambda(\rho_B)$, where it is understood that equality means that one should "pad" whichever vector is of lower dimensionality with extra zero entries so that the two vectors have the same dimensionality.

The other result we need for the solution of the ensemble probability problem is *Schur's theorem*, which relates the eigenvalues of a matrix to the diagonal entries of that matrix:

> *Theorem 5.1.2*: (**Schur's theorem**)
> Let A be a d by d Hermitian matrix. Let $\mathrm{diag}(A)$ denote the vector whose components are the diagonal entries of A with respect

to some orthonormal basis $|j\rangle$. Then

$$\mathrm{diag}(A) \prec \lambda(A). \tag{5.6}$$

Schur's theorem is a special case of a more general result, anticipated in Theorem 1.3.2 on page 7. We state a slight generalization of that result here:

> *Theorem 5.1.3*: Let A be a Hermitian matrix, and P_j a complete set of orthonormal projectors, that is, $\sum_j P_j = I$. Then $A' = \sum_j P_j A P_j$ satisfies $A' \prec A$.

The theorem follows immediately from Uhlmann's theorem and the observation that $\mathcal{E}(\rho) = \sum_j P_j \rho P_j$ is a doubly stochastic operation, that is, $\mathcal{E}$ is both trace-preserving and $\mathcal{E}(I) = I$. To prove Schur's theorem from theorem 5.1.3 we simply choose the projectors $P_j = |j\rangle\langle j|$.

We are now in position to make our first significant progress on the ensemble probability problem, proving a necessary condition for a density matrix ρ to be decomposable as $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. For later purposes it will be convenient to state the result in a slightly more general, though equivalent, form, as a result about arbitrary positive matrices, not just density matrices:

> *Proposition 5.1.4*: Suppose M is a positive matrix, $r = (r_j)$ is a vector of real non-negative numbers, and $|\psi_j\rangle$ is a corresponding set of normalized state vectors, such that $M = \sum_j r_j |\psi_j\langle\rangle\psi_j|$. Then $r \prec \lambda(M)$. Note that if the vector r is of a different dimension to $\lambda(M)$ then we "pad" whichever vector has the smaller dimension with additional zeroes to enable comparison using the majorization relation.

Proof: We will assume that $M = \rho$ is a density matrix, and $(r_j) = (p_j)$ is a probability distribution. The general result follows easily from the observation that $M/\mathrm{tr}(M)$ is a density matrix for any non-zero positive matrix M. We give the system in which ρ lives a name, system A. Furthermore, we introduce an auxiliary quantum system, B. By definition, B's state space is spanned by a set of orthonormal vectors $|j\rangle_B$ whose index set j runs over the same set of values as the probabilities p_j in the probability distribution. The system B is a ficticious system, yet it plays a role in siimplifying the proof of the proposition; you might think of the role of B as being like the

role complex numbers play in providing simple proofs of some proofs in real analysis. We define a pure state, $|\psi\rangle$, of the system AB by

$$|\psi\rangle \equiv \sum_j \sqrt{p_j}|\psi_j\rangle|j\rangle. \tag{5.7}$$

Let ρ_A and ρ_B be the corresponding reduced density matrices. Direct calculation shows that $rho_A = \rho$, while

$$\rho_B = \sum_{jk} \sqrt{p_j p_k}|j\rangle\langle k| \langle\psi_k|\psi_j\rangle, \tag{5.8}$$

so $\mathrm{diag}(\rho_B) = (p_j)$. Thus, $(p_j) = \mathrm{diag}(\rho_B) \prec \lambda(\rho_B)$, where we have applied Schur's lemma. But, as discussed above in connection with the Schmidt decomposition, $\lambda(\rho_B) = \lambda(\rho_A)$, so we deduce that $(p_j) \prec \lambda(\rho_A) = \lambda(\rho)$, which completes the proof. ■

Exercise 5.1.1: This exercise provides a more direct proof of Proposition 5.1, avoiding the use of Schur's theorem or auxiliary systems. Use the ensemble theorem, Theorem 5.1.1, to argue that $p_k = \sum_j |u_{jk}|^2 \lambda_j(\rho)$ for some $l \times m$ matrix u with orthonormal rows, where l is the rank of ρ, and m is the number of elements in the probability distribution. Then use Theorem 3.1.2 to aruge that $(p_j) \prec \lambda(\rho)$.

It turns out that the necessary condition found in Proposition 5.1 is actually sufficient as well. Proving this turns out to be an easy consequence of the rank-two case.

> *Proposition 5.1.5*: Let $r \prec s$ be two-dimensional vectors with non-negative entries, and suppose $|e_1\rangle, |e_2\rangle$ are orthonormal vectors. Then there exist normalized state vectors $|\phi_1\rangle, |\phi_2\rangle$ such that
>
> $$r_1|\phi_1\rangle\langle\phi_1| + r_2|\phi_2\rangle\langle\phi_2| = s_1|e_1\rangle\langle e_1| + s_2|e_2\rangle\langle e_2|. \tag{5.9}$$

Proof: It will be convenient to assume that $r_1, r_2 \neq 0$; we can assume this without loss of generality, since the cases when one or both are zero are trivial. Because $r \prec s$ it follows from Theorem 3.1.2 that there exists t satisfying $0 \leq t \leq 1$ and such that

$$r_1 = ts_1 + (1-t)s_2; \quad r_2 = (1-t)s_1 + ts_2. \tag{5.10}$$

We define $\theta \equiv \arccos(\sqrt{t})$ and define states $|\phi_1\rangle$ and $|\phi_2\rangle$ by

$$\sqrt{r_1}|\phi_1\rangle \equiv \cos(\theta)\sqrt{s_1}|e_1\rangle + \sin(\theta)\sqrt{s_2}|e_2\rangle, \quad (5.11)$$
$$\sqrt{r_2}|\phi_2\rangle \equiv -\sin(\theta)\sqrt{s_1}|e_1\rangle + \cos(\theta)\sqrt{s_2}|e_2\rangle. \quad (5.12)$$

The fact that $|\phi_1\rangle$ is normalized follows by taking the inner product of Equation (5.11) with itself to give

$$r_1\langle\phi_1|\phi_1\rangle = \cos^2(\theta)s_1 + \sin^2(\theta)s_2. \quad (5.13)$$

By definition of θ we have $\cos^2(\theta) = t$ and $\sin^2(\theta) = 1 - t$, so comparing with Equation (5.10) we see that $|\phi_1\rangle$ is normalized. A similar proof shows that $|\phi_2\rangle$ is normalized. Equation (5.9) can be verified either by direct calculation, or by using Theorem 5.1.1. ■

By combining Propositions 5.1 and 5.1, we can obtain a statement about the rank-two case that is stronger than Proposition 5.1.

> *Proposition 5.1.6*: Let $r \prec s$ be two-dimensional vectors with non-negative entries, and suppose $|psi_1\rangle, |psi_2\rangle$ are normalized state vectors. Then there exist normalized state vectors $|\phi_1\rangle, |\phi_2\rangle$ such that
>
> $$r_1|\psi_1\langle\rangle\psi_1| + r_2|\psi_2\langle\rangle\psi_2| = s_1|\phi_1\langle\rangle\phi_1| + s_2|phi_2\langle\rangle\phi_2|. \quad (5.14)$$

Proof: Let $M \equiv s_1|\phi_1\langle\rangle\phi_1| + s_2|phi_2\langle\rangle\phi_2|$. By Proposition 5.1 $s \prec \lambda(M)$. Since the majorization relation is transitive it follows also that $r \prec \lambda(M)$, and the result follows from Proposition 5.1. ■

We're in position to put all the elements together to solve the ensemble probability problem.

> *Theorem 5.1.7*: Let ρ be a density matrix, and $p = (p_j)$ a probability distribution. Then a set of normalized state vectors $|\psi_j\rangle$ such that $\rho = \sum_j p_j|\psi_j\rangle\langle\psi_j|$ exists if and only if $p \prec \lambda(\rho)$. As earlier, if p and $\lambda(\rho)$ have different dimensions then we may pad whichever vector is smaller with extra zeroes to enable comparison using the majorization relation.

Proof: The necessity of the condition $p \prec \lambda(\rho)$ was already proved in Proposition 5.1. To prove sufficiency, we use Theorem 3.1.2 to write $p =$

$T_1 T_2 \ldots T_n \lambda(\rho)$, where n is finite, and the T_j are T-transforms. The result now follows by combining this fact with repeated applications of Proposition 5.1. ■

It is worth observing that the statement of Theorem 5.1.7 can actually be strengthened slightly. If we look at the constructive part of the proof, we see that the ensemble of states $|\psi_j\rangle$ can be formed by taking *real* linear combinations of the eigenstates of ρ. This observation will be useful below in our discussion of Horn's lemma.

The history of Theorem 5.1.7 is interesting. The result was conjectured by Uhlmann [56], who proved that $p \prec \lambda(\rho)$. The proof in the reverse direction was completed by Nielsen [41]; Ruskai had noted the same result prior to the paper of Nielsen, but had not published. Many elements of the proof are implicit in Hughston, Jozsa and Wootters [27], but they do not draw the connection with majorization. In the context of the present proof, it is interesting that the proofs of Nielsen and Ruskai **(CHECK!)** make use of Horn's lemma, which we have not used. The idea for the present line of development was suggested to the author in a personal communication by Kitaev. What is nice about this line of thought is that we can easily *deduce* Horn's lemma as a consequence of Theorem 5.1.7, reversing the order of earlier presentations. This offers a substantial advantage over earlier treatments of Horn's lemma, which tended to be somewhat complex and notationally difficult. By contrast, the present development is simple and enlightening.

Exercise 5.1.2:

Exercise 5.1.3: Let ρ be any rank d density matrix, and suppose $m \geq d$. Show that there exist pure states $|\psi_1\rangle, \ldots, |\psi_m\rangle$ such that ρ is an equal mixture of these states with probability $1/m$,

$$\rho = \sum_k \frac{|\psi_k\rangle\langle\psi_k|}{m}. \tag{5.15}$$

(From [41].)

5.2 Horn's lemma

Theorem 5.2.1: (**(Horn's lemma)**)
Let r and s be d-dimensional real vectors. Suppose $r \prec s$.

Then there exists a $d \times d$ real unitary matrix[1] u such that $r_j = \sum_k |u_{jk}|^2 s_k$. Conversely, for any $d \times d$ unitary matrix, u, real or not, $r_j = \sum_k |u_{jk}|^2 s_k$ implies $r \prec s$.

Note that Horn's lemma immediately implies the weaker statement that $r \prec s$ if and only if $r_j = \sum_k |u_{jk}|^2 s_k$ for some unitary, u. Most often it is this form of Horn's lemma that is useful, and often when speaking of "Horn's lemma" it is this result to which we are referring.

There is a compact way of restating Horn's lemma that is sometimes useful. The *Hadamard product* $A \circ B$ of two $m \times n$ matrices A and B is defined to be just the elementwise product: $(A \circ B)_{jk} \equiv A_{jk} B_{jk}$. Horn's lemma may thus be stated as $r \prec s$ if and only if there exists orthogonal u such that $r = (U \circ U^*)s$. This has the substantial advantage of being notationally simpler than the statement of Horn's lemma above, and has the additional advantage of making explicit the connection between Horn's lemma and the well-studied Hadamard product — see, for example, Chapter 5 of [22] for a survey of the properties of the Hadamard product.

Proof: Suppose $r \prec s$. We suppose initially that both r and s have non-negative entries summing to one. That is, r and s can be thought of as probability distributions. This restriction will be lifted below. Let σ be a density matrix in d dimensions, chosen so that $\lambda(\sigma) = s$. Then by Theorem 5.1.7, we can find a set of states $|\psi_j\rangle$ such that $\sigma = \sum_j r_j |\psi_j\rangle\langle\psi_j|$. By Theorem 5.1.1 and the comments made after the proof of Theorem 5.1.7, we can write

$$\sqrt{r_j}|\psi_j\rangle = \sum_k u_{jk} \sqrt{s_k} |k\rangle, \tag{5.16}$$

where u is a real unitary matrix, and $|k\rangle$ are an orthonormal set of eigenstates for σ. Taking the inner product of Equation (5.16) with itself, we have $r_j = \sum_k |u_{jk}|^2 s_k$, as required.

To prove the converse, just note that the matrix D with entries defined by $D_{jk} \equiv |u_{jk}|^2$ is doubly stochastic, since the rows and columns of u must all be normalized. By Theorem 3.1.2 it follows that $r \prec s$. ■

A doubly stochastic matrix D whose entries can be written $D_{jk} = |u_{jk}|^2$ for unitary u is called *unistochastic*. D is said to be *orthostochastic* if, in addition, U is real. Horn's lemma shows that $r \prec s$ if and only if $r = Ds$ for orthostochastic (or unistochastic) D.

[1] Such matrices are usually referred to as *orthogonal* matrices.

It is not difficult to show that all 2×2 doubly stochastic matrices are also unistochastic, but in higher dimensions this is not true. To see this, suppose we have a unistochastic matrix D, $D_{jk} = |u_{jk}|^2$. Writing out in component form the condition that the first two columns of u are orthonormal to one another we obtain $\sum_j u_{j1}^* u_{j2} = 0$, and thus

$$u_{11}^* u_{12} = -\sum_{j \neq 1} u_{j1}^* u_{j2}. \tag{5.17}$$

Taking the absolute value of both sides and using the triangle inequality we obtain

$$\sqrt{d_{11} d_{12}} \leq \sum_{j \neq 1} \sqrt{D_{j1} D_{j2}}. \tag{5.18}$$

This condition is violated for the 3×3 doubly stochastic matrix

$$D = \begin{bmatrix} 0.7 & 0.3 & 0 \\ 0.3 & 0.2 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix}, \tag{5.19}$$

and thus D cannot be unistochastic.

Exercise 5.2.1: Show that all 2 by 2 doubly stochastic matrices are unitary-stochastic.

5.3 Decomposing the density matrix revisited

This section will be filled in later.

Problems for Lecture 5

Problem 5.3.1: (**(From [41].)**) Suppose $|\psi\rangle$ is a pure state of a composite system AB with Schmidt decomposition

$$|\psi\rangle = \sum_j \sqrt{p_j} |j_A\rangle |j_B\rangle. \tag{5.20}$$

Prove that given a probability distribution q_j there exists an orthonormal basis $|j'_A\rangle$ for system A and corresponding normalized states $|\psi_j\rangle$ of system B such that

$$|\psi\rangle = \sum_j \sqrt{q_j}|j'_A\rangle|\psi_j\rangle \tag{5.21}$$

if and only if $(q_i) \prec (p_i)$.

Hints for Lecture 5

Hint for Exercise 5.1 Set $|\psi_k\rangle \equiv \sum_{j=1}^d \omega^{jk}|j\rangle$, where $|k\rangle$ are orthonormal eigenstates for ρ, and $\omega \equiv \exp(2\pi i/m)$.

Chapter 6

Information acquisition during quantum measurements

In this chapter we apply the theory of majorization to the analysis of the quantum measurement process. Majorization turns out to be remarkably well suited to the analysis of quantum measurements, and we will prove results capturing the intuition that quantum measurements acquire information about the state being measured. Later, in Chapter **??**, we will use these results to analyse the process of transforming one entangled state to another.

The results of the chapter are framed in terms of the so-called *generalized measurement* formalism. Generalized measurements extend the standard von Neumann formalism for quantum measurement taught in most undergraduate quantum mechanics class. They do this by imagining a situation in which one quantum system (the "system being measured") is allowed to unitarily interact with another quantum system (the "measuring device"), and the result of the measurement is then read out by applying a von Neumann measurement to the measuring device. The generalized measurement formalism achieves its simplicity and utitility by describing the effect of this operation on the system being measured alone, not the effect on the combined systems. Thus, the generalized measurement formalism provides an extremely general approach to quantum measurement, describing with ease many situations that are awkward (though not impossible) to describe using the better-known von Neumann approach. This ease of use has resulted in the generalized measurement formalism becoming quite widely used in quantum information theory. A review of the generalized measurement formalism may be found in Appendix B.

The chapter is structured as follows.

Section **??** contains proofs of the static constraints (6.7) and (6.8) on the mixing of quantum states, and the dynamic constraints (6.9) and (6.10) on quantum measurement, and explores some elementary consequences of these results. In Section 6.3 we prove the partial converses to (6.7)-(6.8) and (6.9)-(6.10). Section 10.1 explains how the results of the present paper may be used to obtain simplified proofs of known results about entanglement transformation. Finally, Section 6.4 concludes the paper with a discussion of some open problems and future directions.

6.0.1 Quantum measurement without post-selection

As an application, let $f(\cdot)$ be any Schur-convex function. There is a natural function on Hermitian matrices induced by f, namely $f(A) \equiv f(\lambda_A)$, where A is a Hermitian matrix, and λ_A is its vector of eigenvalues. Clearly $f(\mathcal{E}(A)) \leq f(A)$ for all Hermitian A if and only if $\mathcal{E}$ is a doubly stochastic quantum operation. Thus, for example $\mathrm{tr}(\mathcal{E}(\rho)^2) \leq \mathrm{tr}(\rho^2)$ for all density matrices ρ, and $S(\mathcal{E}(\rho)) \geq S(\rho)$.

Schur's theorem has a beautiful corollary, the *Hadamard determinant theorem*:

> *Theorem 6.0.1*: (**Hadamard determinant theorem**)
> Suppose A is a d by d positive matrix. Then
>
> $$\det(A) \leq \prod_{i=1}^{d} A_{ii}. \tag{6.1}$$

Proof:

We know that $\mathrm{diag}(A) \prec \lambda_A$. By the Schur-concavity of the product function $f(x) \equiv \prod_{i=1}^{d} x_i$ we have

$$\prod_{i=1}^{d} \lambda(A)_i \leq \prod_{i=1}^{d} A_{ii}, \tag{6.2}$$

which gives

$$\det(A) \leq \prod_{i=1}^{d} A_{ii}. \tag{6.3}$$

■

Schur's theorem is equivalent to another extremely useful result, *Ky Fan's maximum principle*:

> *Theorem 6.0.2*: (**Ky Fan's maximum principle**)
> Let A be any d by d Hermitian matrix. Then for any k in the range 1 through d,
>
> $$\sum_{j=1}^{k} \lambda(A)_j^{\downarrow} = \max_P \operatorname{tr}(AP), \tag{6.4}$$
>
> where the maximum is over all k dimensional projectors P.

Proof: Let P be a projector onto a subspace with orthonormal basis $|1\rangle, |2\rangle, \ldots, |k\rangle$. Writing A with respect to this basis, Schur's theorem implies that

$$\operatorname{tr}(PA) = \sum_{i=1}^{k} A_{ii} \leq \sum_{i=1}^{k} \lambda_i^{\downarrow}(A). \tag{6.5}$$

■

Exercise 6.0.1: Use Ky Fan's maximum principle to prove Schur's theorem.

> *Theorem 6.0.3*: (**Minkowski determinant theorem**)
> If A and B are d by d positive matrices then
>
> $$(\det(A+B))^{1/d} \leq (\det A)^{1/d} + (\det B)^{1/d}. \tag{6.6}$$

6.1 Constraints on the mixing of quantum states

Suppose we mix a set of quantum states ρ_i according to the probability distribution p_i. Then we will show that this mixing process must satisfy the constraint equations:

$$\lambda\left(\sum_i p_i \rho_i\right) \prec \sum_i p_i \lambda(\rho_i) \tag{6.7}$$

$$\bigoplus_i p_i \lambda(\rho_i) \prec \lambda\left(\sum_i p_i \rho_i\right). \tag{6.8}$$

A formal definition of majorization appears in Subsection **??**, however for now the essential intuition to grasp is that the relation $x \prec y$ means that the vector x is more "mixed" (or "disordered") than y. Thus, Equation (6.7) captures the intuition that $\sum_i p_i \rho_i$ is more mixed, on average, than the states ρ_i appearing in the ensemble. The intuition behind (6.8) is a little more complex. Imagine that we prepare the state ρ by randomly choosing a value for i according to the probability distribution p_i, and then preparing the corresponding state ρ_i. Our quantum state, including a description of i, may be written as $\sum_i p_i |i\rangle\langle i| \otimes \rho_i$. We then "throw away" the state $|i\rangle$ representing our random choice of i, leaving only the state $\sum_i p_i \rho_i$. The relation (6.8) expresses the fact that when we throw away i, the state of the quantum system becomes less disordered.

Suppose we perform a measurement on a quantum mechanical system initially in the state ρ, obtaining measurement result i with probability p_i, and corresponding posterior state ρ'_i. What constraints are placed on the relationship between ρ, p_i and ρ'_i? We will show that the following two *dynamic constraints* must be satisifed:

$$\lambda(\rho) \prec \sum_i p_i \lambda(\rho'_i) \tag{6.9}$$

$$\bigoplus_i p_i \lambda(\rho'_i) \prec \lambda(\rho). \tag{6.10}$$

The intuition behind (6.9) is that quantum measurements acquire information about the state of the system being measured, and thus after measurement the state of the system is less mixed, on average, than before. The intuition behind (6.10) is a little more complex, but can be understood using Zurek's approach[71] to decoherence and quantum measurement. Recall that in this approach a measurement involves three systems: the system being measured, which starts in the state ρ, and ends in the state ρ'_i; a measuring device, which starts in some standard state, and finishes in a "pointer state" $|i\rangle$ recording the result of the measurement, and an environment which "decoheres" the measuring device, ensuring that it behaves in an essentially classical fashion. The system and measuring device interact unitarily during the measurement, ensuring that there is no change in the amount of disorder present in the system. The subsequent environmental decoherence process can also be thought of as a type of measurement, in which the different outcomes are averaged over. In this view, the environment continually measures the state of the measuring apparatus, resulting in a final state $\sum_i p_i |i\rangle\langle i| \otimes \rho'_i$

for the measuring apparatus and system being measured. This decoherence process causes an increase in the disorder present in the system, which is the intuition behind (6.10). More succinctly, (6.10) may be thought of as capturing the notion that the total ensemble of possible quantum states is more disordered after a measurement than it is before.

The importance of the static constraints (6.7)-(6.8) and the dynamics constraints (6.9)-(6.10) is further reinforced by the fact that in each case there is a type of converse to these equations. In this introduction we focus only on the more interesting case of the converse to the dynamic constraints (6.9) and (6.10), however rather similar remarks hold also for the static constraints (6.7) and (6.8). Suppose p_i is a probability distribution, and ρ and ρ'_i are quantum states such that

$$\lambda(\rho) \quad \prec \quad \sum_i p_i \lambda\left(\rho'_i\right). \tag{6.11}$$

Then we will show that there exists a quantum measurement whose measurement outcomes may be labelled by a *pair* of indices (i, j), such that for any fixed i and for all j the posterior state of the quantum system after measurement is ρ'_i, and the probabilities p_{ij} for the (i, j)th measurement outcome satisfy $\sum_j p_{ij} = p_i$. Unfortunately, this result is not a tight converse to equations (6.9) and (6.10), due to the introduction of the extra index j, however for many purposes it is a sufficiently strong converse. We will show that even the equations (6.9) and (6.10) together do not completely characterize the quantum measurement process, however I believe it likely that there is a simple characterization of the measurement process along similar lines that may be expressed entirely in terms of the eigenvalues of the prior and posterior states, and the probabilities of the different measurement outcomes. Of course, it is true that the quantum measurement formalism already provides such a characterization, in the form of a matrix equation, however equations such as (6.9) and (6.10) provide far more explicit information, and as such, are likely to be more useful in practice. We will demonstrate the utility of this approach by application to the problem of entanglement transformation, simplifying the proofs of several known results about entanglement transformation[40, 61, 62, 31, 19].

There is a striking level of symmetry in the equations (6.7)-(6.8), (6.9)-(6.10), which we will also see in the partial converse results. It is obviously tempting to suggest that this reflects some deeper underlying principle, much as Maxwell's equations may be derived from a deeper action principle based

on the Faraday tensor, or the still deeper principles of gauge invariance and relativity. Unfortunately, I have not yet succeeding in obtaining a satisfactory form for such a deeper principle. Presumably, such a deeper principle might assist in tightening the partial converse results, or perhaps tightening the partial converses may shed light on the origin of Equations (6.7)-(6.8), (6.9)-(6.10).

In explaining the intuitive meanings of the equations (6.7)-(6.8) and (6.9)-(6.10) we have used language such as the "disorder" present in a quantum state. One might wonder if it is possible to write down entropic statements capturing these intuitions. We will show that each of these equations in fact implies an entropic statement whose content corresponds to the intuition we have described. Of course, entropic statements should really only be interpreted in the aymptotic limit where we have a large number of identical copies of a system available; the advantage of Equations (6.7)-(6.8) and (6.9)-(6.10) is that they are stronger forms of these asymptotic statements which may be applied to single quantum systems.

This paper contains six fundamental results (together with a number of applications), expressed in the four constraint equations, (6.7)-(6.8), (6.9)-(6.10), and the partial converses to (6.7)-(6.8) and (6.9)-(6.10). We now review antecedents of these results in the existing literature. Equation (6.7) is an elementary consequence of classic results in the theory of majorization. Equation (6.8) follows as a corollary of work of Uhlmann[56], Ruskai (unpublished, 1993) and Nielsen[41] on the relationship between mixed states and probability distributions. Equations (6.9) and (6.10) are implicit in the work of Vidal[62] on entanglement transformation, and the partial converse to (6.9)-(6.10) is implicit in the work of Jonathan and Plenio[31] on entanglement transformation, building on earlier work by Nielsen[40]. A proof of Equation (6.9) in the context of entanglement transformation has also been previously obtained by Jonathan, Nielsen, Schumacher and Vidal (unpublished, 1999). There are several advantages to the point of view taken in the present paper. First, measurement is in some sense a more fundamental process than entanglement transformation, and Equations (6.9) and (6.10) highlight the fundamental connection between measurement and majorization for the first time, incidentally explaining why there is a connection between entanglement transformation and majorization: it arises as a result of a deeper connection between measurement and majorization. Second, the proofs in the present paper are novel, and have the advantage of proceeding from a more unified point of view than earlier work. As a result they

are, perhaps, more elegant and informative than earlier proofs, especially the proof of the partial converse to (6.9)-(6.10), which is a substantial improvement of and extension to existing constructions. Several other items of related work are also worth pointing out. There is a substantial mathematical literature on the problem of characterizing the properties of sums $A + B$ of Hermitian matrices A and B, and Fulton[17] has written a nice review of recent progress on this problem, which is closely related to the problem of mixing of density matrices. Hardy[19] has introduced techniques in the context of entanglement transformation that can be used to prove (6.9) and the partial converse to (6.9)-(6.10). Fuchs and Jacobs (unpublished, 2000) have obtained a beautiful and quite different proof of (6.9), after hearing of the result from Nielsen. Finally, the procedure described in this paper to prove the partial converse to (6.9)-(6.10) is a generalization of the procedures for entanglement transformation for pure states found by Nielsen in [40], and subsequently improved in independent work by Hardy, Jonathan and Nielsen (described in Chapter 12 of [43]), by Jensen and Schack[29], and by Werner (unpublished, 2000).

6.2 Constraints on the mixing of quantum states

Theorem 6.2.1: Suppose $\rho = \sum_j p_j \rho_j$ is a convex combination of quantum states ρ_j with probabilities p_j. Then

$$\lambda(\rho) \prec \sum_j p_j \lambda(\rho_j) \tag{6.12}$$

$$\bigoplus_j p_j \lambda(\rho_j) \prec \lambda(\rho). \tag{6.13}$$

In the statement of the theorem, the notation $\oplus$ denotes a direct sum of vectors. Note that the vectors on the left- and right-hand sides of (6.13) may therefore be of different dimension. In such cases we extend whichever vector is of lesser dimension by padding it with zero entries, to enable comparison using the majorization relation. As an example of this convention, suppose $p_1 = 1/3, p_2 = 2/3, \rho_1 = \mathrm{diag}(3/4, 1/4)$ and $\rho_2 = \mathrm{diag}(1/5, 4/5)$. Then Equation (6.13) becomes

$$\frac{1}{3}\begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix} \bigoplus \frac{2}{3}\begin{bmatrix} \frac{4}{5} \\ \frac{1}{5} \end{bmatrix} \prec \lambda\left(\frac{1}{3}\begin{bmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} + \frac{2}{3}\begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{4}{5} \end{bmatrix}\right), \tag{6.14}$$

which is equivalent to

$$\begin{bmatrix} \frac{1}{4} \\ \frac{1}{12} \\ \frac{8}{15} \\ \frac{2}{15} \end{bmatrix} \prec \begin{bmatrix} \frac{37}{60} \\ \frac{23}{60} \\ 0 \\ 0 \end{bmatrix}. \tag{6.15}$$

Proof: Equation (6.12) is an immediate consequence of the fact that $\lambda(A+B) \prec \lambda(A) + \lambda(B)$ for any two Hermitian matrices A and B, as proved in Subsection **??**.

Proof of (6.13): As noted in Subsection **??**, if a density matrix ρ can be written as a convex combination of pure states $|\psi_i\rangle$, $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, then it follows that $(p_i) \prec \lambda(\rho)$, where (p_i) denotes the vector whose entries are the probabilities p_i. Equation (6.13) is a corollary of this result. To see this, note that if r_{ij} are the eigenvalues of ρ_i and $|i,j\rangle$ the corresponding orthonormal eigenvectors then (6.13) is equivalent to the equation

$$(p_i r_{ij}) \prec \lambda(\rho), \tag{6.16}$$

which follows from the results of Subsection **??** and the observation that

$$\rho = \sum_i p_i \rho_i = \sum_{ij} p_i r_{ij} |i,j\rangle\langle i,j|. \tag{6.17}$$

This completes the proof of the theorem. ■

6.2.1 Dynamical constraints on quantum measurement

Theorem 2: Suppose $\{E_i\}$ is a set of measurement matrices satisfying the completeness relation $\sum_i E_i^\dagger E_i = I$. Then the quantum measurement described by these matrices must satisfy the following four constraints:

$$\lambda\left(\sum_i E_i \rho E_i^\dagger\right) \prec \sum_i \lambda\left(E_i \rho E_i^\dagger\right) \tag{6.18}$$

$$\bigoplus_i \lambda\left(E_i \rho E_i^\dagger\right) \prec \lambda\left(\sum_i E_i \rho E_i^\dagger\right) \tag{6.19}$$

$$\lambda(\rho) \prec \sum_i \lambda\left(E_i \rho E_i^\dagger\right) \tag{6.20}$$

$$\bigoplus_i \lambda\left(E_i \rho E_i^\dagger\right) \prec \lambda(\rho). \tag{6.21}$$

A slightly different way of stating Theorem 2 is to define p_i to be the probability of obtaining outcome i when the measurement defined by the matrices $\{E_i\}$ is performed on the system, and let $\rho'_i = E_i \rho E_i^\dagger / \mathrm{tr}(E_i \rho E_i^\dagger)$ be the corresponding posterior states. Then the following four equations are equivalent to (6.18)-(6.21):

$$\lambda\left(\sum_i p_i \rho'_i\right) \prec \sum_i p_i \lambda(\rho'_i) \tag{6.22}$$

$$\bigoplus_i p_i \lambda(\rho'_i) \prec \lambda\left(\sum_i p_i \rho'_i\right) \tag{6.23}$$

$$\lambda(\rho) \prec \sum_i p_i \lambda(\rho'_i) \tag{6.24}$$

$$\bigoplus_i p_i \lambda(\rho'_i) \prec \lambda(\rho). \tag{6.25}$$

Theorem 2 is a fundamental constraint on the dynamics that may occur during a quantum measurement. Equations (6.22) and (6.23) are, of course, merely the dynamical expression of the static constraints found earlier in Theorem 1. Equations (6.24) and (6.25) represent novel constraints of an essentially dynamical nature, connecting as they do the prior and posterior states of the quantum measurement. Intuitively, Equation (6.24) captures the notion that a quantum measurement "gains information" (on average) about a quantum state, since it says that the eigenvalues of the initial state ρ are, on average, more disordered than the eigenvalues of the posterior states ρ'_i. Intuitively, the second dynamic constraint, (6.25) captures the notion that the *total ensemble* of possible quantum states is more disordered after the measurement than before. Thus, (6.24) and (6.25) represent complementary constraints on the evolution of a quantum system during a quantum measurement process.

The constraints (6.22)-(6.25) are applicable even for very complex measurement processes. For example, a single mode cavity undergoing direct photodetection by an ideal photodetector can be described by a special case of the generalized measurements formalism known as the *quantum trajectories* or *stochastic Schrödinger equation* picture (see [48, 69] for a review and references). In this picture, if the system is started in the state ρ then the final state of the system is ρ_h, where "h" is used here to denote not just a single measurement outcome, but rather the complete history recorded by the photodetector, that is, all the times at which photocounts occurred.

Then (6.24) and (6.25) may be written as

$$\lambda(\rho) \prec \int d\mu(h)\lambda(\rho_h) \tag{6.26}$$

$$\bigoplus_h d\mu(h)\lambda(\rho_h) \prec \lambda(\rho), \tag{6.27}$$

where the integral is a functional integral over all possible photodetection histories, and $d\mu(h)$ is the corresponding measure on histories.

Proof of Theorem 2: The first two equations of Theorem 2, (6.18) and (6.19), are immediate consequences of the deeper *static* constraints on quantum mechanics introduced in Theorem 1; here we are merely enumerating the implications these static constraints have for dynamics. The remaining constraints, (6.20) and (6.21), are genuine quantum dynamical constraints relating the prior and posterior states of a quantum measurement.

Proof of (6.20): Suppose ρ is a positive matrix which can be written in the block form:

$$\rho = \begin{bmatrix} A & X \\ X^\dagger & B \end{bmatrix}. \tag{6.28}$$

For our purposes ρ will most often be a density matrix (and thus satisfy $\mathrm{tr}(\rho) = 1$), but the results we prove hold for a general positive matrix. We will show that $\lambda(\rho) \prec \lambda(A) + \lambda(B)$. (Recall our conventions on padding, which imply that the vectors of eigenvalues for A and B are to be extended by zeroes in such a way that they contain as many entries as the vector of eigenvalues of ρ). ρ is a positive matrix, so there must exist a matrix $D = [D_1|D_2]$ such that $\rho = D^\dagger D$, where the matrices D_1 and D_2 have the same number of columns as A and B, respectively, and both have the same number of rows as ρ. Thus we have

$$\begin{bmatrix} A & X \\ X^\dagger & B \end{bmatrix} = D^\dagger D = \begin{bmatrix} D_1^\dagger D_1 & D_1^\dagger D_2 \\ D_2^\dagger D_1 & D_2^\dagger D_2 \end{bmatrix}, \tag{6.29}$$

from which we read off $A = D_1^\dagger D_1$ and $B = D_2^\dagger D_2$. Using the results of Subsection **??** and the fact that the eigenvalues of a product EF of matrices E and F are the same as the eigenvalues of FE, up to padding by zeroes, we see that

$$\lambda(\rho) = \lambda(D^\dagger D) \tag{6.30}$$

$$
\begin{aligned}
&= \lambda(DD^\dagger) && (6.31)\\
&= \lambda(D_1D_1^\dagger + D_2D_2^\dagger) && (6.32)\\
&\prec \lambda(D_1D_1^\dagger) + \lambda(D_2D_2^\dagger) && (6.33)\\
&= \lambda(D_1^\dagger D_1) + \lambda(D_2^\dagger D_2) && (6.34)\\
&= \lambda(A) + \lambda(B), && (6.35)
\end{aligned}
$$

and thus $\lambda(\rho) \prec \lambda(A) + \lambda(B)$, as claimed. This method for eliminating off-diagonal block terms was introduced by Wielandt to connect the Weyl and Aronszajn inequalities (cited as [67] in Chapter 3 of [8].)

As a straightforward consequence we see by induction that for any positive matrix ρ and complete set of orthogonal projectors $\{P_i\}$:

$$\lambda(\rho) \prec \sum_i \lambda(P_i \rho P_i) \tag{6.36}$$

Extending even further, suppose $\{E_i\}$ is any set of measurement matrices defining a generalized measurement, and ρ is a positive matrix. As in Subsection **??** we can introduce an ancilla system with an orthonormal basis $|i\rangle$ in one-to-one correspondence with the indices on the measurement matrices E_i and define a unitary matrix U which has the action

$$U|\psi\rangle|0\rangle = \sum_i E_i|\psi\rangle|i\rangle, \tag{6.37}$$

where $|0\rangle$ is some standard state of the ancilla. Then we have $\lambda(\rho) = \lambda(\rho \otimes |0\rangle\langle 0|)$, since the non-zero eigenvalues of ρ and $\rho \otimes |0\rangle\langle 0|$ are the same. Simple algebra and (6.36) imply that

$$
\begin{aligned}
\lambda(\rho) &= \lambda(U(\rho \otimes |0\rangle\langle 0|)U^\dagger) && (6.38)\\
&\prec \sum_i \lambda((I \otimes |i\rangle\langle i|)U(\rho \otimes |0\rangle\langle 0|)U^\dagger(I \otimes |i\rangle\langle i|)) && (6.39)\\
&= \sum_i \lambda(E_i \rho E_i^\dagger \otimes |i\rangle\langle i|) && (6.40)\\
&= \sum_i \lambda(E_i \rho E_i^\dagger), && (6.41)
\end{aligned}
$$

where in the last line we used $\lambda(E_i\rho E_i^\dagger \otimes |i\rangle\langle i|) = \lambda(E_i \rho E_i^\dagger)$, since the non-zero entries agree. This completes the proof of (6.20).

Proof of (6.21): Again, let U be the unitary matrix constructed in Subsection **??** to implement the measurement described by the measurement

matrices $\{E_i\}$, namely, any unitary matrix having the action

$$U|\psi\rangle|0\rangle = \sum_i E_i|\psi\rangle|i\rangle. \tag{6.42}$$

Again, we have $\lambda(\rho) = \lambda(\rho \otimes |0\rangle\langle 0|)$, since the non-zero eigenvalues of ρ are the same as those of $\rho \otimes |0\rangle\langle 0|$, and thus $\lambda(\rho) = \lambda\left(U(\rho \otimes |0\rangle\langle 0|)U^\dagger\right)$. It follows from Equation (B.9) that

$$\lambda\left(\sum_i (I \otimes |i\rangle\langle i|)U(\rho \otimes |0\rangle\langle 0|)U^\dagger(I \otimes |i\rangle\langle i|)\right) \prec \lambda(\rho), \tag{6.43}$$

and thus

$$\lambda\left(\sum_i E_i \rho E_i^\dagger \otimes |i\rangle\langle i|\right) \prec \lambda(\rho). \tag{6.44}$$

This last equation is obviously equivalent to the statement we set out to prove,

$$\bigoplus_i \lambda\left(E_i \rho E_i^\dagger\right) \prec \lambda(\rho), \tag{6.45}$$

which concludes the proof of Theorem 2. ■

6.2.2 Consequences of the constraint equations

The constraints proved in Theorems 1 and 2 are very strong and, not surprisingly, have many interesting consequences. We now elucidate a few of these consequences using the notions of *Schur-concavity* and *Schur-convexity*. A Schur-convex function $f(\cdot)$ is a real-valued function which preserves the majorization relation, in the sense that if $x \prec y$ then $f(x) \leq f(y)$. Simple necessary and sufficient conditions for a function to be Schur-convex are known [8], and many interesting functions are Schur-convex. These include, for example, the function $x \to f(x) \equiv \sum_{j=1}^d x_j^k$, for any $k \geq 1$. Similarly, a *Schur-concave* function $f(\cdot)$ is one such that if $x \prec y$ then $f(x) \geq f(y)$. Equivalently, $f(\cdot)$ is Schur-concave if $-f(\cdot)$ is Schur-convex. Perhaps the canonical example of a Schur-concave function is the Shannon entropy $H(x) = -\sum_j x_j \log_2(x_j)$, so that whenever $x \prec y$ it follows that $H(x) \geq H(y)$, giving further justification to the intuitive notion that $x \prec y$ means that x is more disordered

than y. Applying the Schur-concavity of Shannon's entropy to the results of Theorems 1 and 2 we obtain an attractive suite of results. First, applying the Schur-concavity of $H(\cdot)$ to (6.12) gives

$$S(\rho) \quad \geq \quad H\left(\sum_i p_i \lambda(\rho_i)\right). \tag{6.46}$$

Applying the concavity of the Shannon entropy to the right hand side, we obtain as a corollary the concavity of the von Neumann entropy,

$$S(\rho) \quad \geq \quad \sum_i p_i S(\rho_i). \tag{6.47}$$

Applying the Schur-concavity of $H(\cdot)$ to (6.13) and doing some simple algebra gives

$$\sum_i p_i S(\rho_i) + H(p_i) \geq S(\rho). \tag{6.48}$$

This result was obtained previously by Lanford and Robinson[34] using different techniques. Applying the Schur-concavity of $H(\cdot)$ to (6.24), followed by the concavity of the Shannon entropy, gives

$$S(\rho) \geq \sum_i p_i S(\rho'_i). \tag{6.49}$$

Essentially the same result has been obtained previously in the context of entanglement transformation [6], where it expresses the fact that local processes cannot increase the amount of entanglement present in a system. Finally, applying the Schur-concavity of $H(\cdot)$ to (6.25) gives the beautiful inequality

$$H(p_i) + \sum_i p_i S(\rho'_i) \geq S(\rho), \tag{6.50}$$

which implies that in order to lower the entropy of a system by an amount Δ, on average, the information $H(p_i)$ collected by the measurement must be at least as large as Δ. This fact can be seen as a quantum mechanical expression of the principle, expressed by Landauer[33] and fleshed out by Bennett[4] and Zurek[70], that measurement of a physical system carries with it a thermodynamic cost when the measurement record is erased, and proper accounting of this cost enables one to solve the conundrum posed by Maxwell's demon. (See [5] for a review.)

Applying the Schur-convexity of the functions $f(x) = \sum_i x_i^k$ for $k \geq 1$ to the results of Theorems 1 and 2 also give a number of interesting constraints. The arguments used are analogous to those given above for the Shannon entropy, so the details will be omitted, and we merely state the results:

$$\sum_i p_i^k \mathrm{tr}\left(\rho_i^k\right) \leq \mathrm{tr}\left(\rho^k\right) \leq \sum_i p_i \mathrm{tr}\left(\rho_i^k\right) \tag{6.51}$$

$$\sum_i p_i^k \mathrm{tr}\left((\rho_i')^k\right) \leq \mathrm{tr}\left(\rho^k\right) \leq \sum_i p_i \mathrm{tr}\left((\rho_i')^k\right). \tag{6.52}$$

6.3 Partial converses to the constraints on mixing and measurement

Given the constraints on mixing and measurement described in Theorems 1 and 2 it is natural to ask if these constraints completely *characterize* the processes of mixing and measurement, respectively. We will show below that the answer to this question is *no*. However, partial progress towards achieving simple characterizations of mixing and measurement may be reported in the form of a partial converse to Theorem 1, described below in Subsection 6.3.1, and a partial converse to Theorem 2, described in Subsection 6.3.2.

6.3.1 Partial converse to the constraints on mixing

Given the constraints Theorem 1 imposes on mixing it is natural to ask whether these constraints completely characterize the mixing process. That is, given a density matrix ρ, probabilities p_i and vectors λ_i with non-negative, non-increasing components which sum to one, and such that

$$\lambda(\rho) \prec \sum_i p_i \lambda_i \tag{6.53}$$

$$\bigoplus_i p_i \lambda_i \prec \lambda(\rho), \tag{6.54}$$

does it follow that there exist density matrices ρ_i such that $\lambda(\rho_i) = \lambda_i$ and $\rho = \sum_i p_i \rho_i$?

We will show below that the answer to this question is no, however I suspect that some characterization along similar lines is possible. Progress towards such a characterization can be reported in the form of a partial converse to Theorem 1, which states that provided (6.53) holds then there

exist states ρ_{ij} and a probability distribution p_{ij} such that $\lambda(\rho_{ij}) = \lambda_i$, independent of the value of the index j, and $p_i = \sum_j p_{ij}$ for each i, as well as $\rho = \sum_{ij} p_{ij}\rho_{ij}$. That is, in order to obtain a converse to (6.53) we need to introduce an extra index, j. We will show below that it is necessary to introduce the extra index if only (6.53) is assumed as a hypothesis for the converse. Let's state and prove the partial converse as Theorem 3.

Theorem 3: Suppose ρ is a density matrix and λ_i are vectors with non-negative, non-increasing components summing to one. Suppose p_i are probabilities such that

$$\lambda(\rho) \prec \sum_i p_i \lambda_i. \tag{6.55}$$

Then there exist density matrices ρ_{ij} and a probability distribution p_{ij} such that $p_i = \sum_j p_{ij}$, $\lambda(\rho_{ij}) = \lambda_i$, and $\rho = \sum_{ij} p_{ij}\rho_{ij}$.

To prove Theorem 3 we need the result stated in Subsection **??** that $x \prec y$ if and only if there exist probabilities q_j and permutation matrices P_j such that $x = \sum_j q_j P_j y$. Applying this result with the assumption (6.55) we obtain

$$\lambda(\rho) = \sum_{ij} p_i q_j P_j \lambda_i. \tag{6.56}$$

Working in the basis in which ρ is diagonal, and defining Λ_i to be the diagonal matrix with diagonal entries λ_i, we may set $p_{ij} \equiv p_i q_j$ and $\rho_{ij} \equiv P_j \Lambda_i P_j^\dagger$, obtaining $p_i = \sum_j p_{ij}$ and $\lambda(\rho_{ij}) = \lambda_i$. Finally, the equation $\rho = \sum_{ij} p_{ij}\rho_{ij}$ follows immediately from these definition and (6.56), completing the proof.

What of a tight converse to Theorem 1? It is easy to see that it is not possible to obtain a tight converse to (6.53) alone, as follows. Suppose we choose $\rho = I/2$ to be the completely mixed state of a single qubit, and define a probability distribution on just one outcome, the trivial distribution $p_1 = 1$, with corresponding vector $\lambda_1 = (1, 0)$. Clearly, $\lambda(\rho) \prec \sum_i p_i \lambda_i$, yet it is not possible to find a state ρ_1 such that $\rho = p_1 \rho_1$ and $\lambda(\rho_1) = \lambda_1$. Thus, in this example, it is necessary to introduce extra indices, just as was done in Theorem 3.

Might it be that conditions (6.53) and (6.54) together completely characterize the mixing process? The following example, due to Julia Kempe, shows that this is not the case. Suppose we consider a qubit system, and choose $\rho = \mathrm{diag}(5/12, 7/12)$, $p_1 = p_2 = 1/2$, and $\lambda_1 = (1, 0), \lambda_2 = (1/2, 1/2)$. It is easy to verify that conditions (6.53) and (6.54) are satisfied with these

choices. Unfortunately, it is not possible to find states ρ_1 and ρ_2 with vectors of eigenvalues λ_1 and λ_2 such that $\rho = p_1\rho_1 + p_2\rho_2$, since with these choices for λ_1 and λ_2 it follows that ρ_1 must be a pure state and $\rho_2 = I/2$ the completely mixed state, so $p_1\rho_1 + p_2\rho_2$ has eigenvalues $3/4$ and $1/4$, which are not equal to $5/12$ and $7/12$. Despite this example, I believe it likely that conditions along the lines of (6.53) and (6.54) may be used to completely characterize the process of mixing in quantum mechanics.

6.3.2 Partial converse to the constraints on measurement

Given the constraints Theorem 2 imposes on the quantum measurement process it is natural to ask whether these constraints completely characterize the possible posterior states and probabilities which may occur in such a measurement? That is, supposing ρ is a density matrix, p_i is a probability distribution, and ρ'_i are density matrices such that

$$\lambda(\rho) \prec \sum_i p_i \lambda(\rho'_i) \tag{6.57}$$

$$\bigoplus_i p_i \lambda(\rho'_i) \prec \lambda(\rho), \tag{6.58}$$

does it follow that there exist measurement matrices $\{E_i\}$ satisfying the completeness relation $\sum_i E_i^\dagger E_i = I$ and giving the states ρ'_i as posterior states, with probabilities p_i, when the measurement is performed on a system initially prepared in the state ρ?

We will show below that the answer to this question is no, however I suspect that some characterization along similar lines is possible. Progress towards such a characterization can be reported in the form of a partial converse to Theorem 2, which states that provided the relation (6.57) holds, then there is a quantum measurement described by measurement matrices $\{E_{ij}\}$ such that the corresponding posterior states ρ'_{ij} satisfy $\rho'_{ij} = \rho_i$ for every j, and the measurement probabilities p_{ij} satisfy $\sum_j p_{ij} = p_i$. Thus, in order to obtain a converse to (6.57) we need to introduce an extra index, j, just as we did earlier in the partial converse to Theorem 1. Also analogously to that case, we show below that it is necessary to introduce the extra index with only (6.57) as hypothesis for the converse. Let's state and prove the partial converse as Theorem 4.

Theorem 4: Suppose ρ is a density matrix with vector of eigenvalues λ, and σ_i are density matrices with vectors of eigenvalues λ_i. Suppose p_i are probabilities such that

$$\lambda \prec \sum_i p_i \lambda_i \tag{6.59}$$

Then there exist matrices $\{E_{ij}\}$ and a probability distribution p_{ij} such that

$$\sum_{ij} E_{ij}^\dagger E_{ij} = I \tag{6.60}$$

$$E_{ij} \rho E_{ij}^\dagger = p_{ij} \sigma_i \tag{6.61}$$

$$\sum_j p_{ij} = p_i. \tag{6.62}$$

To prove Theorem 4, we again use the result that $x \prec y$ if and only if there exist probabilities q_j and permutation matrices P_j such that $x = \sum_j q_j P_j y$. By assumption we have $\lambda \prec \sum_i p_i \lambda_i$ and thus there exist permutation matrices P_j and probabilities q_j such that

$$\lambda = \sum_{ij} p_i q_j P_j \lambda_i. \tag{6.63}$$

Without losss of generality we may assume that ρ and σ_i are all diagonal in the same basis, with non-increasing diagonal entries, since if this is not the case then it is an easy matter to prepend or append unitary matrices to the measurement matrices to obtain the correct transformation. With this convention, we define matrices E_{ij} by

$$E_{ij} \sqrt{\rho} \equiv \sqrt{p_i q_j} \sqrt{\sigma_i} P_j^\dagger. \tag{6.64}$$

In order for E_{ij} to be well-defined by this formula alone it is necessary that ρ be invertible. If this is not the case then the E_{ij} are defined on the support of ρ by the formula (6.64), and to act as the zero operator on the orthocomplement of the support of ρ. It is convenient to let P be the projector onto the support of ρ. Note that we have

$$\sqrt{\rho} \left(\sum_{ij} E_{ij}^\dagger E_{ij} \right) \sqrt{\rho} = \sum_{ij} p_i q_j P_j \sigma_i P_j^\dagger. \tag{6.65}$$

Comparing with (6.63) we see that the right-hand side of the last equation is just ρ and thus

$$\sqrt{\rho}\left(\sum_{ij} E_{ij}^\dagger E_{ij}\right)\sqrt{\rho} = \rho, \tag{6.66}$$

from which we deduce that $\sum_{ij} E_{ij}^\dagger E_{ij} = P$, the projector onto the support of ρ. Letting $Q \equiv I - P$ be the projector onto the orthocomplement of the support, we can append an additional measurement matrix $E_{00} \equiv Q$ to the collection E_{ij} to ensure that the completeness relation $\sum_{ij} E_{ij}^\dagger E_{ij} = I$ is satisifed. Furthermore, from the definition (6.64) it follows that

$$E_{ij}\rho E_{ij}^\dagger \;=\; p_i q_j \sigma_i, \tag{6.67}$$

and thus upon performing a measurement defined by the measurement matrices $\{E_{ij}\}$ the result (i,j) occurs with probability $p_{ij} = p_i q_j$, $\sum_j p_{ij} = p_i$, and the post-measurement state is σ_i. This completes the proof of Theorem 4.

Theorem 4 is not a sharp converse to the condition of Equation (6.57) because of the extra index j. Introducing some such index is certainly necessary with the present hypotheses, as may be seen by considering an example with $\lambda = (1/2, 1/2)$, and the trivial probability distribution on one outcome, $p_1 = 1$, with $\lambda_1 = (1, 0)$. Then $\lambda \prec p_1\lambda_1$, but it is clear that there does not exist an E_1 such that $E_1\rho E_1^\dagger = \rho_1$, where $\lambda(\rho) = \lambda, \lambda(\rho_1) = \lambda_1$ and $E_1^\dagger E_1 = I$, because the last equation implies that E_1 must be unitary. It is not difficult to construct more complex examples to convince oneself that this behaviour is generic.

Might it be that the conditions (6.57) and (6.58) together characterize the posterior states and probabilities achievable through a quantum measurement? The following argument, due to Julia Kempe and the author, shows that this is not the case. Suppose we consider a qubit system, and choose $\rho = \mathrm{diag}(5/12, 7/12)$, $p_1 = p_2 = 1/2$, and $\rho_1' = \mathrm{diag}(1, 0)$, $\rho_2' = \mathrm{diag}(1/2, 1/2)$. It is easy to verify that conditions (6.57) and (6.58) are satisfied with these choices. Unfortunately, it is not possible to find measurement matrices E_1 and E_2 satisfying $\sum_i E_i^\dagger E_i = I$ and giving posterior states ρ_1' and ρ_2' with equal probabilites $1/2$, when the state ρ is measured. This can be seen in a variety of ways. A simple direct way is to note that the purity of ρ_1' implies that E_1 must have the form $E_1 = \alpha|a\rangle\langle b|$ for normalized states $|a\rangle$ and $|b\rangle$, and some $\alpha > 0$. Thus

$$E_2^\dagger E_2 \;=\; I - E_1^\dagger E_1 \tag{6.68}$$

$$= I - \alpha^2 |b\rangle\langle b| \tag{6.69}$$
$$= (1-\alpha^2)|b\rangle\langle b| + |c\rangle\langle c|, \tag{6.70}$$

where $|c\rangle$ is orthonormal to $|b\rangle$. The polar decomposition gives $E_2 = U\sqrt{E_2^\dagger E_2}$ for some unitary U, so

$$E_2 = \sqrt{1-\alpha^2}U|b\rangle\langle b| + U|c\rangle\langle c|. \tag{6.71}$$

We are requring that $E_2 \rho E_2^\dagger = I/4$, so it must be the case that E_2 is non-singular, and thus $\alpha < 1$. Premultiplying by E_2^{-1} and postmultiplying by $(E_2^\dagger)^{-1}$ gives

$$\rho = \frac{1}{4(1-\alpha^2)}|b\rangle\langle b| + \frac{1}{4}|c\rangle\langle c|. \tag{6.72}$$

Since $|b\rangle$ and $|c\rangle$ are orthonormal it follows that such a ρ cannot be equal to $\mathrm{diag}(5/12, 7/12)$, which is the desired contradiction. Despite this example, I believe it likely that conditions along the lines of (6.57) and (6.58) may be used to characterize the process of measurement in quantum mechanics.

6.4 Conclusion

We have shown that there are strong fundamental constraints on the processes of mixing and measurement in quantum mechanics that may be naturally expressed in the language of majorization. Although the results in the present paper don't completely characterize these processes, they suggest that there may exist a simple set of conditions which substantially simplify the usual characterization of these processes via operator equations. Another interesting direction for further research is to generalize the constraints on measurements obtained in this paper to better understand how two or more states may transform simultaneously under a measurement. Once again, although this problem is in principle already "solved", in the sense that there is an operator equation specifying exactly what transformations may occur, results such as those in the present paper and in [13] indicate that much more explicit characterizations may be possible. Such explicit conditions are likely to have applications to fundamental problems such as the problem of transformation of mixed state entanglement[6], and to the problem of determining to what extent the acquisition of information about the identity of a quantum state disturbs the system being measured[16].

Part III

Advanced theory of majorization

Chapter 7

Submajorization

In the usual definition of majorization, we say that $r \prec s$ if $\sum_{j=1}^{k} r_j^{\downarrow} \leq \sum_{j=1}^{k} s_j^{\downarrow}$ for all k, with equality when $k = d$, the dimension of the space in which r and s live. We have seen that this definition arises naturally in a number of contexts, yet there are other contexts in which it helps to make use of the related concept of *submajorization*. We say that r is submajorized by s, and write $r \prec_w s$, if

$$\sum_{j=1}^{k} r_j^{\downarrow} \leq \sum_{j=1}^{k} s_j^{\downarrow} \tag{7.1}$$

for $k = 1, \ldots, d$. The difference between majorization and submajorization is that we *do not* require equality when $k = d$.

The purpose of this chapter is to explore the basic properties of submajorization, and to provide some illustrative applications. In Section 7.1 we explain a connection between submajorization and the so-called *doubly substochastic* matrices, paralleling Theorem 3.1.2, which characterized majorization in terms of doubly stochastic matrices. Section 7.2 applies this characterization to the problem of characterizing an important set of invariants, the singular values, of a sum of two matrices. Finally, in Section 7.3 we apply the results of the previous sections to obtain insight into the properties of quantum entanglement.

Note that our development will rely heavily on Birkhoff's theorem, which was introduced in Section 3.1 on page 22, and developed in detail in Appendix A. In particular, the discussion below will be couched, to some extent, in the language of convex analysis introduced in the appendix, so the

reader who has not already done so is advised to look at the appendix to familiarize themselves with the approach taken there. In particular, it will help to understand the first few paragraphs of material in Section A.3, up to the statement of Birkhoff's theorem, if not the actual details of the proofs.

7.1 Double substochasticity and submajorization

A matrix $D = (D_{ij})$ is said to be *doubly substochastic* if it has non-negative entries, and all row and column sums are *at most* 1. Note that all doubly stochastic matrices are also doubly substochastic. Furthermore, it is clear that any square submatrix of a doubly stochastic matrix is doubly substochastic. The following theorem shows that the converse is true.

> *Theorem 7.1.1*: Suppose D is doubly substochastic. Then D is a submatrix of a doubly stochastic matrix E. We say that E is a *dilation* of D.

Proof: Let R be a diagonal matrix with entries 1 minus the corresponding row sums of D. Let C be a diagonal matrix with entries 1 minus the corresponding column sums of I. Then a dilation of D with the required property is

$$E \equiv \begin{bmatrix} D & R \\ C & D^T \end{bmatrix}. \tag{7.2}$$

■

The set of d by d doubly substochastic matrices is obviously convex. The following theorem characterizes the extremal points of this set. However, instead of needing to go through a complex argument, as was needed for Birkhoff's theorem, we can instead use the embedding of the doubly substochastic matrices in the doubly stochastic matrices to apply Birkhoff's theorem directly.

> *Theorem 7.1.2*: The extremal points of the d by d doubly substochastic matrices are the matrices having at most one entry 1 in each row and column, and 0s elsewhere.

Proof: It is clear that matrices of this sort are extreme points in the set of doubly substochastic matrices. To show that they exhaust the extreme points, let D be doubly substochastic. Dilate D to a doubly stochastic E. By Birkhoff's theorem $E = \sum_j p_j P_j$ for probabilities p_j and permutation matrices P_j. Observe that any submatrix of P_j is a matrix having at most one entry 1 in each row and column, and 0s elsewhere. ■

We have already seen one way of embedding the doubly substochastic matrices in the doubly stochastic matrices. Here is another useful technique:

> *Theorem 7.1.3*: A d by d matrix $D = (D_{jk})$ is doubly substochastic if and only if there exists a d by d doubly stochastic $E = (E_{jk})$ such that $D_{jk} \leq E_{jk}$ for all j and k.

Proof: The reverse implication is clear, so we need only prove the forward implication. Write $D = \sum_j p_j Q_j$, where the p_j are probabilities, and the Q_j are matrices with at most one entry 1 in each row and column, and 0s elsewhere. Choose permutation matrices P_j such that $Q_j \leq P_j$ elementwise. E defined by $E \equiv \sum_j p_j P_j$ is the required matrix. ■

We can now prove an analogue of Theorem 3.1.2, characterizing submajorization in terms of doubly substochastic matrices.

> *Theorem 7.1.4*: Let r and s be real d-dimensional vectors with non-negative entries. Then $r \prec_w s$ if and only if there exists a doubly substochastic matrix D such that $r = Ds$.

Proof: Suppose $r \prec_w s$, and let $\Delta \equiv \sum_{j=1}^{d}(s_j - r_j)$, so $\Delta \geq 0$. Choose n sufficiently large that Δ/n is smaller than the smallest positive component of r. Suppose we dilate r and s to $d+n$-dimensional vectors r' and s' as follows:

$$r' = \left(r_1, \ldots, r_d, \frac{\Delta}{n}, \ldots, \frac{\Delta}{n}\right) \tag{7.3}$$

$$s' = (s_1, \ldots, s_d, 0, \ldots, 0)\,. \tag{7.4}$$

Then a little thought shows that $r' \prec d'$, and thus by Theorem 3.1.2, $r' = D's'$, for some doubly stochastic matrix D'. It follows that $r = Ds$, where D is a submatrix of D', and thus is doubly substochastic.

Conversely, suppose $r = Ds$, where D is doubly substochastic. By Theorem 7.1.3, there exists a doubly stochastic E such that $D_{jk} \leq E_{jk}$ for all

pairs of indices j and k. Theorem 3.1.2 implies that $Es \prec s$. Furthermore, it is easy to see that $Ds \leq Es$, where the relation $x \leq y$ on vectors means that $x_j \leq y_j$ for all indices j. Thus $Ds \leq Es \prec s$, from which we can verify directly that $Ds \prec_w s$. ■

7.2 The singular values of a sum of matrices

Consider

7.3 Majorization, submajorization, and the Schmidt decomposition for entangled quantum states

This

Problems for Lecture 7

Problem 7.3.1: Let r and s be real d-dimensional vectors with non-negative entries. Then $r \prec_w s$ if and only if there exist unitary matrices u and v such that $r_j = \sum_k |u_{jk}||v_{jk}|s_k$.

Chapter 8

Functions preserving majorization

8.0.1 Isotone functions

The theory of *isotone functions* is concerned with maps that preserve the majorization relation. For example, given two probability distributions p_i and q_i such that $(p_i) \prec (q_i)$, we will show that $-H(p_i) \leq -H(q_i)$, where $H(\cdot)$ is the Shannon entropy.

This example illustrates the simplest type of isotone function, a *Schur-convex* function. A function $f : \mathbf{R}^d \to \mathbf{R}$ is said to be *Schur-convex* if

$$x \prec y \;\Rightarrow\; f(x) \leq f(y). \tag{8.1}$$

A function f is *Schur-concave* if $-f$ is Schur-convex.

There is a close connection between Shur-convexity and more familiar notions of convexity:

Theorem 8.0.1: The following conditions are equivalent:

1. $x \prec y$.
2. $F(x) \leq F(y)$, where $F(x) \equiv \sum_{i=1}^{d} f(x_i)$, for all convex functions $f : \mathbf{R} \to \mathbf{R}$.

Proof:

Suppose $x \prec y$ and $f(\cdot)$ is a convex function. Then $x = \sum_i p_i P_i y$ for some set of probabilities p_i and permutation matrices P_i. $F(\cdot)$ is a sum of convex

functions, and thus convex, so

$$F(x) \leq \sum_i p_i F(P_i y). \tag{8.2}$$

But $F(\cdot)$ is manifestly permutation invariant, so $F(P_i y) = F(y)$, and thus

$$F(x) \leq \sum_i p_i F(y) = F(y), \tag{8.3}$$

as required.

Conversely, suppose $F(x) \leq F(y)$ for all convex functions $f(\cdot)$. Define $f(x) \equiv |x - t|$, where t is some real parameter. Then $f(\cdot)$ is convex, so $F(x) \leq F(y)$, that is

$$\sum_{i=1}^{d} |x_i - t| \leq \sum_{i=1}^{d} |y_i - t|. \tag{8.4}$$

Since this holds for any t, by the order-free characterization of majorization (Theorem 2.2.1 on page 14), it follows that $x \prec y$.

■

One consequence of this result is that the Shannon entropy is Schur-concave; to see this, just note that $f(x) \equiv x \log(x)$ is convex. Similarly, the convexity of x^k (for any $k \geq 1$) implies that $\sum_i x_i^k$ is a Schur-convex function. Indeed, it is possible to give a complete characterization of the differentiable Schur-convex functions:

> *Theorem 8.0.2*: (**Characterization of Schur-convexity**)
> Suppose $f : \mathbf{R}^d \to \mathbf{R}$ is a differentiable function. Then $f(\cdot)$ is Schur-convex if and only if the following two conditions are satisfied:
>
> 1. $f(\cdot)$ is *permutation invariant*, that is, $f(Px) = f(x)$ for any permutation P.
> 2. For each $x \in \mathbf{R}^d$ and for any pair of indices i and j,
>
> $$(x_i - x_j)\left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}\right) \geq 0. \tag{8.5}$$

Proof:

Suppose $f(\cdot)$ is Schur-convex. We will show that conditions 1 and 2 follow. Let P be any permutation. Then $x \prec Px \prec x$, so $f(x) \leq f(Px) \leq f(x)$, whence $f(x) = f(Px)$, and f is permutation invariant. This shows condition 1. Without loss of generality, we will prove condition 2 for the case where $i = 1, j = 2$. Note that

$$((1-t)x_1 + tx_2, tx_1 + (1-t)x_2, x_3, \ldots, x_d) \prec (x_1, \ldots, x_d). \tag{8.6}$$

Defining $\Delta \equiv x_1 - x_2$ this may be rewritten

$$(x_1 - t\Delta, x_2 + t\Delta, x_3, \ldots, x_d) \prec (x_1, \ldots, x_d). \tag{8.7}$$

By the Schur-convexity of f this gives

$$f(x_1 - t\Delta, x_2 + t\Delta, x_3, \ldots, x_d) \leq f(x_1, \ldots, x_d). \tag{8.8}$$

Thus

$$\begin{aligned} 0 &\leq \lim_{t \to 0^+} \frac{f(x_1, \ldots, x_d) - f(x_1 - t\Delta, x_2 + t\Delta, x_3, \ldots, x_d)}{t} && (8.9) \\ &= \Delta \frac{\partial f}{\partial x_1} - \Delta \frac{\partial f}{\partial x_2}. && (8.10) \end{aligned}$$

Substituting $\Delta = x_1 - x_2$ gives the result.

Conversely, suppose $f(\cdot)$ satisfies conditions 1 and 2. We will show that $f(\cdot)$ is Schur-convex. To do this, it is sufficient to show that $f(Ty) \leq f(y)$, where T is a T-transform acting on the first two components of y. The reason this is sufficient is because the permutation invariance of $f(\cdot)$ ensures that the result is then true for any T-transform T, and if $x \prec y$ then $x = T_1 T_2 \ldots T_m y$ for some sequence of T-transforms $T_1, \ldots, T_m$.

Define

$$y(t) \equiv ((1-t)y_1 + ty_2, ty_1 + (1-t)y_2, y_3, \ldots, y_d). \tag{8.11}$$

Then

$$\begin{aligned} f(y(t)) - f(y(0)) &= \int_0^t \frac{df(y(s))}{ds} ds && (8.12) \\ &= \int_0^t \left(\frac{\partial f}{\partial y_1} \frac{dy_1}{ds} + \frac{\partial f}{\partial y_2} \frac{dy_2}{ds} \right) ds. && (8.13) \end{aligned}$$

But $y_1(s) \equiv (1-s)y_1 + sy_2$ and $y_2(s) \equiv (sy_1 + (1-s)y_2$. Thus

$$\begin{aligned} f(y(t)) - f(y(0)) &= \int_0^t \left(\frac{\partial f}{\partial y_1}(y_2 - y_1) + \frac{\partial f}{\partial y_2}(y_1 - y_2) \right) ds \quad (8.14) \\ &= \int_0^t - \left[(y_1 - y_2) \left(\frac{\partial f}{\partial y_1} - \frac{\partial f}{\partial y_2} \right) \right] ds. \quad (8.15) \end{aligned}$$

Since the integrand is negative, by assumption, it follows that $f(y(t)) - f(y(0)) \leq 0$, that is, $f(y(t)) \leq f(y(0))$, as we desired to show.

■

Example: The product $f(x) \equiv \prod_{i=1}^d x_i$ of the components of a vector is Schur-concave in the region of $\mathbf{R}^d$ where $x_i \geq 0$ for all components of x. To see this, note that $f(\cdot)$ is permutation invariant, and

$$(x_i - x_j) \left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j} \right) = (x_i - x_j) \left(\frac{1}{x_i} - \frac{1}{x_j} \right) f(x). \quad (8.16)$$

Example: Suppose $1 \leq k \leq d$. Define the *kth elementary symmetric polynomial* $S_k : \mathbf{R}^d \to \mathbf{R}$ by

$$S_k(x) \equiv \sum_{1 \leq i_1 \leq i_2 \leq \ldots \leq i_k \leq d} x_{i_1} x_{i_2} \ldots x_{i_k} \leq 0. \quad (8.17)$$

Then $S_k(x)$ is Schur-concave.

Exercise 8.0.1: Prove that the elementary symmetric polynomials are Schur-concave.

Exercise 8.0.2: Find a Schur-convex function that is not convex.

An *isotone* function $f : \mathbf{R}^d \to \mathbf{R}^m$ is a function such that $x \prec y$ implies $f(x) \prec_w y$. A function is *strongly isotone* if $x \prec_w y$ implies $f(x) \prec_w f(y)$. A function is *strictly isotone* if $x \prec y$ implies $f(x) \prec f(y)$. The following theorem gives a useful sufficient condition for isotonicity:

> *Theorem 8.0.3*: Let $f : \mathbf{R}^d \to \mathbf{R}^m$ be a convex map such that for any permutation matrix P in d dimensions there exists a permutation P' in m dimensions such that
>
> $$f(Px) = P'f(x). \quad (8.18)$$
>
> Then f is an isotone function.

By *convex* we mean that $f(\sum_i p_i x_i) \leq \sum_i p_i f(x_i)$, where p_i is any probability distribution, and $\leq$ means that the inequality holds for each component. *Proof:*

Suppose $x \prec y$. Then $x = \sum_i p_i P_i y$ for some probability distribution p_i and permutation matrices P_i. By the convexity of f,

$$f(x) \leq \sum_i p_i f(P_i y) \tag{8.19}$$

$$= \sum_i p_i P_i' f(y). \tag{8.20}$$

Setting $\tilde{x} \equiv \sum_i p_i P_i' f(y)$ we have

$$f(x) \leq \tilde{x} \prec f(y), \tag{8.21}$$

and thus $f(x) \prec_w f(y)$.

■

The *Schur-convex functions* are real-valued functions f such that $x \prec y$ implies $f(x) \leq f(y)$. Examples of Schur-convex functions include $f(x) \equiv \sum_i x_i \log(x_i)$, $f(x) \equiv \sum_i x_i^k$ (for any constant $k \geq 1$), $f(x) \equiv -\prod_i x_i$, and $f(x) \equiv -x_1^{\downarrow}$. More examples and a characterization of the Schur-convex functions may be found in [8, 38]. Each such Schur-convex function gives rise to an inequality relating the vector of probabilities (p_i) in Equation (**??**) to the vector λ^ρ. For example, we see from the Schur-convexity of $\sum_i x_i \log(x_i)$ the useful inequality that $H(p_i) \geq S(\rho)$, where $H(\cdot)$ is the Shannon entropy, and $S(\cdot)$ is the von Neumann entropy. (This result was obtained by Lanford and Robinson [34] using different techniques.) In general, any Schur-convex function will give rise to a similar inequality relating (p_i) and λ^ρ. A similar property related to convex functions has previously been noted (see the review [65] for an overview, as well as the original references [56, 57, 58, 59, 64]), however those results are a special case [8] of the more general result given here based upon Schur-convex functions, which may be obtained by noting that if $f(x)$ is convex then the map $(p_i) \to \sum_i f(p_i)$ is Schur-convex.

8.0.2 Binary functions and majorizations

A map $f : \mathbf{R}^2 \to \mathbf{R}$ is said to be *lattice superadditive* if it satisfies the condition

$$f(s_1, t_1) + f(s_2, t_2) \leq f(\min(s_1, s_2), \min(t_1, t_2)) + f(\max(s_1, s_2), \max(t_1, t_2)). \tag{8.22}$$

The connection between lattice superadditivity and majorization will be made shortly; for now we content ourselves with reformulating lattice superadditivity in a manner that is somewhat easier to deal with. Suppose we take two points, s and t, and some positive displacements $\epsilon, \delta > 0$. Then the condition of lattice superadditivity can be rewritten as

$$f(s+\delta, t) + f(s, t+\epsilon) \leq f(s,t) + f(s+\delta, t+\epsilon). \tag{8.23}$$

This rewriting of the definition also allows us to give a beautiful characterization of lattice superadditive fuctions using calculus:

> *Theorem 8.0.4*: (**Characterization of lattice superadditivity**)
> A twice differentiable function $f : \mathbf{R}^2 \to \mathbf{R}$ is lattice superadditive if and only if
>
> $$\frac{\partial^2 f}{\partial s \partial t} \geq 0. \tag{8.24}$$
>
> at all points.

Proof:

The result follows from the observation that

$$\int_s^{s+\delta} dx \int_t^{t+\epsilon} dy \, \frac{\partial f}{\partial x \partial y} = f(s+\delta, t+\epsilon) - f(s, t+\epsilon) - f(s+\delta, t) + f(s,t). \tag{8.25}$$

If f is lattice superadditive then the right hand side must be non-negative. Since δ and ϵ were arbitrary, the integrand must also be non-negative. Conversely, if the integrand is non-negative, then so is the right hand side, and thus f is lattice superadditive.

■

As examples of lattice superadditive functions we have $f(s,t) = s + t$ and $f(s,t) = st$ (on $\mathbf{R}^2_+$ for the latter function). These follow by the differentiability criteria.

Exercise 8.0.3: Suppose g is twice differentiable, convex, and increasing. Suppose f is twice differentiable, monotone increasing, and lattice superadditive. Show that the composition $g \circ f$ is monotone increasing and lattice superadditive.

Exercise 8.0.4: Show that $f(s,t) \equiv -(s-t)^2$ is lattice superadditive.

Exercise 8.0.5: Show that $f(s,t) \equiv |s-t|$ is not lattice superadditive.

The significance of lattice superadditivity follows from the following theorem. To state the theorem, we need a couple of conventions. First, we say that $f(\cdot,\cdot)$ is *monotone* if it is monotone in each of its arguments. Second, given $f : \mathbf{R}^2 \to \mathbf{R}$ we define $F : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d$ by

$$F(x,y) \equiv (f(x_1,y_1), f(x_2,y_2), \ldots, f(x_d,y_d)). \tag{8.26}$$

Theorem 8.0.5: If f is monotone and lattice superadditive, then for all x and y in $\mathbf{R}^d$,

$$F(x^{\downarrow}, y^{\uparrow}) \prec_w F(x,y) \prec^w F(x^{\downarrow}, y^{\downarrow}). \tag{8.27}$$

Proof:

■

The functions $f(x,y) = (x+y)$ and $f(x,y) = xy$ are both examples of monotone lattice superadditive functions. As a consequence we have the useful results:

$$(x^{\downarrow} + y^{\uparrow}) \prec_w x + y \prec^w x^{\downarrow} + y^{\downarrow} \tag{8.28}$$

$$x^{\downarrow} \cdot y^{\uparrow} \leq x \cdot y \leq x^{\downarrow} \cdot y^{\downarrow}. \tag{8.29}$$

The next result provides another connection between quantum mechanics and majorization that is especially useful in the present context.

Theorem 8.0.6:

Suppose A and B are Hermitian matrices. Then there exists a doubly stochastic matrix D depending only on the bases in which A and B are diagonal (and not on their eigenvalues) such that

$$\mathrm{tr}(AB) = (\lambda_A, D\lambda_B). \tag{8.30}$$

Conversely, for any vector b such that $b \prec \lambda_B$, there exists an operator $\tilde{B}$ unitarily equivalent to B such that

$$\mathrm{tr}(A\tilde{B}) = (\lambda_A, b). \tag{8.31}$$

Proof:

There exist unitary U and V such that

$$\mathrm{tr}(AB) = \mathrm{tr}(U\Lambda(A)U^\dagger V\Lambda(B)V^\dagger. \tag{8.32}$$

Defining $W \equiv V^\dagger U$ this may be rewritten

$$\begin{aligned} \mathrm{tr}(AB) &= \mathrm{tr}(W\Lambda(A)W^\dagger\Lambda(B)) & (8.33)\\ &= \sum_{ij} |W_{ij}|^2 \lambda(A)_j \lambda(B)_i. & (8.34) \end{aligned}$$

Defining $D_{ij} \equiv |W_{ij}|^2$ gives

$$\mathrm{tr}(AB) = (\lambda(A), D\lambda_B). \tag{8.35}$$

The converse now follows from the existence of an orthostochastic D such that $b = D\lambda_B$.

■

This is a very interesting result. It tells us that the average of an observable A can be written

$$\langle A\rangle = \mathrm{tr}(\rho A) = (\lambda_\rho, D\lambda_A), \tag{8.36}$$

for some doubly stochastic D depending only on the bases ρ and D are diagonal in.

One useful corollary of this result is that

$$(\lambda_A^\downarrow, \lambda_B^\uparrow) \leq \mathrm{tr}(AB) \leq (\lambda_A^\downarrow, \lambda_B^\downarrow). \tag{8.37}$$

This line of thought can be used to put useful bounds on the fidelity. The fidelity between density matrices ρ and σ is defined by

$$F(\rho,\sigma) \equiv \mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}. \tag{8.38}$$

It can be shown that

$$F(\rho,\sigma) = \max_U |\mathrm{tr}(\sqrt{\rho}\sqrt{\sigma}U)|, \tag{8.39}$$

where the maximization is over all unitary U. Thus

$$\begin{aligned} F(\rho,\sigma) &\geq \mathrm{tr}(\rho^{1/2}\sigma^{1/2}) & (8.40)\\ &\geq F(\lambda_\rho^\downarrow, \lambda_\sigma^\uparrow). & (8.41) \end{aligned}$$

Chapter 9

Lidskii's theorem

Recall the simple consequence of the Fan maximum principle that for Hermitian matrices R and S, $\lambda(R+S) \prec \lambda(R) + \lambda(S)$. There is a stronger perturbation result for eigenvalues known as *Lidskii's theorem*, which has a rather similar form:

$$\lambda(R+S) - \lambda(R) \prec \lambda(S). \tag{9.1}$$

It is easy to see that Lidskii's theorem implies the weaker result $\lambda(R+S) \prec \lambda(R) + \lambda(S)$. However, the converse implication does not follow easily (at least so far as is known). According to Bhatia [8], Lidskii's theorem was originally noted by Lidskii in 1950 [35], who provided an elementary matrix-analytic proof of a result of Berezin and Gel'fand [7]. The proof given in this appendix is a simplification of a proof given by Smiley [55].

Our proof of Lidskii's theorem relies on a simple lemma known as *Cauchy's interlacing theorem*, which states that if A is a d by d matrix and $B = PAP$ where P is a projector onto a $d-k$-dimensional space, then for $j = 1, 2, \ldots, d-k$

$$\lambda_j(A) \geq \lambda_j(B) \geq \lambda_{j+k}(A). \tag{9.2}$$

Cauchy's interlacing theorem is easily proved directly, and we omit the details. (See Section III.1 of [8] for a detailed proof.)

The proof of Lidskii's theorem is by induction on the dimension d the matrices R and S live in. For $d = 1$ the result is trivial, so we assume the result is correct in dimension $d-1$ and prove the result for dimension d, that is,

$$A \tag{9.3}$$

We used the result $\lambda(R+S) \prec \lambda(R) + \lambda(S)$ to establish a set of dynamical constraints on the amount of information obtained through a quantum measurement. Might it possible to further strengthen this result by using Lidskii's theorem? I don't know the answer to this question, but it does seem an interesting possibility for further work.

Part IV

Majorization and entanglement

Chapter 10

Entanglement transformation

10.1 Entanglement transformation

The problem of *entanglement transformation* is a natural context in which the results of the present paper may be applied. The problem of entanglement transformation arises as a consequence of the fundamental question of how may we convert one type of physical resource into another, and there has been considerable effort devoted to determining when it is possible to convert one type of entanglement to another. In [40] a connection was noted between entanglement transformation and majorization, namely, that if $|\psi\rangle$ and $|\phi\rangle$ are pure states of a bipartite quantum system with components belonging to Alice (A) and Bob (B) respectively, then Alice and Bob can transform the state $|\psi\rangle$ into the state $|\phi\rangle$ using local operations on their respective systems and classical communication between Alice and Bob, if and only if

$$\lambda_\psi \prec \lambda_\phi, \tag{10.1}$$

where λ_ψ (respectively λ_ϕ) is the vector of eigenvalues of the reduced density matrix for Alice's system when the joint system is in the state $|\psi\rangle$ ($|\phi\rangle$). As per usual, the components of such vectors are ordered into non-increasing order. This result has subsequently been generalized by Vidal[61] to the case of conclusive transformation, and even further by Jonathan and Plenio[31] to the problem where Alice and Bob are supplied with a state $|\psi\rangle$ and wish to tranform this state into an *ensemble* of states in which the state $|\phi_i\rangle$ occurs with probability p_i. (See also Hardy[19] for an instructive alternative approach to results of this type.) The necessary and sufficient condition for

such a transformation to be possible is that[31]:

$$\lambda_\psi \prec \sum_i p_i \lambda_{\phi_i}. \tag{10.2}$$

We now explain how this result can be seen as an easy consequence of the results proved in the present paper, and thus the connection between majorization and entanglement is really a consequence of a deeper connection between majorization and measurement. By a result of Lo and Popescu[37], it is possible to transform $|\psi\rangle$ into the ensemble $\{p_i, |\phi_i\rangle\}$ by local operations and classical communication if and only if it is possible to make the transformation via the following simplified procedure: first, Alice performs a generalized measurement on her state, then sends the result to Bob, who performs a unitary operation on his system conditional on the outcome of the measurement Alice made. Let $\rho = \mathrm{tr}_B(|\psi\rangle\langle\psi|)$ be the initial state of Alice's system, and suppose Alice performs a quantum measurement described by measurement matrices E_i, so that outcome i occurs with probability p_i and $(E_i \otimes U_i)|\psi\rangle = \sqrt{p_i}|\phi_i\rangle$, for some unitary operator U_i acting on Bob's system. Considering Alice's system alone and observing that that $E_i \rho E_i^\dagger = \sigma_i$, where $\sigma_i = p_i \mathrm{tr}(|\phi_i\rangle\langle\phi_i|)$, we deduce from Theorem 2 that

$$\lambda_\rho \prec \sum_i p_i \lambda_{\sigma_i}, \tag{10.3}$$

which is equivalent to (10.2). To prove the converse, suppose (10.2) holds. Then by Theorem 4 there exists a quantum measurement described by measurement matrices E_{ij}, and probabilities p_{ij} such that

$$E_{ij} \rho E_{ij}^\dagger = p_{ij} \sigma_i; \quad \sum_j p_{ij} = p_i. \tag{10.4}$$

The procedure for Alice and Bob to produce the ensemble is for Alice to perform the measurement described by the set E_{ij}. The post-measurement state $|\phi_{ij}\rangle$ is then a purification [43] of the state σ_i, and it can be shown (see [27] or Section 2.5 of [43]) that by performing an appropriate unitary transformation Bob can convert the state $|\phi_{ij}\rangle$ into the state $|\phi_i\rangle$, with total probability p_i of obtaining the state $|\phi_i\rangle$. Thus Equation (10.2) represents a necessary and sufficient condition for it to be possible to transform the state $|\psi\rangle$ into the ensemble $\{p_i, |\phi_i\rangle\}$ by local operations and classical communication.

Chapter 11

Application to separability

A remarkable feature of quantum entanglement is that an entangled state of two parties, Alice (A) and Bob (B), may be more disordered locally than globally. That is, $S(A) > S(A, B)$, where $S(\cdot)$ is the von Neumann entropy. It is known that satisfaction of this inequality implies that a state is non-separable. In this paper we prove the stronger result that for separable states the vector of eigenvalues of the density matrix of system AB is majorized by the vector of eigenvalues of the density matrix of system A alone. This gives a strong sense in which a separable state is more disordered globally than locally and a new *necessary* condition for separability of bipartite states in arbitrary dimensions. We also investigate the extent to which these conditions are *sufficient* to characterize separability, exhibiting examples that show separability cannot be characterized solely in terms of the local and global spectra of a state. We apply our conditions to give a simple proof that non-separable states exist sufficiently close to the completely mixed state of n qudits.

Quantum mechanics harbours a rich structure whose investigation and explication is the goal of quantum information science[43, 50]. At present only a limited understanding of the fundamental static and dynamic properties of quantum information has been obtained, and many major problems remain open. In particular, we would like a detailed ontology and quantitative methods of description for the different types of information and dynamical processes afforded by quantum mechanics. An example of the pursuit of these goals has been the partial development of a theory of quantum entanglement; see, e.g., [6, 40, 23, 60, 61] and references therein.

The *separability* or non-separability of a quantum state is a question that

has received much attention in the development of a theory of entanglement. The notion of separability captures the idea that a quantum state's static properties can be explained entirely by classical statistics, and is sometimes claimed to be equivalent to the notion that a state is "not entangled". More precisely, a state ρ_{AB} of Alice and Bob's system is separable[66] if it can be written in the form $\rho_{AB} = \sum_j q_j \rho_j \otimes \sigma_j$, for some probability distribution $\{q_j\}$, and density matrices ρ_j and σ_j of Alice and Bob's systems, respectively. Thus, we can think of Alice and Bob's systems as having a local, pseudo-classical description, as a mixture of the product states $\rho_j \otimes \sigma_j$ with probabilities q_j. Note that separability is equivalent to the condition

$$\rho_{AB} = \sum_j p_j |\psi_j\rangle\langle\psi_j| \otimes |\phi_j\rangle\langle\phi_j|, \tag{11.1}$$

where $\{p_j\}$ is a probability distribution and $|\psi_j\rangle, |\phi_j\rangle$ are pure states of Alice and Bob's systems, respectively.

One reason for interest in separability is a deep theorem due to M., P. and R. Horodecki connecting separability to positive maps on operators[23]. The Horodeckis used this theorem to prove that the "positive partial transpose" criterion for separability introduced by Peres[46] is a necessary and sufficient condition for separability of a state ρ_{AB} of a system consisting of a qubit in Alice's possession, and either a qubit or qutrit in Bob's possession. More precisely, if we define $\rho_{AB}^{T_B}$ to be the operator that results when the transposition map is applied to system B alone, then the Horodeckis showed that ρ_{AB} is separable if and only if $\rho_{AB}^{T_B}$ is a positive operator. Unfortunately, this criterion, while necessary for a state to be separable in higher dimensions[46], is not sufficient.

A hallmark of quantum entanglement is the remarkable fact that individual components of an entangled system may exhibit *more* disorder than the system as a whole. The canonical example of this phenomenon is a pair of qubits A and B prepared in the maximally entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$. The von Neumann entropy $S(A)$ of qubit A is equal to one bit, compared with a von Neumann entropy $S(A, B)$ of zero bits for the joint system. Classically, of course, such behaviour is impossible, and the Shannon entropy $H(X)$ of a single random variable is never larger than the Shannon entropy of two random variables, $H(X), H(Y) \leq H(X, Y)$. It has been shown [26] (see Chapter 8 of [39] and [12, 25] for related results) that an analogous relation holds for separable states,

$$S(A), S(B) \leq S(A, B). \tag{11.2}$$

This result is a consequence of the concavity of $S(A,B) - S(A)$ [36, 43], since when $\rho_{AB} = \sum_j q_j \rho_j \otimes \sigma_j$ we have $S(A,B) - S(A) \geq \sum_j q_j (S(\rho_j \otimes \sigma_j) - S(\rho_j)) \geq 0$. Unfortunately, the inequalities (11.2) are insufficient to characterize separability. To see this, consider the Werner state of two qubits $\rho_p = p|\Psi\rangle\langle\Psi| + (1-p)I/4$ $(0 \leq p \leq 1)$ and $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. The positive partial transpose criterion implies that the state is separable iff $p \leq 1/3$. The marginal density matrices being fully mixed for all p, however, one obtains $S(A) = S(B) = 1 \leq S(A,B) = H(\frac{1+3p}{4}, \frac{1-p}{4}, \frac{1-p}{4}, \frac{1-p}{4})$ for $0 \leq p \leq 0.747...$, so the condition (11.2) is fulfilled for a range of inseparable states.

The notion of von Neumann entropy is a valuable notion of disorder in a quantum state, however more sophisticated tools for quantifying disorder exist. One such tool is the theory of majorization, whose basic elements we now review (see Chapters 2 and 3 of [8], [38] or [1] for more extensive background). Suppose $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ are two d-dimensional real vectors; we usually suppose in addition that x and y are probability distributions, that is, the components are non-negative and sum to one. The relation $x \prec y$, read "x is majorized by y", is intended to capture the notion that x is more "mixed" (i.e. disordered) than y. Introduce the notation $\downarrow$ to denote the components of a vector rearranged into decreasing order, so $x^\downarrow = (x_1^\downarrow, \ldots, x_d^\downarrow)$, where $x_1^\downarrow \geq x_2^\downarrow \geq \ldots \geq x_d^\downarrow$. Then we define $x \prec y$, if

$$\sum_{j=1}^{k} x_j^\downarrow \leq \sum_{j=1}^{k} y_j^\downarrow, \tag{11.3}$$

for $k = 1, \ldots, d-1$, and with the inequality holding with equality when $k = d$. To understand how this definition connects with disorder consider the following result (see Chapter 2 of [8] for a proof): $x \prec y$ if and only $x = Dy$, where D is a doubly stochastic matrix. Thus, when $x \prec y$ we can imagine that y is the input probability distribution to a noisy channel described by the doubly stochastic matrix D, inducing a more disordered output probability distribution, x. Majorization can also be shown [8] to be a more stringent notion of disorder than entropy in the sense that if $x \prec y$ then it follows that $H(x) \geq H(y)$.

Given the known connections between measures of disorder such as the von Neumann entropy and separability, it is natural to conjecture that there might be some relationship between separability and the vectors $\lambda(\rho_{AB}), \lambda(\rho_A), \lambda(\rho_B)$ of eigenvalues for ρ_{AB} and the corresponding reduced density matrices. Ma-

jorization suggests the following theorem as a natural way of strengthening the necessary conditions for separability, Equation (11.2):

Theorem 11.0.1: If ρ_{AB} is separable then

$$\lambda(\rho_{AB}) \prec \lambda(\rho_A) \quad \textit{and} \quad \lambda(\rho_{AB}) \prec \lambda(\rho_B). \tag{11.4}$$

By convention we append zeroes to the vectors $\lambda(\rho_A)$ and $\lambda(\rho_B)$ so they have the same dimension as $\lambda(\rho_{AB})$.

Theorem 1 is the main result of this paper. Note that it provides a more stringent criterion for separability than (11.2), since for any two states ρ and σ, $\lambda(\rho) \prec \lambda(\sigma)$ implies that $S(\rho) \geq S(\sigma)$, but not necessarily conversely.
Proof: If ρ_{AB} is separable, it may be written in the form of (11.1). Let $\rho_{AB} = \sum_k r_k |e_k\rangle\langle e_k|$ be a spectral decomposition for ρ_{AB}. By the classification theorem for ensembles (Theorem 2.6 in [43]) it follows that there is a unitary matrix u_{kj} such that

$$\sqrt{r_k}|e_k\rangle \;=\; \sum_j u_{kj}\sqrt{p_j}|\psi_j\rangle|\phi_j\rangle. \tag{11.5}$$

Next we trace out system B in (11.1) to give $\rho_A = \sum_j p_j |\psi_j\rangle\langle\psi_j|$. Letting $\rho_A = \sum_l a_l |f_l\rangle\langle f_l|$ be a spectral decomposition and applying the classification theorem for ensembles we see that there is a unitary matrix v_{jl} such that $\sqrt{p_j}|\psi_j\rangle = \sum_l v_{jl}\sqrt{a_l}|f_l\rangle$. Substituting into (11.5) gives $\sqrt{r_k}|e_k\rangle = \sum_{jl}\sqrt{a_l}u_{kj}v_{jl}|f_l\rangle|\phi_j\rangle$. Multiplying this equation by its adjoint and using the orthonormality of the vectors $|f_l\rangle$ we obtain

$$r_k = \sum_l D_{kl} a_l. \tag{11.6}$$

where

$$D_{kl} \equiv \sum_{j_1 j_2} u^*_{kj_1} u_{kj_2} v^*_{j_1 l} v_{j_2 l} \langle\phi_{j_1}|\phi_{j_2}\rangle. \tag{11.7}$$

To complete the proof all we need to do is show that D_{kl} is doubly stochastic. The fact that $D_{kl} \geq 0$ follows by defining $|\gamma_{kl}\rangle \equiv \sum_j u_{kj}v_{jl}|\phi_j\rangle$ and noting that $D_{kl} = \langle\gamma_{kl}|\gamma_{kl}\rangle \geq 0$. From (11.7) and by the unitarity of u we have

$$\sum_k D_{kl} = \sum_{j_1 j_2} \delta_{j_1 j_2} v^*_{j_1 l} v_{j_2 l} \langle\phi_{j_1}|\phi_{j_2}\rangle = \sum_j v^*_{jl} v_{jl} = 1.$$

Similarly, $\sum_l D_{kl} = 1$, and thus D is a doubly stochastic matrix. ■

The separability criterion (11.4) is strictly stronger than the entropic criterion (11.2). Indeed, for Bell-diagonal states of two qubits, it follows from the positive partial transpose criterion and a straightforward calculation that condition (11.4) is equivalent to separability, whereas as remarked earlier the condition $S(A), S(B) \leq S(A, B)$ is not sufficient to characterize separability even for the more restricted case of Werner states. More generally, the separability criterion (11.4) completely characterizes the separability properties of Werner states in arbitrary (d) dimensions. More precisely, states of the form $\rho_p = p|\Psi\rangle\langle\Psi| + (1-p)/d^2 I$ where $|\Psi\rangle = (|00\rangle + |11\rangle + \ldots + |(d-1)(d-1)\rangle)/\sqrt{d}$ are known to be separable iff $p \leq 1/(d+1)$ [14]. The marginal density matrices of these states are completely mixed and the criterion (11.4) thus becomes

$$\frac{1}{d^2}(1 + (d^2 - 1)p, 1 - p, \ldots, 1 - p) \prec \frac{1}{d}(1, \ldots, 1), \tag{11.8}$$

which is easily seen to be equivalent to $p \leq 1/(d + 1)$.

Another interesting application of the conditions (11.4) is to the problem of finding non-separable states near the completely mixed state $I^{\otimes n}/d^n$ of n qudits (d-dimensional quantum systems). Consider the state $\rho \equiv (1 - \epsilon)I^{\otimes n}/d^n + \epsilon|\psi\rangle\langle\psi|$, where $|\psi\rangle$ is the cat state of n qudits. Partitioning the n qudits so that the first $n - 1$ belong to Alice, and the final qudit to Bob, a straightforward calculation shows that the conditions (11.4) are violated whenever $\epsilon > 1/(1 + d^{n-1})$, and thus ρ must be inseparable when ϵ satisfies this condition. Note that this result has previously been obtained by other techniques [47, 52] (see also [15, 63, 11, 10]), however the utility of the conditions (11.4) is demonstrated in this application by the ease with which they are applied and their generality, as compared to the more complex and state-specific arguments used previously to study the separability of ρ.

It is natural to conjecture the converse to Theorem 1, that if both the conditions in (11.4) hold then ρ_{AB} is separable. Unfortunately, this is not the case, as the following two qubit example shows.
Example 1: Let $\rho^p_{AB} \equiv p|00\rangle\langle 00| + (1 - p)|\Phi\rangle\langle\Phi|$ with the Bell state $|\Phi\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$. Then the partial transpose criterion implies that this state is non-separable whenever $p \neq 1$. However $\lambda(\rho^p_{AB}) = (p, 1 - p) \prec \lambda(\rho^p_{A,B}) = ((1 + p)/2, (1 - p)/2)$ for $1/3 \leq p$, that is, criterion (11.4) is fulfilled for this non-separable state.

More generally, we now show that attempts to characterize separability based only upon the eigenvalue spectra $\lambda(\rho_{AB}), \lambda(\rho_A)$ and $\lambda(\rho_B)$ can never

work. We will demonstrate this by exhibiting a pair of two qubit states ρ_{AB} and σ_{AB} such that all these vectors of eigenvalues are the same (i.e., the states are globally and locally *isospectral*), yet ρ_{AB} is not separable, while σ_{AB} is.
Isospectral Example:

$$\rho_{AB} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \sigma_{AB} = \begin{bmatrix} \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} \end{bmatrix} \tag{11.9}$$

The isospectrality of the these states may be checked by direct calculation, and the fact that ρ_{AB} is non-separable while σ_{AB} is follows from the partial transpose criterion. (Note that similar examples have also been found by Richard Davis (private communication).) It is worth emphasizing how remarkable such examples are: these density matrices have the same spectra, both globally and locally, yet one is separable, while the other is not. This runs counter to the often-encountered wisdom that a complete understanding of a quantum system can be obtained by studying the local and global properties of the spectra of that system. This is the point of view apparently adopted, for instance, in the theory of quantum phase transitions[53], perhaps leading to the disregard of important physical effects in that theory.

Given the isospectral example it is natural to ask under what conditions a separable state exists, given specified global and local spectra. We can report the following result in this direction.
Theorem 2: If ρ_{AB} is a density matrix such that $\lambda(\rho_{AB}) \prec \lambda(\rho_A)$, then there exists a separable density matrix σ_{AB} such that $\lambda(\sigma_{AB}) = \lambda(\rho_{AB})$ and $\lambda(\sigma_A) = \lambda(\rho_A)$.

Proof:

Suppose $(r_j) = \lambda(\rho_{AB})$ and $(s_k) = \lambda(\rho_A)$. By Horn's lemma[21, 41], there is a unitary matrix u_{jk} such that $s_j = \sum_k |u_{jk}|^2 r_k$. Introduce orthonormal bases $|j\rangle$ for system B and $|k\rangle$ for system A, and for each non-zero r_j define

$$|\psi_j\rangle \equiv \frac{\sum_k u_{jk}\sqrt{s_k}|k\rangle}{\sqrt{r_j}}. \tag{11.10}$$

Then define $\sigma \equiv \sum_j r_j |\psi_j\rangle\langle\psi_j| \otimes |j\rangle\langle j|$. Note that σ is manifestly separable with spectrum $\lambda(\rho_{AB})$, while a simple calculation shows that $\mathrm{tr}_B(\sigma) = \sum_k s_k |k\rangle\langle k|$, and thus $\lambda(\sigma_A) = \lambda(\rho_A)$, completing the proof. ■

A stronger conjecture is that whenever *both* $\lambda(\rho_{AB}) \prec \lambda(\rho_A)$ *and* $\lambda(\rho_B)$, then there exists a separable state σ_{AB} which is isospectral to ρ_{AB}. Unfortunately, the following theorem shows that this is not true.

Theorem 3: For the class of states ρ^p_{AB} in Example 1 (which are non-separable when $1 > p > 1/3$) the separability conditions (11.4) are fulfilled yet there is no separable σ_{AB} (globally and locally) isospectral to ρ^p_{AB} when $1 > p \geq 1/2$.

Proof: Suppose $\sigma \equiv \sigma_{AB}$ is a separable state isospectral to ρ^p_{AB}. Then $\sigma = p|s_1\rangle\langle s_1| + (1-p)|s_2\rangle\langle s_2|$ for orthonormal states $|s_1\rangle$ and $|s_2\rangle$. We suppose for now that σ can be given a separable decomposition with only two terms, $\sigma = q|a_1\rangle\langle a_1| \otimes |b_1\rangle\langle b_1| + (1-q)|a_2\rangle\langle a_2| \otimes |b_2\rangle\langle b_2|$. We show later that this is the only case that need be considered. Define angles α, β and ϕ by $|\langle a_1|b_1\rangle| \equiv \cos(\alpha); |\langle a_2|b_2\rangle| \equiv \cos(\beta); \cos(\phi) \equiv \cos(\alpha)\cos(\beta)$. Then the global and local spectra for σ are easily calculated,

$$\lambda(\sigma_{AB}) = \left(\frac{1 \pm \sqrt{1 - 4q(1-q)\sin^2(\phi)}}{2} \right), \tag{11.11}$$

with similar expressions for $\lambda(\sigma_A)$ and $\lambda(\sigma_B)$, with α and β appearing in place of ϕ. Assuming $1/2 \leq p$ this gives $\sin^2(\alpha) = \sin^2(\beta) = (1-p^2)/4q(1-q)$ and $p(1-p) = q(1-q)\sin^2(\phi)$. Using $\sin^2(\phi) = 1 - (1 - \sin^2(\alpha))(1 - \sin^2(\beta))$ to substitute the former expression into the latter, we find $q(1-q) = (1+p)^2/8$. For $p > \sqrt{2}-1 \approx 0.41$ there is no q in the range 0 to 1 satisfying this equation, so we deduce that no such separable state σ can exist.

To complete the proof we show that any separable decomposition $\sigma = \sum_j q_j |a_j\rangle\langle a_j| \otimes |b_j\rangle\langle b_j|$ can be assumed to have two terms. Without loss of generality we assume that there is no redundancy in the decomposition, that is, there do not exist values $j \neq k$ such that $|a_j\rangle|b_j\rangle = |a_k\rangle|b_k\rangle$ (up to phase). We show that assuming the decomposition has three or more terms leads to a contradiction. Note that the decomposition must contain contributions from at least two linearly independent states, say $|a_1\rangle|b_1\rangle$ and $|a_2\rangle|b_2\rangle$. Furthermore, because $\mathrm{rank}(\sigma) = 2$ any other state in the sum must be a linear combination of these two states, $|a_j\rangle|b_j\rangle = \alpha_j|a_1\rangle|b_1\rangle + \beta_j|a_2\rangle|b_2\rangle$. By the no-redundancy assumption neither $|\alpha_j| = 1$ nor $|\beta_j| = 1$, so we must have $0 < |\alpha_j|, |\beta_j| < 1$. Consider now three possible cases. In the first case, $|a_1\rangle = |a_2\rangle$ (up to phase), in which case $|a_j\rangle = |a_1\rangle$ (up to phase) for all j, and thus $\lambda(\sigma_A) = (1,0) \neq \lambda(\rho^p_A)$, a contradiction. A similar contradiction arises when $|b_1\rangle = |b_2\rangle$ up to phase. The third and final case is when neither $|a_1\rangle = |a_2\rangle$ nor $|b_1\rangle = |b_2\rangle$ up to phase. In this case $\alpha_j|a_1\rangle|b_1\rangle + \beta_j|a_2\rangle|b_2\rangle$

cannot be a product state, a contradiction. ■

Given that attempts to characterize separability based on the local and global spectra are doomed to failure, it is still interesting to ask whether the conditions $\lambda(\rho_{AB}) \prec \lambda(\rho_A)$ and $\lambda_(\rho_{AB}) \prec \lambda(\rho_B)$ are equivalent to some other interesting physical condition? We have tried to find such an equivalence, with little success, but can identify several plausible possibilities which these conditions are *not* equivalent to. They are not equivalent to the property of violating a Bell inequality, of having a positive partial transpose, or of being distillable. Another interesting idea is to find states which have positive partial transposition, but which violate (11.4). Such a state will necessarily be bound-entangled [24]. We have not yet identified any such states, despite searching through several of the known classes of bound-entangled states, and doing numerical searches.

In summary, we have connected two central notions in the theory of entanglement, using majorization to obtain a simple set of necessary conditions for a state to be separable in arbitrary dimensions. Understanding the physical import of these conditions and their relationship to criteria such as the positive partial transpose conditon remains an interesting problem for further research.

Problems for Lecture 11

Problem 11.0.1: Let p_j be probabilities and $|\psi_j\rangle$ be states of some system A. Choose B to be a system with orthonormal basis $|j\rangle$ labelled by the index j for the states $|\psi_j\rangle$, and define the separable state

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| \otimes |j\rangle\langle j|. \tag{11.12}$$

Apply the criterion of Theorem 11.0.1 to the state ρ to conclude that $(p_j) \prec \lambda\left(\sum_j p_j |\psi_j\rangle\langle\psi_j|\right)$.

Chapter 12

Open problems

Exercise 12.0.1: (**Entanglement transformation with no communication**) Suppose Alice and Bob can convert $|\psi\rangle$ to $|\phi\rangle$ by local operations with *no* classical communication. Show that a necessary and sufficient condition for this to be possible is that there exist a vector r such that $q \otimes r = p$. Describe an algorithm to check whether such an r exists. Find an example of entanglement conversion possible with local operations and classical communication that is not possible without classical communication.

Now is an exciting time to be working on majorization and its applications to quantum information theory, because there are so many interesting open problems to be solved. In the next few sections I will describe some of the problems, conjectures as to the nature of their solution, and describe some partial progress I have made. Please contribute by supplying solutions (preferably complete, but I'm not picky), and suggestions for more problems, or other ideas for proof-techniques. The main line of thought concerns problems related to entanglement transformation.

12.0.1 Entanglement catalysis

Let's start with *entanglement catalysis*. Jonathan and Plenio[30] have discovered an interesting procedure for doing entanglement transformations. The idea is that there exist pure states $|\psi\rangle$ and $|\phi\rangle$ such that neither $|\psi\rangle \rightarrow |\phi\rangle$ nor $|\phi\rangle \rightarrow |\psi\rangle$, but there may be a *catalysing state* $|l\rangle$ such that $|\psi\rangle|l\rangle \rightarrow |\phi\rangle|l\rangle$. As an explicit example of this, Jonathan and Plenio offer as an example the

states

$$|\psi\rangle = \sqrt{0.4}|00\rangle + \sqrt{0.4}|11\rangle + \sqrt{0.1}|22\rangle + \sqrt{0.1}|33\rangle \quad (12.1)$$
$$|\phi\rangle = \sqrt{0.5}|00\rangle + \sqrt{0.25}|11\rangle + \sqrt{0.25}|11\rangle \quad (12.2)$$
$$|l\rangle = \sqrt{0.6}|00\rangle + \sqrt{0.4}|11\rangle. \quad (12.3)$$

It is easy to check the appropriate majorization conditions and verify that it is possible to make the transformation $|\psi\rangle|l\rangle \rightarrow |\phi\rangle|l\rangle$.

Problems for Lecture 12

Problem 12.0.2: (**37**) What are necessary and sufficient conditions on $|\psi\rangle$ and $|\phi\rangle$ for the existence of a state $|l\rangle$ catalysing the transformation $|\psi\rangle \rightarrow |\phi\rangle$?

Another way of stating this problem is in purely mathematical terms,

Problems for Lecture 12

Problem 12.0.3: (**37**) What are necessary and sufficient conditions on real vectors x and y such that there exists a real vector z with $x \otimes z \prec y \otimes z$?

I have not solved this problem, but I can report some partial results and ideas about how to approach the problem that may yield fruit. What we are looking to characterize is a new partial order, $\prec_T$, on vectors x and y; we say $x \prec_T y$ if there exists z such that $x \prec_T y$. We say that x is *trumped* by y, with z as catalyst. Equivalently, y *trumps* x, with z as catalyst.

Exercise 12.0.2: (**Jonathan and Plenio (1999)**) Show that $x \prec_T y$ and $y \prec_T x$ if and only if $x^{\downarrow} = y^{\downarrow}$.

Additive Schur-convex functions and catalysis

One approach to the catalysis problem is via the theory of Schur-convex functions. This gives rise to some necessary conditions for $x \prec_T y$:

Theorem 12.0.2: Suppose $f(\cdot)$ is a family of Schur-convex functions that is *additive* in the sense that $f(x \otimes z) = f(x) + f(z)$. Then $x \prec_T y$ implies $f(x) \leq f(y)$.

Proof:

Suppose $x \prec_T y$. Then there exists z such that $x \otimes z \prec y \otimes z$. Since f is Schur-convex, we obtain

$$f(x \otimes z) \leq f(y \otimes z), \tag{12.4}$$

which by additivity is equivalent to

$$f(x) + f(z) \leq f(y) + f(z). \tag{12.5}$$

Thus $f(x) \leq f(y)$.

■

There are many examples of additive Schur-convex functions: minus the Shannon entropy $-H(x)$, minus the Renyi entropies $\log(\sum_i x_i^k)$ for $k \geq 1$, the log of the largest eigenvalue, minus the log of the smallest eigenvalue, and the product of eigenvalues. All of these families give rise to interesting constraints on the trumping relation. Suppose that $x \prec_T y$. Then these additive Schur-convex functions give rise to the following constraints on x and y:

$$\begin{aligned}
H(x) &\geq H(y) && (12.6)\\
\sum_{i=1}^{d} x_i^k &\leq \sum_{i=1}^{d} y_i^k \quad \text{for } k \geq 1 && (12.7)\\
x_1^{\downarrow} &\leq y_1^{\downarrow} && (12.8)\\
x_d^{\downarrow} &\geq y_d^{\downarrow} && (12.9)\\
\prod_{i=1}^{d} x_i &\geq \prod_{i=1}^{d} y_i. && (12.10)
\end{aligned}$$

I think it is reasonably likely that there is a simple family of additive Schur-convex functions whose monotonicity provides a simple set of necessary and sufficient conditions for the trumping relation, much as the family of functions $\sum_i |x_i - t|$ characterizes majorization. One might take many different approaches to this problem. For example: **Problems for Lecture 12**

Problem 12.0.4: Prove that $x \prec_T y$ if and only if $f(x) \leq f(y)$ for all families of additive Schur-convex functions f.

Convexity and catalysis

Another approach to the catalysis problem is to try to study the convex structure of the problem. My guess, based on the partial results reported below, is that this will ultimately be the approarch that is most fruitful.

Recall that the set $S(y)$ of points x such that $x \prec y$ is just the convex hull of all vectors of which may be obtained by permutation of the components of y. One approach to the problem of entanglement catalysis is to study the set $T(y)$ of vectors x such that $x \prec_T y$. The following result shows that this set is also convex.

> *Theorem 12.0.3*:
>
> Let $T(y)$ denote the set of vectors x such that $x \prec_T y$. Then $T(y)$ is a convex set.

Proof:

Let $0 \leq p \leq 1$, and $x_1, x_2 \in T(y)$. We will show that $\tilde{x} \equiv px_1 + (1-p)x_2$ is also an element of $T(y)$. To do this, let z_1 and z_2 be catalysts for x_1 and x_2 respectively, and let D_1 and D_2 be doubly stochastic matrices such that $D_i(y \otimes z_i) = x_i$ for $i = 1, 2$. We claim that $z_1 \otimes z_2$ is a catalyst for $\tilde{x}$. To see this, define $\tilde{D} \equiv p\tilde{D}_1 + (1-p)\tilde{D}_2$, where it is understood that D_1 acts non-trivially on the first and second terms in the tensor product, and D_2 acts non-trivially on the first and third terms in the tensor product, so $\tilde{D}(y \otimes z_1 \otimes z_2) = px_1 \otimes z_1 \otimes z_2 + (1-p)x_2 \otimes z_1 \otimes z_2 = \tilde{x} \otimes z_1 \otimes z_2$, and thus $\tilde{x} \in T(y)$.

■

Exercise 12.0.3: Show that $S(y) \subseteq T(y)$.

Exercise 12.0.4: Show that y is a boundary point of $T(y)$.

Problems for Lecture 12

Problem 12.0.5: (**35-40**) What are the extreme points of the set $T(y)$?

I can report some rather curious and perhaps surprising progress towards solution of this problem. It relies on the result that:

> *Lemma 12.0.4*: Suppose $x_i \prec y$, and p_i is a probability distribution with no zero entries. Then if $y = \sum_i p_i x_i$, it follows that $x_i = y$ for all i.

Proof:

y is extremal in $S(y)$, and the x_i are elements of $S(y)$. Thus $x_i = y$ for all i.

■

> *Theorem 12.0.5*: y and permutations of y are extreme points of $T(y)$.

Proof:

Suppose $y = px_1 + (1-p)x_2$ is a convex combination of points x_1 and x_2 in $T(y)$. Let z_1 and z_2 be the corresponding catalysts, and D_1 and D_2 be doubly stochastic matrices chosen so that $D_i(y \otimes z_i) = x_i \otimes z_i$. Then

$$y \otimes z_1 \otimes z_2 = [(px_1 + (1-p)x_2] \otimes z_1 \otimes z_2. \tag{12.11}$$

By the lemma, $y \otimes z_1 \otimes z_2 = x_1 \otimes z_1 \otimes z_2$, from which we conclude that $x_1 = y$. Similarly, $x_2 = y$, and we see that y is extremal. Similarly, permutations of y are also extreme points oy $T(y)$.

■

What this result implies is that $T(y)$ must have a more complicated geometric structure than $S(y)$, since the extreme points of $T(y)$ contain (sometimes strictly, otherwise $\prec$ would be the same as $\prec_T$) the extreme points of $S(y)$.

As an example of convex sets with a similar property, consider a square inscribed on a circle. The set of points C inside the circle is convex, as is the set of points inside the square, S. The extreme points of C are all the points on the circle, a continuum of points. The extreme points of the square are the four corner points, a strict (indeed finite) subset of the extreme points of the circle.

I conjecture that $S(y)$ and $T(y)$ are related in a similar way. $S(y)$ has only a finite number of extreme points. Intuitively, the reason this is so is because $S(y)$ is formed by intersecting a finite number of half-plans, corresponding

to the finite set of inequalities that have to be satisified. (Strictly speaking, there is an ambiguity because of the ordering in the inequalities, but there are only a finite number of orderings as well.) By contrast, there is an infinite set of inequalities, corresponding to the continuum of possible catalysts, that may potentially be checked in order to determine whether $x \prec_T y$. For this reason I ask that you prove the following conjecture: **Problems for**

Lecture 12

Problem 12.0.6: Show that $T(y)$ has a continuum of extreme points.

Computational approach to convexity

In order to attack problems on the convex structures associated with $\prec_T$, it may help to use some of the ideas of computational geometry. An introduction to these ideas and further references are given in the book by Preparata and Shamos [49], which you can borrow from me. In particular, there are algorithms in that book which allow one to numerically find the convex hull of a set of points.

One way of doing so is to look at the set of points $T(y, z)$, defined to be the set x such that $x \otimes z \prec y \otimes z$. It is not difficult to prove that $T(y, z)$ is convex. Moreover, $T(y, z)$ is quite amenable to numerical study. By fixing y and z, and then doing a search for x such that $x \otimes z \prec y \otimes z$, we can numerically characterize $T(y, z)$. We can algorithmically determine the extremal points of $T(y, z)$. Hopefully, by inspection of the extremal points it will become possible to discern some structure in the etxremal points of $T(y, z)$.

Other computational questions

We can generate computational questions related to catalysis pretty much *ad infinitum*, and attempt to answer them. How much power does it add to add extra dimensions to the catalyst?

Double stochasticity

Another approach to the problem of catalysis is via the connection with double stachasticity: $x \prec_T y$ if and only if there exists z and a doubly

stochastic D such that $x \otimes z = d(y \otimes z)$. Unfortunately, I don't have any really good ideas on how to exploit this approach. Instead of listing good ideas, I'll list a few half-baked ones.

First, one can try to specialize. What does it say if we restrict D to be ortho-stochastic? Unitary-stochastic? A product of T-transforms? What if we specialize further, and look at the actual structure of the relevant proofs. One apporach which I think has some promise is to use the proof that it is sufficient to consider products of T-transforms. There's actually quite a bit of structure in the sequence of T-transforms generated by that proof, which might be exploitable here. (I have a sneaking suspicion that conjugating the sequence of T-transforms in just the right way might lead to a substantial simplification of the sequence of T-transforms on the tensor product. No luck yet, however.)

Another approach is to try looking at environmental models for doubly stochastic maps. This might help by reducing the problem more to being one about permutations on product spaces — the study of which would also be a direction in which to head.

Miscellaneous questions

There's a bundle of other open problems that don't really fit in any of the categories we've looked at so far. I've collected them here, in a more or less disorganized fashion. Hopefully, as we solve the problems a better structure to lay all this out will become apparent.

Probably my favourite open problem is the following[1]:

Problems for Lecture 12

Problem 12.0.7: A vector z with non-negative real components summing to 1 is said to be *non-uniform* if it has two non-zero components that are not equal. I conjecture that if z is any non-uniform vector then there exist x and y such that $x \not\prec y$, but $x \otimes z \prec y \otimes z$.

I think it very likely that this conjecture is true. What's better, I also think it's very important if true, and probably not too difficult to prove! This is why I like the problem.

[1] This conjecture has recently been proved. (Daftuar *et al* (to appear).)

The following two easy exercises demonstrate some related applications of the notion of uniformity.

Exercise 12.0.5: Suppose $x \prec_T y$, and x is uniform. Show that $x \prec y$.

Exercise 12.0.6: Suppose $x \prec_T y$, and y is uniform. Show that $x \prec y$.

Another problem is[2]: **Problems for Lecture 12**

Problem 12.0.8: Suppose the primary system is d dimensional. Is it possible to put an upper bound on the size of the Hilbert space for the catalyst?

A good way of getting insight into this problem is to do numerical searches. My own bet is that if there is a bound, it is d or d^2 dimensions. (However, my thinking has shifted, and I now think it likely that more dimensions will always give more catalysts that catalyse transformations not previously possible. It would be very exciting to find a four dimensional example which needed a five or seventeen dimensional catalyst, as such an example would rule out the d and d^2 conjectures.)

12.0.2 Entanglement banking

So far, I haven't had any luck solving the problem of when entanglement catalysis is possible. However, it may be possible to make progress by using other problems to bridge the gap between what we know of majorization and the catalysis problem.

A useful metaphor for generating other problems is a financial metaphor for entanglement catalysis. Imagine that Alice and Bob share a state $|\psi\rangle$ which they wish to turn into another state $|\phi\rangle$, by local operations and classical communication. To do so, they go to Entanglement Banking Corporation, and ask for the loan of an entangled state $|l\rangle$. In standard entanglement catalysis, they must repay the loan (with no interest) after performing the transformation $|\psi\rangle|l\rangle \rightarrow |\phi\rangle|l\rangle$.

Is it possible to perform the transformation in such a way that Alice and Bob *pay interest* to the bank, in the sense that they make the transformation $|\psi\rangle|l\rangle \rightarrow |\phi\rangle|r\rangle$, and the "repayment" $|r\rangle$ is more entangled than the loan $|l\rangle$ in the sense that $|r\rangle \rightarrow |l\rangle$ but not vice-versa.

[2]This problem has recently been settled in the negative by Klimesh *et al* (to appear).

12.0.3 Other directions

We've mostly so far been focused fairly inward on the problem of entanglement transformation. Let's look outward.

Vidal has solved the following problem: given $|\psi\rangle$, what is the maximum probablitiy of obtaining a state $|\phi\rangle$ for $|\psi\rangle$, by local operations and classical communication? Not surprisingly, the answer involves majorization. It also gives rise to numerous interesting problems: **Problems for Lecture 12**

Problem 12.0.9: (**17**) Can you find states $|\psi\rangle$ and $|\phi\rangle$ of entangled qubits, and a catalysing state $|c\rangle$ such that the probability of being able to transform from $|\psi\rangle$ to $|\phi\rangle$ is enhanced by the presence of the catalyst?

Wootters and collaborators [68, 20, 6] have studied the problem of the *entanglement of formation* — how many Bell states are needed to generate many copies of an entangled mixed state. It is possible that the results of Jonathan and Plenio can be used to get some insight into this work.

Appendix A

Birkhoff's theorem

Birkhoff's theorem is a structure theorem characterizing the extremal points of the convex set of doubly stochastic matrices. It plays a role in the theory of double stochasticity analogous, for example, to the spectral theorem in the theory of Hermitian matrices. In other words, it's a very powerful representation theorem!

The statement of Birkhoff's theorem is very simple. It says that any d by d doubly stochastic matrix, D, can be represented as a convex combination of d by d permutation matrices. That is, $D = \sum_j p_j P_j$, for some set of probabilities, p_j, and corresponding permutation matrices, P_j. Conversely, any convex combination of permutation matrices is doubly stochastic. Thus, Birkhoff's theorem provides us with a way of representing doubly stochastic matrices in terms of objects that, *a priori*, are much simpler to deal with, namely, probability distributions and permutation matrices.

One reason for interest in Birkhoff's theorem is Theorem 3.1.2, on page 20, which implies that $r \prec s$ if and only if there exists a doubly stochastic matrix D such that $r = Ds$.

Note that in the statement and proof of Theorem 3.1.2, not only did we prove that $r \prec s$ if and only if $r = \sum_j p_j P_j s$, we also proved that $r \prec s$ if and only if $r = Ds$ for some doubly stochastic matrix D. Thus, in some sense, applying Birkhoff's theorem in the context of Theorem 3.1.2 does not add anything immediately to our knowledge of majorization. We will come back to this point below.

You might reasonably wonder whether or not we can deduce Birkhoff's theorem from Theorem 3.1.2. We can certainly deduce that, given a doubly stochastic matrix, D, and for any *particular* vector, s, there exist probabilities

p_j and permutation matrices P_j such that $Ds = \sum_j p_j P_j s$. This does not mean, however, that $Ds = \sum_j p_j P_j s$ for *all* vectors, s. Unfortunately, so far as I am aware, there is not any direct way of going from a proof of Theorem 3.1.2 to Birkhoff's theorem. (Although finding such a path might make a nice research problem!)

One might ask why we need Birkhoff's theorem at all, given that it does not appear to add anything to our knowledge of majorization beyond what we already know from Theorem 3.1.2? From a utilarian point of view, there is some truth to this point of view. In the main text, all we will ever need are the results of Theorem 3.1.2. Indeed, this utilarian view is why the proof of Birkhoff's theorem is in an appendix, and not the main text! Nonetheless, from the point of view of deepening our understanding of *why* Theorem 3.1.2 is true, Birkhoff's theorem serves a valuable purpose. Furthermore, the ideas used in the proof of Birkhoff's theorem are beautiful, useful, and stimulate many other interesting questions and connections, both within the theory of majorization, and in other areas of mathematics, making it worthwhile to spend time in the study of the proof. In particular, in Chapter 3 Birkhoff's theorem will stimulate us to ask for an analogous quantum result, while in Chapter 7 we will see how Birkhoff's theorem enables us to make some powerful statements about a concept related to majorization known as *sub-majorization*.

Birkhoff's theorem is not trivial to prove, and our route to the proof is somewhat indirect. We begin with a combinatorial result known as *Hall's theorem* in Section A.1. In Section A.2 we use Hall's theorem to prove an analogous theorem about matrices, known as the *König-Frobenius* theorem. Finally, in Section A.3 we prove Birkhoff's theorem.

A.1 The marriage problem and Hall's theorem

To prove Birkhoff's theorem we're going to take a route through a problem in combinatorics, known variously as the *marriage problem* or as the *matching problem*. The marriage problem involves two equally sized and finite sets B and G of "boys" and "girls" respectively, and a relation R on $B \times G$. You can think of this relation as representing the fact that boy b and girl g love each other if $R(b, g)$ is true, and don't if it is false. (There is no unrequited

love in the marriage problem.) The marriage problem is to determine when it is possible to marry every boy to a single girl in such a way that no girl has more than one husband, and so that each married boy and girl love one another. Such a scheme, if it exists, is called a *compatible matching* for $B \times G$ and R.

We are going to prove Hall's theorem, which completely solves the marriage problem. Perhaps surprisingly, Hall's theorem also gives rise to a simple proof of Birkhoff's theorem.

> *Theorem A.1.1*: (**(Hall's theorem)**)
> There exists a compatible matching for $B \times G$ and R if and only if each group of k boys loves at least k girls, for $k = 1, \ldots, |B|$.

Proof: The forward implication is clear.

To prove the reverse implication, we induct on $d = |B|$. The case $d = 1$ is obvious, so we assume the result is true up to $|B| = d$, and try to prove it for $|B| = d + 1$. We split the analysis into two cases.

Case (a): There exists k such that $1 \leq k \leq d$, and a group β of d boys that loves a group of exactly d girls, γ. By the inductive hypothesis, β and γ can be compatibly matched. We will use the inductive hypothesis to show that the complementary sets β^c and γ^c can also be compatibly matched, and thus B can be compatibly matched with G. Let S be a subset of β^c containing h members. The set $\beta \cup S$ of $k + h$ boys must love at least $k + h$ girs, and thus the boys in S must love at least h girls in γ^c. The inductive hypothesis implies that β^c and γ^c can be compatibly matched.

Case (b): All groups of k boys ($1 \leq k \leq d$) love at least $k+1$ girls. Pick any boy-girl pair and marry them off. Then the remaining d boys and d girls satisfy the inductive hypothesis, and thus may be compatibly matched. ■

Exercise A.1.1: (**Algorithm for the marriage problem**) Directly checking the conditions of Hall's theorem requires that we check $2^{|B|}$ subsets of B, and thus is inefficient. Find an efficient algorithm to solve the marriage problem, that is, an algorithm which requires a number of operations polynomial in $|B|$ to find a compatible matching, or to demonstrate that no such matching exists.

Exercise A.1.2: (**Generalized Hall's theorem**) Suppose $B \times G$ and R specify a marriage problem. Fix a number s in the range 0 through n. Show that a compatible matching exists if and only if every group of

$k \leq s$ boys loves at least k girls, and every group of $k \leq |B| - s$ girls loves at least k boys.

A.2 The König-Frobenius theorem

To apply Hall's theorem to the study of doubly stochastic matrices, we first need to translate it into a matrix form. One way of achieving this it the *König-Frobenius theorem*. To state this theorem, we define a *diagonal* of a d by d matrix A to be the vector $(A_{1\pi(1)}, A_{2\pi(2)}, \ldots, A_{d\pi(d)})$, where π is some permutation of $1, \ldots, d$, and A_{jk} is the (j,k)th element of A.

> *Theorem A.2.1*: (**König-Frobenius theorem**)
> There exists a diagonal of a d by d matrix A with no zero elements if and only if every l by m zero submatrix of A satisfies $l+m \leq d$.

Proof: The proof is a simple application of Hall's theorem. We identify the set of boys with the rows of A, and the girls with the columns of A. By definition, boy j and girl k love one another if and only if $A_{jk} \neq 0$. With these definitions, a potential matching corresponds to a diagonal of A, and the matching is compatible if and only if there are no zero elements on that diagonal. Note also that a l by m zero submatrix corresponds to a group of l boys loving at most $n - m$ girls.

Applying Hall's theorem and these facts, there exists a diagonal of A with no zero elements iff a compatible matching exists iff every group of l boys loves at least l girls iff every l by $m - l$ zero submatrix satisfies $l \leq d - m$, that is, $l + m \leq d$. ■

A.3 Birkhoff's theorem

We now have the main technical ingredients needed to prove Birkhoff's theorem. Before stating Birkhoff's theorem formally, it is helpful to note a couple of other facts. First, it is not difficult to show that the set of doubly stochastic matrices is convex, that is, convex combinations of doubly stochastic matrices are also doubly stochastic. Second, the permutation matrices are extreme points of the set of doubly stochastic matrices. This means that (a) all permutation matrices are doubly stochastic, and (b) a permutation matrix

can't be written as a convex combination of two distinct doubly stochastic matrices. Thus, a permutation matrix is "on the edge" of the convex set of doubly stochastic matrices, which is why it is called an extreme point. The proof of these facts is left as an exercise for the reader.

Exercise A.3.1: (**Convexity of the doubly stochastic matrices**) Show that the set of d by d doubly stochastic matrices is convex.

Exercise A.3.2: Show that permutation matrices are extreme points of the set of doubly stochastic matrices, that is, if P is a permutation matrix, then it is a doubly stochastic matrix, but it is not a convex combination of two distinct doubly stochastic matrices.

Theorem A.3.1: (**Birkhoff's theorem (Birkhoff 1946 [9])**) The set of d by d doubly stochastic matrices is a convex set whose extreme points are the permutation matrices.

Proof: You have already demonstrated in the exercises that the permutation matrices are extreme points of the convex set of doubly stochastic matrices. What remains to be shown is that any doubly stochastic matrix D can be written as a convex combination of permutation matrices,

$$D = \sum_j p_j P_j. \tag{A.1}$$

We will prove this by induction on $n(D)$, which is defined to be the number of non-zero elements in D. Note that $d \leq n(D)$, since every column must contain at least one non-zero entry. For the case $n(D) = d$, D must have a single 1 in each row and column, and thus is a permutation matrix, establishing the result for $n(D) = d$.

Next we do the inductive step. Let $s(D)$ denote the sum of all elements in D, which must be the sum of all the row sums, so double stochasticity implies that $s(D) = d$. Suppose D has a l by m zero sub-matrix. Then the sum of all the elements in D must be greater than or equal to the sum of the elements in the l rows corresponding to the zero submatrix, plus the m rows corresponding to the zero submatrix, since no non-zero element in D is counted more than once in this sum. Thus $l+m \leq d$, so the König-Frobenius theroem implies that there exists a diagonal of D with no non-zero elements.

Let p be the smallest element on the diagonal, and P the permutation matrix with ones on the diagonal. There are two cases to consider.

Case (a): $p = 1$. Then D must be a permutation matrix and we are done.

Case (b): $0 < p < 1$. Define

$$Q \equiv \frac{D - pP}{1 - p}. \tag{A.2}$$

Then $D = (1-p)Q+pP$. Clearly Q is doubly stochastic and $n(Q) \leq n(D)-1$. By the inductive hypothesis

$$Q = \sum_j p_j P_j \tag{A.3}$$

for some set of probabilities p_j and permutation matrices P_j, from which it follows that

$$D \sum_j (1-p) p_j P_j + pP, \tag{A.4}$$

so D is a convex combination of permutation matrices, as required. ■

As a simple example of Birkhoff's theorem we can express a 2 by 2 doubly stochastic matrix as a convex combination of permutations:

$$\begin{bmatrix} t & 1-t \\ 1-t & t \end{bmatrix} = t \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (1-t) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{A.5}$$

More generally, Carathéodory's theorem [51] guarantees that a point in an m-dimensional compact convex set may be expressed as a convex combination of at most $m + 1$ extremal points of that set. The d by d doubly stochastic matrices form a d^2-2d+1-dimensional set, so an arbitrary doubly stochastic matrix may be expressed as a convex combination of at most $d^2 - 2d + 2$ permutation matrices.

Exercise A.3.3: Find an algorithm which decomposes a d by d doubly stochastic matrix into a convex combination of permutation matrices, taking time no more than polynomial in d.

Exercise A.3.4: An m by n real matrix is said to be *column stochastic* if it has non-negative entries and all the columns sum to one. A special class

of column stochastic matrices is the *deterministic* matrices, which have a single one in each column and are zero elsewhere. (The nomenclature stems from the interpretation of the matrix as a noisy channel — a deterministic matrix corresponds to a noisy channel whose action is deterministic.) Prove that the set of m by n column stochastic matrices is convex with extremal points the deterministic matrices.

Appendix B

Generalized measurements and quantum operations

This appendix has two closely related purposes. The first purpose is to describe an approach to quantum measurements that is more general than the projective measurements taught in most introductory classes on quantum mechanics. The second purpose is to describe the quantum operations formalism, a general formalism that can be used to describe a very large class of quantum dynamics. The processes that can be described using quantum operations include the unitary evolution and projective measurement processes described in introductory courses. They also include generalized measurements, as well as noise processes, like spontaneous emmission, that occur when a quantum system is coupled to its environment in an uncontrolled way. More detailed introductions to the theory of generalized measurements and of quantum operations may be found in [43, 32, 18].

The appendix begins in Section B.1 with a discussion of the need for generalized measurements. Section B.2 describes in detail an especially important class of generalized measurements known as *ideal* generalized measurements. We conclude in Section B.3 with a discussion of the quantum operations formalism. This formalism generalizes in a natural way the earlier discussion of ideal generalized measurements. Also in Section B.3 we explain how the quantum operations formalism provides a general approach to both quantum measurements and quantum noise processes.

B.1 The need for generalized measurements

There is a standard approach to quantum measurements taught as part of most undergraduate classes on quantum mechanics. According to this approach, a quantum measurement is described by a set of operators P_j acting on the state space of the system being measured, with the subscript j indexing the possible measurement outcomes. The only restrictions put on the operators P_j are that: (a) they should be *projectors*, that is, P_j should be Hermitian, with $P_j^2 = P_j$; and (b) the P_j should satisfy the completeness relation $\sum_j P_j = I$. If the quantum system is in the state $|\psi\rangle$ immediately prior to the measurement, then the outcome j is obtained with probability

$$\Pr(j) = \langle\psi|P_j|\psi\rangle, \tag{B.1}$$

and the posterior state of the system after the measurement is

$$|\psi'_j\rangle = \frac{P_j|\psi\rangle}{\sqrt{\langle\psi|P_j|\psi\rangle}}. \tag{B.2}$$

Note that the completeness relation $\sum_j P_j = I$ ensures that the probabilities $\Pr(j)$ sum to one, as we would expect.

Exercise B.1.1: Suppose the projectors P_j describe a quantum measurement process. Suppose we use that process to measure a quantum state $|psi\rangle$, and then repeat the process again immediately after the first measurement is complete. Show that the outcome of the second measurement will, with probability one, be the same as the outcome of the first measurement.

This description of quantum measurements works extremely well in many situations. A standard textbook example is a spin-$\frac{1}{2}$ system, such as an electron, with basis states $|\uparrow\rangle$ and $|\downarrow\rangle$ corresponding to spin up and spin down in the z direction, respectively. Then a measurement of the spin in the z direction is well described by the projectors $P_\uparrow = |\uparrow\rangle\langle\uparrow|, P_\downarrow = |\downarrow\rangle\langle\downarrow|$.

Unfortunately, there are also many situations in which this formalism does not describe the effect of a quantum measurement on a system. For example, measuring the number of photons in a particular mode of an electromagnetic cavity usually involves coupling the cavity mode to modes of the electromagnetic field external to the cavity, which in turn couple to a

photodetector. This photodetector determines the number of photons inside the cavity by absorbing some of the energy transferred from the cavity mode to the external modes. Thus, to function effectively, the entire process must result in the photon number in the cavity decreasing. It turns out that this process is not well-described as a projective measurement on the cavity mode. One way of seeing this is to note that the procedure is obviously not repeatable, in the sense of Exercise B.1, and thus cannot be described by a projective measurement on the cavity mode.

Of course, we could describe the measurement as a projective measurement on the combined system containing the cavity mode, the modes of the electromagnetic field external to the cavity, and the photodetector. However, if all we're interested in is the state of the cavity mode, that description seems like overkill. The generalized measurement formalism, and more generally, the quantum operations formalism, provides an elegant way of describing the effect of the measurement on the cavity mode alone.

B.2 Ideal generalized measurements

We begin by introducing a particular type of generalized measurement, the *ideal* generalized measurements. The theory of non-ideal generalized measurements is easily understood once the ideal case has been mastered. For this reason, when we refer to generalized measurements in this section, we really mean ideal generalized measurements. Our approach is to begin by explaining the mathematical formalism of (ideal) generalized measurements, and then to explain how that formalism can be uderstood in terms of the von Neumann projective measurements.

Mathematically, a generalized measurement is specified by a set $\{E_j\}$ of *measurement matrices* satisfying the *completeness relation* $\sum_j E_j^\dagger E_j = I$. The index j on the measurement matrices is in one-to-one correspondence with the possible measurement outcomes. The measurement matrices play a role in the theory of generalized measurements analogous to the role played by the projectors P_j in the von Neumann formalism.

The rule used to connect the measurement matrices to physics is that if the prior state of the quantum system is ρ then the outcome j occurs with probability

$$\Pr(j) = \mathrm{tr}(E_j \rho E_j^\dagger), \tag{B.3}$$

and the posterior state is given by

$$\rho'_j = \frac{E_j \rho E_j^\dagger}{\mathrm{tr}(E_j \rho E_j^\dagger)}. \tag{B.4}$$

Projective measurements are obviously a special case of generalized measurements, corresponding to the case when each of the measurement matrices $E_j = P_j$ is a projector: the completeness relation $\sum_j E_j^\dagger E_j = I$ for measurement matrices is equivalenet to the completeness relation $\sum_j P_j = I$ for projectors.

However, generalized measurements also enable us to give simple descriptions of measurements that are rather awkward to describe in standard quantum mechanics. Consider, for example, a measurement on a single qubit described by measurement matrices

$$E_1 = \frac{1}{\sqrt{1+\sqrt{1/2}}} |0\rangle\langle 0|; \tag{B.5}$$

$$E_2 = \frac{1}{2\sqrt{1+\sqrt{1/2}}} (|0\rangle + |1\rangle)(\langle 0| + \langle 1|; \tag{B.6}$$

$$E_3 = \sqrt{I - E_1^\dagger E_1 - E_2^\dagger E_2}, \tag{B.7}$$

where E_3 is defined to be that positive matrix satisfying $E_3^2 = I - E_1^\dagger E_1 - E_2^\dagger E_2$; note that such a matrix exists because $I - E_1^\dagger E_1 - E_2^\dagger E_2$ is itself a positive matrix. Furthermore, note that the trhee

Th

Generalized measurements are obviously more general than the projective measurements described in most textbooks. Projective measurements have the feature that they are *repeatable*, in the sense that if one performs a projective measurement twice in a row on a quantum system, then one will obtain the same result both times. By contrast, most real measurements don't have this feature of being repeatable, which tips us off to the need for the formalism of generalized measurements. Nevertheless, even the generalized measurement formalism can be understood in terms of projective measurements as follows: the effect of a generalized measurement on a quantum system is *equivalent* to a unitary interaction between the system being measured and another "ancilla" system, followed by a projective measurement on the ancilla system. More precisely, suppose $\{E_i\}$ is a set of

measurement matrices satisfying the completeness relation $\sum_i E_i^\dagger E_i = I$. We introduce an ancilla system with orthonormal basis elements $|i\rangle$ indexed by the possible measurement outcomes. Define a matrix U acting on the joint quantum system-ancilla by the action:

$$U|\psi\rangle|0\rangle \equiv \sum_i E_i|\psi\rangle|i\rangle, \tag{B.8}$$

where $|0\rangle$ is some standard state of the ancilla and $|\psi\rangle$ is an arbitrary state of the quantum system being measured. It is easy to show using the completeness relation $\sum_i E_i^\dagger E_i = I$ that U can be extended to a unitary matrix acting on the entire state space of the joint system. Suppose we perform the unitary transformation U on the joint quantum system-ancilla, and then do a projective measurement of the ancilla in the $|i\rangle$ basis. It is then easily checked that the result of the measurement is i with probability $p_i = \mathrm{tr}(E_i \rho E_i^\dagger)$ and the corresponding post-measurement state of the system is $\rho_i' = E_i \rho E_i^\dagger / \mathrm{tr}(E_i \rho E_i^\dagger)$. Thus, the effect on the quantum system is exactly as we have described above for a generalized quantum measurement. Conversely, it is not difficult to verify that the effect of a unitary interaction between system and ancilla followed by a projective measurement on the ancilla can always be understood in terms of a generalized measurement (see for example Chapter 8 of [43]).

This notion of "padding" vectors of unequal dimension so they can be compared by the majorization relation is surprisingly useful, and we adopt the general convention that when x and y are of different dimension then $x \prec y$ means that $\tilde{x} \prec \tilde{y}$, where $\tilde{x}$ and $\tilde{y}$ are padded with extra zero components to ensure that they have the same dimension. For example, $(1/3, 1/3, 1/3) \prec (1/2, 1/2)$ since $(1/3, 1/3, 1/3) \prec (1/2, 1/2, 0)$. It is easy to check that this extended notion of majorization is well-defined, provided x and y both have non-negative components, and this will be the case for all the applications in this paper. Similarly, it is often useful to write $x = y$ provided the padded versions of x and y are equal, that is, the non-zero entries of x and y are equal. With these conventions, it is easy to see that algebraic manipulations proceed exactly as one would expect. For example, for non-negative real vectors w, x, y, z if $w \prec x, x = y, y \prec z$ then obviously $w \prec z$, even if all four vectors have different dimensionality. We occasionally make use of such elementary observations in proofs, without explicit comment.

The final result about majorization we shall need is that if P_i are a set of orthogonal projectors such that $\sum_i P_i = I$, and ρ is a density matrix, then

[8]

$$\lambda\left(\sum_i P_i \rho P_i\right) \prec \lambda(\rho). \tag{B.9}$$

Intuitively, if a projective measurement of a quantum system is performed, but we do not learn the result of the measurement, then the state of the system after measurement is more mixed than it was before. One way of proving this relation is via Horn's lemma; a sketch follows. First, note that it suffices to prove that $\lambda(P\rho P + Q\rho Q) \prec \lambda(\rho)$, where P and $Q = I - P$ are two orthogonal projectors satisfying $P + Q = I$. Once this is proved, the general relation (B.9) follows by a simple induction. However, if we define a unitary matrix $U \equiv P - Q$ then it is easy to verify that

$$P\rho P + Q\rho Q = \frac{\rho + U\rho U^\dagger}{2}. \tag{B.10}$$

Applying Horn's lemma and the easily proved fact that if $x_1 \prec y$ and $x_2 \prec y$ then $(x_1 + x_2)/2 \prec y$, it follows with a little simple linear algebra that $\lambda(P\rho P + Q\rho Q) \prec \lambda(\rho)$.

B.3 Quantum operations

Appendix C

Classification of ensembles for a density matrix

The density matrix is a tool used in quantum mechanics to deal with the case where we only have incomplete knowledge of a quantum state. A detailed discussion of the density matrix formalism may be found in most advanced quantum mechanics textbooks, and we assume the reader is familiar with the basic properties of the density matrix. Recall that if a quantum system is in state $|\psi_j\rangle$ with probability p_j, then the density matrix describing that situation is defined to be

$$\rho \equiv \sum_j p_j |\psi_j\rangle\langle\psi_j|. \tag{C.1}$$

For example, a single qubit in the state $|0\rangle$ with probability $1/2$ and state $(|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1/2$ may be described by the density matrix

$$\begin{aligned} \rho &= \frac{1}{2}|0\rangle\rangle 0| + \frac{1}{2}\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0| + \langle 1|}{\sqrt{2}}\right) && \text{(C.2)} \\ &= \frac{1}{4}\begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}, && \text{(C.3)} \end{aligned}$$

where we have written the matix with respect to the computational basis, $|0\rangle, |1\rangle$.

We call the collection $\{p_j, |\psi_j\rangle\}$ consisting of the probabilities p_j and their corresponding states $|\psi_j\rangle$ an *ensemble*, and we say that $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ is the density matrix generated by the ensemble $\{p_j, |\psi_j\rangle\}$.

One of the most useful results about density matrices is the characterization theorem guaranteeing that any ensemble $\{p_j, |\psi_j\rangle\}$ generates a density matrix, ρ, that (a) has unit trace, $\mathrm{tr}(\rho) = 1$, and (b) is a positive matrix, that is, $\langle\psi|\rho|\psi\rangle \geq 0$ for all $|\psi\rangle$. Conversely, for any positive matrix ρ with unit trace, there exists an ensemble $\{p_j, |\psi_j\rangle\}$ generating ρ. This characterization theorem thus justifies defining a matrix to be a density matrix precisely when it is a positive matrix with unit trace.

Exercise C.0.1: The characterization theorem for density matrices often includes the apparently supplementary condition that the density matrix is Hermitian. Show that this condition is unnecessary by proving that any positive matrix is automatically Hermitian.

Given a density matrix, ρ, that is, a positive matrix with unit trace, it turns out that, in general, there are many different ensembles $\{p_j, |\psi_j\rangle\}$ generating ρ. The most commonly used ensemble is, of course, the eigenensemble: letting $\lambda_j(\rho)$ be the eigenvalues of ρ, and $|j\rangle$ the corresponding eigenvectors, we have $\rho = \sum_j \lambda_j(\rho)|j\rangle\langle j|$. The conditions that ρ be positive and have unit trace imply that the $\lambda_j(\rho)$ form a probability distribution. Thus $\{\lambda_j(\rho), |j\rangle\}$ is an ensemble generating ρ.

However, the eigenensemble may not be the only ensemble generating a particular density matrix, as the following example shows. A qubit in the state $|0\rangle$ with probability $3/4$ and $|1\rangle$ with probability $1/4$ has density matrix

$$\rho = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|. \tag{C.4}$$

Suppose instead that we prepared the qubit in the state $|a\rangle$ or $|b\rangle$, with respective probabilities $1/2$, where $|a\rangle$ and $|b\rangle$ are defined by

$$|a\rangle \equiv \sqrt{\frac{3}{4}}|0\rangle + \sqrt{\frac{1}{4}}|1\rangle \tag{C.5}$$

$$|b\rangle \equiv \sqrt{\frac{3}{4}}|0\rangle - \sqrt{\frac{1}{4}}|1\rangle. \tag{C.6}$$

Direct calculation shows that the density matrix generated by this ensemble is

$$\rho = \frac{1}{2}|a\rangle\langle a| + \frac{1}{2}|b\rangle\langle b| = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|, \tag{C.7}$$

and thus is identical to the density matrix generated by the ensemble considered earlier, in which $|0\rangle$ is prepared with probability 3/4, and $|1\rangle$ is prepared with probability 1/4.

What this example shows is that two (or perhaps more) different ensembles of quantum states may give rise to the same density matrix. It is natural, therefore, to try to characterize exactly which ensembles $\{p_j, |\psi_j\rangle\}$ give rise to a particular density matrix, and we will now prove a theorem giving such a characterization. Interestingly, this theorem has been independently discovered several different times — I am aware of independent discoveried by Schrödinger[54], by Jaynes[28], and by Hughston, Jozsa and Wootters[27]; there may well be others. Indeed, Schrödinger, in his 1936 paper[54], does not even claim any particular priority for the result, presumably assuming that the result was already known to others.

We prove the characterization theorem by first proving an intermediate lemma which is just the desired characterization theorem in a simpler notation. To state and prove the lemma it helps to first introduce a little more nomenclature. A set of (possibly un-normalized) vectors $|\psi_j\rangle$ *generates* the operator $\rho \equiv \sum_j |\psi_j\rangle\langle\psi_j|$. The lemma gives necessary and sufficient conditions for two sets of vectors $|\psi_j\rangle$ and $|\phi_k\rangle$ to generate the same matrix. The connection with density matrices is then made by As a corollary we shall characterize the set of ensembles consistent with a given density matrix.

**** Done up to here *****

> *Lemma C.0.1*: **(Ensemble classification theorem)** The sets $|\psi_i\rangle$ and $|\phi_j\rangle$ generate the same density matrix if and only if
>
> $$|\psi_i\rangle = \sum_j u_{ij} |\phi_j\rangle, \tag{C.8}$$
>
> where u_{ij} is a unitary matrix of complex numbers, with indices i and j, and we "pad" whichever set of vectors $|\psi_i\rangle$ or $|\phi_j\rangle$ contains fewer elements with extra **0** vectors to ensure that the two sets have the same number of elements.

As an immediate consequence of the theorem note that $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| = \sum_k q_k |\phi_k\rangle\langle\phi_k|$ for *normalized* quantum states $|\psi_j\rangle$ and $|\phi_k\rangle$ if and only if

$$\sqrt{p_j}|\psi_j\rangle = \sum_k u_{jk} \sqrt{q_k} |\phi_k\rangle \tag{C.9}$$

for some unitary matrix u with entries u_{jk}, and we pad the smaller ensemble with entries having probability zero to ensure that the two ensembles have the same number of elements. It is easily checked that our earlier example, Equation (C.7), of a density matrix with two different ensemble decompositions, is a special case of this general result. Another useful consequence is the following simple exercise, due to Jaynes[28].

Exercise C.0.2: Let ρ be a density matrix and suppose $|\psi_i\rangle$ is some linearly independent set of pure states spanning the support of ρ. Show that there is a unique probability distribution (p_i) such that $\rho = \sum_i p_i |\psi_\rangle\langle\psi|$, given by

$$p_i = \frac{1}{\langle\psi_i|\rho^{-1}|\psi_i\rangle}, \tag{C.10}$$

where ρ^{-1} is defined to be the inverse of ρ on the support of ρ, and otherwise is zero. (This removes the problem that ρ may not have an inverse.)

The proof of Theorem C is a simple exercise in linear algebra:
Proof: The reverse implication is trivial: we suppose $|\psi_i\rangle = \sum_j u_{ij}|\phi_j\rangle$ for some unitary u_{ij} and then use straightforward algebra to show that $\sum_i |\psi_i\rangle\langle\psi_i| = \sum_j |\phi_j\rangle\langle\phi_j|$. The converse is a trifle more difficult. Suppose

$$\rho = \sum_i |\psi_i\rangle\langle\psi_i| = \sum_j |\phi_j\rangle\langle\phi_j|, \tag{C.11}$$

and let $|\psi\rangle$ be any vector orthonormal to the space spanned by the $|\phi_j\rangle$, so $\langle\psi|\phi_j\rangle\langle\phi_j|\psi\rangle = 0$ for all j and thus from (C.11) we see that

$$0 = \sum_i \langle\psi|\psi_i\rangle\langle\psi_i|\psi\rangle = \sum_i |\langle\psi|\psi_i\rangle|^2. \tag{C.12}$$

We deduce that $\langle\psi|\psi_i\rangle = 0$ for all i and all $|\psi\rangle$ orthonormal to the space spanned by the $|\phi_j\rangle$. It follows that each $|\psi_i\rangle$ may be expressed as a linear combination of the $|\phi_j\rangle$, $|\psi_i\rangle = \sum_j c_{ij}|\phi_j\rangle$, for some matrix of complex numbers c_{ij}. Thus

$$\sum_j |\phi_j\rangle\langle\phi_j| = \sum_{jk} \left(\sum_i c_{ij} c_{ik}^* \right) |\phi_j\rangle\langle\phi_k|, \tag{C.13}$$

from which we can see that $\sum_i c_{ij} c_{ik}^* = \delta_{jk}$, so c is unitary, as claimed. ■

Hints for Lecture C

Hint for Exercise C Show that any matrix M can be decomposed $M = A+iB$, where A and B are Hermitian. Then argue that if M is positive, then B must be zero.

Bibliography

[1] P. M. Alberti and A. Uhlmann. *Stochasticity and partial order: doubly stochastic maps and unitary mixing*. Dordrecht, Boston, 1982.

[2] T. Ando. Majorization, doubly stochastic matrices, and comparison of eigenvalues. *Linear Algebra and Its Applications*, 118:163–248, 1989.

[3] T. Ando. Majorizations and inequalities in matrix theory. *Linear Algebra and Its Applications*, 199:17–67, 1994.

[4] C. H. Bennett. The thermodynamics of computation – a review. *Int. J. Theor. Phys.*, 21(12):905–939, 1982.

[5] C. H. Bennett. Demons, engines and the second law. *Sci. Am.*, 295(5):108, 1987.

[6] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824, 1996. arXiv:quant-ph/9604024.

[7] F. Berezin and I. M Gel'fand. Some remarks on the theory of spherical functions on symmetric riemannian manifolds. *Trudi Moscow Math. Ob.*, 5:311–351, 1956.

[8] R. Bhatia. *Matrix analysis*. Springer-Verlag, New York, 1997.

[9] G. Birkhoff. Tres observaciones sobres el algebra lineal. *Univ. Nac. Tucumán Rev. Ser. A*, 5:147–151, 1946.

[10] S. L. Braunstein, C. M. Caves, R. Jozsa, N. Linden, S. Popescu, and R. Schack. Separability of very noisy mixed states and implications for NMR quantum computing. *Phys. Rev. Lett.*, 83(5):1054–1057, 1999. arXive:quant-ph/9811018.

[11] C. M. Caves and G. J. Milburn. Qutrit entanglement. *Optics Communications*, 179(1–6):439–446, 2000. arXiv:quant-ph/9910001.

[12] N. J. Cerf and C. Adami. Quantum extension of conditional probability. *Phys. Rev. A*, 60(2):893–897, 1999. appeared as part of arXiv:quant-ph/9710001.

[13] A. Chefles. Deterministic quantum state transformations. *arXiv:quant-ph/9911086*, 1999.

[14] W. Dür and J. I. Cirac. Classification of multi-qubit mixed states: separability and distillability properties. *Phys. Rev. A*, 61:042314, 2000. arXiv:quant-ph/9911044.

[15] W. Dür, J. I. Cirac, and R. Tarrach. Separability and distillability of multiparticle quantum systems. *Phys. Rev. Lett.*, 83(17):3562–3565, 1999.

[16] C. A. Fuchs and A. Peres. Quantum state disturbance vs. information gain: Uncertainty relations for quantum information. *Phys. Rev. A*, 53:2038–2045, 1996. arXiv:quant-ph/9512023.

[17] W. Fulton. Eigenvalues, invariant factors, highest weights, and Schubert calculus. *Bull. Amer. Math. Soc.*, 37(3):209–249, 2000.

[18] C. W. Gardiner. *Quantum Noise*. Springer-Verlag, Berlin, 1991.

[19] L. Hardy. Method of areas for manipulating the entanglement properties of one copy of a two-particle pure entangled state. *Phys. Rev. A*, 60(3):1912–1923, 1999.

[20] S. Hill and W. K. Wootters. Entanglement of a pair of quantum bits. *Phys. Rev. Lett.*, 78(26):5022–5025, 1997. arXiv:quant-ph/9703041.

[21] A. Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *Amer. J. Math.*, 76:620–630, 1954.

[22] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1991.

[23] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223(1-2):1–8, 1996. arXiv:quant-ph/9605038.

[24] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? *Phys. Rev. Lett.*, 80(24):5239–5242, 1998.

[25] P. Horodecki, R. Horodecki, and M. Horodecki. Entanglement and thermodynamical analogies. *Acta Phys. Slov.*, 48:141–156, 1998. arXiv:quant-ph/9805072.

[26] R. Horodecki, P. Horodecki, and M. Horodecki. Quantum α-entropy inequalities: independent condition fo local realism? *Phys. Lett. A*, 210:377–381, 1996.

[27] L. P. Hughston, R. Jozsa, and W. K. Wootters. A complete classification of quantum ensembles having a given density matrix. *Phys. Lett. A*, 183:14–18, 1993.

[28] E. T. Jaynes. Information theory and statistical mechanics. ii. *Phys. Rev.*, 108(2):171–190, 1957.

[29] J. G. Jensen and R. Schack. A simple algorithm for local conversion of pure states. *arXiv:quant-ph/0006049*, 2000.

[30] D. Jonathan and M. B. Plenio. Entanglement-assisted local manipulation of pure states. *arXiv:quant-ph/9905071*, 1999.

[31] D. Jonathan and M. B. Plenio. Minimal conditions for local pure state entanglement manipulation. *Phys. Rev. Lett.*, 83(7):1455–1458, 1999. arXiv:quant-ph/9903054.

[32] Karl Kraus. *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Lecture Notes in Physics, Vol. 190. Springer-Verlag, Berlin, 1983.

[33] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.*, 5:183, 1961.

[34] O. E. Lanford and D. Robinson. Mean entropy of states in quantum-statistical mechanics. *J. Math. Phys.*, 9(7):1120–1125, 1968.

[35] V. B. Lidskii. On the proper values of a sum and product of symmetric matrices. *Dokl. Akad. Nauk SSSR*, 75:769–772, 1950.

[36] E. H. Lieb and M. B. Ruskai. Proof of the strong subadditivity of quantum mechanical entropy. *J. Math. Phys.*, 14:1938–1941, 1973.

[37] H.-K. Lo and S. Popescu. Concentrating local entanglment by local actions – beyond mean values. *arXiv:quant-ph/9707038*, 1997.

[38] A. W. Marshall and I. Olkin. *Inequalities: theory of majorization and its applications*. Academic Press, New York, 1979.

[39] M. A. Nielsen. *Quantum Information Theory*. PhD thesis, University of New Mexico, 1998. arXiv:quant-ph/0011036.

[40] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83(2):436–439, 1999. arXiv:quant-ph/9811053.

[41] M. A. Nielsen. Probability distributions consistent with a mixed state. *Phys. Rev. A*, 62:052308, 2000. arXiv:quant-ph/9909020.

[42] M. A. Nielsen, M. J. Bremner, J. L. Dodd, A. M. Childs, and C. M. Dawson. Universal simulation of Hamiltonian dynamics for qudits. *Phys. Rev. A*, 66(2):022317, 2002. arXiv:quant-ph/0109064.

[43] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000.

[44] M. Ohya and D. Petz. *Quantum entropy and its use*. Springer-Verlag, Berlin, 1993.

[45] A. Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic, Dordrecht, 1993.

[46] A. Peres. *Phys. Rev. Lett.*, 77:1413, 1996.

[47] A. O. Pittenger and M. H. Rubin. Separability and Fourier representations of density matrices. *Phys. Rev. A*, 62(3):032313, 2000. arXiv:quant-ph/0001014.

[48] M. B. Plenio and P. L. Knight. The quantum-jump approach to dissipative dynamics in quantum optics. *Rev. Mod. Phys.*, 70(1):101–144, 1998.

[49] F. P. Preparata. *Computational geometry: an introduction*. Springer-Verlag, New York, 1985.

[50] J. Preskill. *Physics 229: Advanced mathematical methods of physics — Quantum computation and information*. California Institute of Technology, Pasadena, CA, 1998. http://www.theory.caltech.edu/people/preskill/ph229/.

[51] R. T. Rockafeller. *Convex Analysis*. Princeton University Press, Princeton, 1970.

[52] P. Rungta, W. J. Munro, K. Nemoto, P. Deuar, G. J. Milburn, and C. M. Caves. Qudit entanglement. *arXiv:quant-ph/0001075*, 2000.

[53] S. Sachdev. *Quantum phase transitions*. Cambridge University Press, Cambridge, 1999.

[54] Erwin Schrödinger. Probability relations between separated systems. *Proc. Cambridge Phil. Soc.*, 32:446–452, 1936.

[55] M. F. Smiley. Inequalities related to lidskii's. *Proc. Amer. Math. Soc.*, 19:1029–1034, 1968.

[56] A. Uhlmann. On the shannon entropy and related functionals on convex sets. *Rep. Math. Phys.*, 1(2):147–159, 1970.

[57] A. Uhlmann. Sätze über dichtematrizen. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 20:633–637, 1971.

[58] A. Uhlmann. Endlic-dimensionale dichtematrizen i. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 21:421–452, 1972.

[59] A. Uhlmann. Endlich-dimensionale dichtematrizen ii. *Wiss. Z. Karl-Marx-Univ. Leipzig*, 22:139–177, 1973.

[60] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57(3):1619–1633, 1998.

[61] G. Vidal. Entanglement of pure states for a single copy. *Phys. Rev. Lett.*, 83(5):1046–1049, 1999.

[62] G. Vidal. Entanglement monotones. *J. Mod. Opt.*, 47(2–3):355–376, 2000.

[63] G. Vidal and R. Tarrach. Robustness of entanglement. *Phys. Rev. A*, 59(1):141–155, 1999. arXiv:9806094.

[64] A. Wehrl. How chaotic is a state of a quantum system. *Rep. Math. Phys.*, 6(1):15–28, 1974.

[65] A. Wehrl. General properties of entropy. *Rev. Mod. Phys.*, 50:221, 1978.

[66] R. F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40(8):4277–4281, 1989.

[67] H. W. Wielandt. Topics in the analytic theory of matrices. University of Wisconsin mimeographed lecture notes (1967), 1967.

[68] W. K. Wootters. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.*, 80(10):2245–2248, 1998.

[69] P. Zoller and C. W. Gardiner. Quantum noise in quantum optics: the stochastic schrödinger equation. In S. Reynaud, E. Giacobino, and J. Zinn-Justin, editors, *Quantum fluctuations: Les Houches Summer School LXIII*, Amsterdam, 1997. Elsevier.

[70] W. H. Zurek. Algorithmic randomness and physical entropy. *Phys. Rev. A*, 40:4731, 1989.

[71] W. H. Zurek. Decoherence and the transition from quantum to classical. *Phys. Tod.*, 44(10):36–44, 1991.