

Lectures Notes on Quantum Information Theory

Michael Walter and Maris Ozols, University of Amsterdam

Spring 2021

Summary

This course gives an introduction to the mathematics of quantum information.

Notation

- Σ – finite set, Section 1.1
- $P(\Sigma)$ – probability distributions on Σ , Eq. (1.12)
- $\mathcal{H}, \mathcal{K}, \dots$ – finite-dimensional Hilbert spaces, Section 1.1
- $L(\mathcal{H}, \mathcal{H}')$ – linear operators from \mathcal{H} to \mathcal{H}' , Eq. (1.5)
- $\text{PSD}(\mathcal{H})$ – cone of positive semidefinite operators on \mathcal{H} , Eq. (1.8)
- $A \leq B$ – (Loewner) partial order defined using the PSD cone, Eq. (1.9)
- $\text{Tr}[M]$ – trace of matrix M , Eq. (1.4)
- $\text{Tr}_B[M_{AB}]$ – partial trace of M_{AB} over system B , Definition 2.8
- $\mathcal{D}(\mathcal{H})$ – quantum states on \mathcal{H} , Definition 1.7
- $|\Phi_{AB}^+\rangle$ – maximally entangled state, Definition 3.5
- $|\Phi^{(x,z)}\rangle$ – Bell states, Eq. (3.4)
- $U(\mathcal{H})$ – unitary operators on \mathcal{H} , Eq. (2.21)
- $U(\mathcal{H}, \mathcal{K})$ – isometries from \mathcal{H} to \mathcal{K} , Eq. (2.20)
- $f(M), \sqrt{M}, \log M, \dots$ – functions of Hermitian operators, Definition 1.6
- $\|x\|_p$ – ℓ^p -norm of vectors, Eq. (4.1)
- $\|M\|_p$ – Schatten p -norm of operator M , Definition 4.1
- $\|M\|_1$ – trace norm of operator M , Eq. (4.2)
- $\|M\|_2$ – Frobenius norm of operator M , Eq. (4.3)
- $\|M\|_\infty$ – operator norm of M , Eq. (4.5)
- $\langle M, N \rangle$ – Hilbert-Schmidt inner product, Eq. (4.4)
- $T(p, q)$ – trace distance between distributions p and q , Definition 7.4
- $T(\rho, \sigma)$ – trace distance between states ρ and σ , Definition 4.6
- $F(\rho, \sigma)$ – fidelity between states ρ and σ , Definition 4.9
- \mathcal{I}_A – identity channel on \mathcal{H}_A , Eq. (4.26)
- $\text{CP}(\mathcal{H}_A, \mathcal{H}_B)$ – completely positive maps from \mathcal{H}_A to \mathcal{H}_B , Definition 4.15
- $\mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$ – quantum channels from \mathcal{H}_A to \mathcal{H}_B , Definition 4.15
- J_{AB}^Φ – Choi operator of $\Phi_{A \rightarrow B}$, Eq. (5.1)
- Δ – completely dephasing channel, Eq. (5.3)
- $H(p), H(X), H(X)_p$ – Shannon entropy, Definitions 6.1 and 6.4
- $T_{n,\varepsilon}(p)$ – typical set, Definition 6.10
- $H(\rho), H(A), H(A)_\rho$ – von Neumann entropy, Definitions 7.1 and 8.1
- $F(\mathcal{T}, \rho)$ – channel fidelity of channel \mathcal{T} and state ρ , Definition 7.7
- $S_{n,\varepsilon}(\rho)$ – typical subspace, Definition 7.11
- $I(A : B)_\rho$ – mutual information of state ρ_{AB} , Definition 8.3
- $\chi(\{p_x, \rho_x\})$ – Holevo χ -quantity of ensemble $\{p_x, \rho_x\}$, Definition 9.1
- $D(p \| q)$ – relative entropy of distribution p with respect to distribution q , Definition 9.6
- $D(\rho \| \sigma)$ – quantum relative entropy of state ρ with respect to state σ , Definition 9.7
- $\text{Sep}(\mathcal{H}_A : \mathcal{H}_B)$ – separable operators on $\mathcal{H}_A \otimes \mathcal{H}_B$, Definition 10.8
- $\text{SepD}(\mathcal{H}_A : \mathcal{H}_B)$ – separable states on $\mathcal{H}_A \otimes \mathcal{H}_B$, Definition 3.1
- $\text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$ – separable completely positive maps, Definition 10.4
- $\text{SepC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$ – separable quantum channels, Definition 10.4
- $\text{LOCC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$ – LOCC channels, Definition 10.3
- $\text{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$ – operators of entanglement rank at most r on $\mathcal{H}_A \otimes \mathcal{H}_B$, Definition 10.10
- $x \prec y$ – majorization of vectors, Definition 11.2

- P_π – permutation matrices, Definition [11.3](#)
- $\lambda(A)$ – vector of eigenvalues of a Hermitian operator, Section [11.2](#)
- $A \prec B$ – majorization of Hermitian operators, Definition [11.12](#)
- $E_D(A : B)_\rho$ – distillable entanglement of ρ , Definition [12.1](#)
- $E_C(A : B)_\rho$ – entanglement cost of ρ , Definition [12.2](#)
- $\text{Sym}^n(\mathcal{H})$ – symmetric subspace of $\mathcal{H}^{\otimes n}$, Definition [13.1](#)

Acknowledgements

We would like to thank Dmitry Grinko, Harold Nieuwboer, Alvaro Piedrafita, and Freek Witteveen for all their help, feedback, and corrections. Thanks also to Christiaan van Asperen, Khallil Berrekkal, Robert Cañellas, Egor Cickovskij, Kjartan van Driel, Jasper van Egeraat, Dylan Feenstra, Titas Geryba, Thijs Jenneskens, Dimitrios Loupas, Andrea Mazzocco, Arend-Jan Quist, Misha Schram, and Jens de Vries for spotting typos.

Last updated: July 1, 2021.

Contents

1	Introduction to quantum information, states, and measurements	7
1.1	Hilbert space and Dirac notation	7
1.2	Operators, eigenvectors, eigenvalues	9
1.3	Quantum states	12
1.4	States of a single quantum bit: Bloch ball	14
1.5	Measurements	16
1.6	Exercises	19
2	Joint systems, reduced states, purifications	23
2.1	Joint or composite systems	23
2.2	Measurements on subsystems	25
2.3	Partial trace and reduced states	26
2.4	Purifications	30
2.5	Schmidt decomposition	32
2.6	Exercises	34
3	Entanglement	37
3.1	Separable and entangled states	37
3.2	Bell states and superdense coding	40
3.3	Teleportation	41
3.4	CHSH game (optional)	43
3.5	Exercises	47
4	Trace distance and fidelity, classical and quantum channels	50
4.1	Norms of operators	50
4.2	Trace distance and fidelity	53
4.3	Channels in probability theory	57
4.4	Quantum channels	59
4.5	Exercises	64
5	Structure of quantum channels	66
5.1	Superoperators and complete positivity	66
5.2	Characterizing quantum channels	69
5.3	Exercises	71
6	Shannon entropy and data compression	75
6.1	Shannon entropy	75
6.2	Lossy and lossless compression	78
6.3	Block codes, Shannon's source coding theorem, typical sets	79
6.4	Exercises	82
7	From classical to quantum compression	84

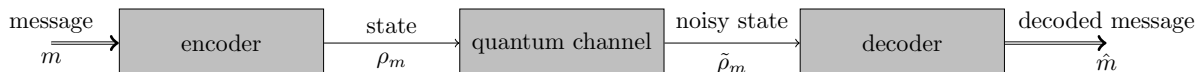
7.1	Von Neumann Entropy	84
7.2	Motivation: Classical compression and correlations	85
7.3	Quantum codes and compression	87
7.4	Channel fidelity	89
7.5	Schumacher's theorem and typical subspaces	89
7.6	Exercises	92
8	Entropy and subsystems	95
8.1	Entropies of subsystems	95
8.2	Mutual information	97
8.3	Exercises	100
9	Holevo bound and relative entropy	103
9.1	Holevo bound	103
9.2	Relative entropy	105
9.3	Quantum relative entropy	107
9.4	Exercises	109
10	LOCC and separable channels	111
10.1	Instruments and LOCC channels	111
10.2	Separable channels and operators	113
10.3	Entanglement rank	116
10.4	Separable and LOCC measurements	117
10.5	Exercises	119
11	Majorization and Nielsen's theorem	121
11.1	Wealth inequality and majorization	121
11.2	Majorization for Hermitian operators	126
11.3	LOCC and Nielsen's theorem	128
11.4	Exercises	131
12	Distillable entanglement and entanglement cost	133
12.1	Conversion, distillable entanglement, entanglement cost	133
12.2	Distillable entanglement equals entanglement cost for pure states	136
12.3	Exercises	137
13	Monogamy of entanglement	138
13.1	Sharing classical vs quantum correlations	138
13.2	The symmetric subspace	140
13.3	The quantum de Finetti theorem	143
13.4	Exercises	146

Lecture 1

Introduction to quantum information, states, and measurements

This course gives an introduction to the mathematical theory of quantum information. We will learn the basic formalism and toolbox that allows us to reason about states, channels, and measurements, discuss important notions such as entropy and entanglement, and see how these can be applied to solve fundamental mathematical problems that relate to the storage, estimation, compression, and transmission of quantum information.

To make this concrete, suppose that we would like to transmit a message through a communication channel (think of an optical fiber with some loss). To achieve this, we might try to encode our message m into a quantum state ρ_m , which we then send through the channel. The receiver receives some noisy state $\tilde{\rho}_m$ and wants to apply a measurement that allows them to recover m with high probability. This situation is visualized in the following figure:



What is the optimal way of encoding the message when the channel is quantum mechanical? To even make sense of this question, we first have to learn how to mathematically model quantum states and channels. We will do so in the first weeks of the course. In the remainder of the course, we will learn a variety of mathematical tools that will eventually allow us to attack information processing problems such as the above.

Throughout the course, we will use some linear algebra (in finite dimensions) and probability theory (of distributions with finitely many outcomes). See Chapter 2.1 and Appendix 1 of “Quantum Computation and Quantum Information” by Nielsen and Chuang for a good summary. We will recap the most important bits in these lecture notes.

1.1 Hilbert space and Dirac notation

Today, we start with an introduction to the axioms (rules, laws, postulates) of quantum information. Some of the axioms may look differently from (or more general than) what you remember from a previous course on quantum mechanics, and we will discuss this carefully. The first axiom is the following:

Axiom 1.1 (System). To every quantum system, we associate a *Hilbert space* \mathcal{H} .

Throughout this course we will restrict to finite-dimensional Hilbert spaces. Recall that a finite-dimensional Hilbert space is nothing but a complex vector space together with an inner

product, which we denote by $\langle\phi|\psi\rangle$. We will always take our inner product to be anti-linear in the *first* argument! Any Hilbert space carries a natural norm, defined by $\|\psi\| := \sqrt{\langle\psi|\psi\rangle}$.

Throughout this course we will use Dirac’s “bra-ket” notation, with “kets” $|\psi\rangle$ denoting vectors in \mathcal{H} and “bras” $\langle\psi|$ denoting the corresponding dual vector in \mathcal{H}^* , i.e., $\langle\psi| := \langle\psi|\cdot\rangle$. The latter means that $\langle\psi|$ is the linear functional that sends a vector $|\phi\rangle$ to the inner product $\langle\psi|\phi\rangle$. Thus, “bra” and “ket” together give the inner product $\langle\psi|\phi\rangle = \langle\psi||\phi\rangle$. A unit vector is a vector $|\psi\rangle$ whose norm (or norm squared) is equal to one, i.e., $\langle\psi|\psi\rangle = 1$.

A well-known example is the Hilbert space $\mathcal{H} = \mathbb{C}^d$ with the standard inner product $\langle\phi|\psi\rangle = \sum_{i=1}^d \overline{\phi_i} \psi_i$ and norm $\|\psi\| = (\sum_{i=1}^d |\psi_i|^2)^{1/2}$. Any d -dimensional Hilbert space can be identified with \mathbb{C}^d by choosing an orthonormal basis. When we speak of a *basis* of a Hilbert space we always mean an orthonormal basis. One can think of kets as column vectors and bras as row vectors. Hence, if $|\psi\rangle$ is a column vector, then $\langle\psi|$ denotes the row vector obtained by taking the *conjugate transpose* of the column vector. The following compares Dirac notation with the corresponding expression in coordinates:

$$\begin{aligned} |\psi\rangle &= \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_d \end{pmatrix}, & \langle\psi| &= (\overline{\psi_1} \quad \cdots \quad \overline{\psi_d}), \\ \langle\phi|\psi\rangle &= \sum_{i=1}^d \overline{\phi_i} \psi_i, & |\psi\rangle\langle\phi| &= \begin{pmatrix} \psi_1 \overline{\phi_1} & \cdots & \psi_1 \overline{\phi_d} \\ \vdots & & \vdots \\ \psi_d \overline{\phi_1} & \cdots & \psi_d \overline{\phi_d} \end{pmatrix}. \end{aligned} \tag{1.1}$$

As a first nontrivial example of using Dirac notation, let $|\psi\rangle$ be a unit vector. Then

$$P = |\psi\rangle\langle\psi| \tag{1.2}$$

is the *orthogonal projection* (‘projector’) onto the one-dimensional space $\mathbb{C}|\psi\rangle$. For example, if $|\psi\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, then $|\psi\rangle\langle\psi| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, which clearly projects on the first coordinate. To prove the general claim, we only need to verify that $P|\psi\rangle = |\psi\rangle\langle\psi|\psi\rangle = |\psi\rangle$, since $\langle\psi|\psi\rangle = \|\psi\|^2 = 1$, while $P|\phi\rangle = |\psi\rangle\langle\psi|\phi\rangle = 0$ for any $|\phi\rangle$ that is orthogonal to $|\psi\rangle$. From this, it is also clear that

$$\sum_i |e_i\rangle\langle e_i| = I \tag{1.3}$$

is the identify operator for any choice of orthonormal basis $|e_i\rangle$.

Another useful formula is that the trace of any $X \in L(\mathcal{H})$ can be calculated as follows:

$$\text{Tr}[X] = \sum_i \langle e_i | X | e_i \rangle. \tag{1.4}$$

Indeed, the right-hand side terms are just the diagonal entries of X when represented as a matrix with respect to the basis $|e_i\rangle$. Exercises 1.1 and 1.2 allow you to sharpen your Dirac notation skills some more.

1.1.1 Qubits and qudits

The simplest quantum system is the *qubit* – short for *quantum bit*. It corresponds to the two-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2$. We denote its *standard (or computational) basis* by

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

These two vectors together make up a classical *bit* inside a quantum bit: $\{0, 1\} \ni x \mapsto |x\rangle \in \mathbb{C}^2$.

More generally, a quantum system with Hilbert space $\mathcal{H} = \mathbb{C}^d$ is called a *qudit*. We denote its *standard basis* by $|x\rangle$ for $x \in \{0, 1, \dots, d-1\}$, that is,

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots$$

In quantum information, it is often useful to work with Hilbert spaces that have a privileged basis labeled by some set Σ . The reason is that we might want to use this basis to store some classical data (such as a message that one would like to transmit, encrypt, or compute with, or the outcomes of measurement, etc). In principle we could simply use $\mathcal{H} = \mathbb{C}^d$, as discussed above, but it is often convenient to allow Σ to be some arbitrary finite set. This is completely analogously to how in probability theory one usually considers probability distributions over an arbitrary alphabet.

Therefore, given a finite set Σ , we denote by \mathbb{C}^Σ the Hilbert space with orthonormal basis $\{|x\rangle\}_{x \in \Sigma}$. That is, $\langle x|y\rangle = \delta_{x,y}$. Note that the Hilbert space is uniquely determined by this requirement. As before, we call the basis $\{|x\rangle\}_{x \in \Sigma}$ the *standard (or computational) basis* of \mathbb{C}^Σ .

Remark 1.2. If you would like to picture this vector space concretely, enumerate the elements of Σ in some arbitrary way, say $\Sigma = \{x_1, \dots, x_d\}$, where $d = |\Sigma|$. Then you can identify \mathbb{C}^Σ with \mathbb{C}^d (the basis vector $|x_j\rangle$ of the former corresponds to the basis vector $|j\rangle$ of the latter).

Formally, the vector space \mathbb{C}^Σ can be defined as the vector space of functions $\Sigma \rightarrow \mathbb{C}$, equipped with the inner product $\langle f|g\rangle := \sum_{x \in \Sigma} \overline{f(x)}g(x)$. In this picture, the standard basis vector $|x\rangle$ corresponds to the function $f_x: \Sigma \rightarrow \mathbb{C}$, $f_x(y) = \delta_{x,y}$ which sends x to 1 and all other $y \neq x$ to 0.

1.2 Operators, eigenvectors, eigenvalues

Throughout these lectures we will often deal with operators on Hilbert spaces, so it is useful to introduce some notation and recall some concepts from linear algebra. For Hilbert spaces \mathcal{H} and \mathcal{K} , define

$$\mathcal{L}(\mathcal{H}, \mathcal{K}) := \{A: \mathcal{H} \rightarrow \mathcal{K} \text{ linear}\}, \quad \mathcal{L}(\mathcal{H}) := \mathcal{L}(\mathcal{H}, \mathcal{H}) = \{A: \mathcal{H} \rightarrow \mathcal{H} \text{ linear}\}. \quad (1.5)$$

We write I or $I_{\mathcal{H}}$ for the identity operator on a Hilbert space \mathcal{H} . **Recall that any operator $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ has an *adjoint*.** This is the operator $A^\dagger \in \mathcal{L}(\mathcal{K}, \mathcal{H})$ defined by the property that

$$\langle \phi|A^\dagger|\psi\rangle = \overline{\langle \psi|A|\phi\rangle} \quad \forall |\phi\rangle \in \mathcal{H}, |\psi\rangle \in \mathcal{K}.$$

If you write A as a matrix with respect to an arbitrary orthonormal basis, then the adjoint is given by the conjugate transpose matrix:

$$A^\dagger = \overline{A}^\top = (\overline{A})^\top.$$

Note that this is the same rule that we used to go from a ‘ket’ to the corresponding ‘bra’, see Eq. (1.1). Indeed, if we think of $|\psi\rangle \in \mathcal{H}$ as an operator $\mathbb{C} \rightarrow \mathcal{H}$ then it is not hard (but perhaps still slightly confusing) to verify that $\langle \psi| = |\psi\rangle^\dagger$ – so this makes perfect sense!

An operator $A \in \mathcal{L}(\mathcal{H})$ is called *Hermitian* if $A = A^\dagger$. The set of Hermitian operators forms a *real* vector space of dimension d^2 , where $d = \dim \mathcal{H}$. Hermitian operators are diagonalizable with real eigenvalues and orthonormal eigenvectors.¹ This is the content of the following important result.

¹Recall that $|\psi\rangle$ is called an *eigenvector* of A with *eigenvalue* a if $A|\psi\rangle = a|\psi\rangle$.

Theorem 1.3 (Spectral theorem for Hermitian operators). *Let $A \in L(\mathcal{H})$ be a Hermitian operator, where $d = \dim \mathcal{H}$. Then there exist real numbers $a_1, \dots, a_d \in \mathbb{R}$ and an orthonormal basis $|\psi_1\rangle, \dots, |\psi_d\rangle$ of \mathcal{H} such that each $|\psi_j\rangle$ is an eigenvector of A , with eigenvalue a_j . Moreover, we have the following “eigendecomposition”:*

$$A = \sum_{j=1}^d a_j |\psi_j\rangle \langle \psi_j|. \quad (1.6)$$

Conversely, any operator that has a decomposition of the form of Eq. (1.6), with real a_i and orthonormal $|\psi_j\rangle$, is necessarily Hermitian with these eigenvectors and eigenvalues. The latter can be seen by verifying that

$$A|\psi_k\rangle = \sum_{j=1}^d a_j |\psi_j\rangle \langle \psi_j | \psi_k \rangle = \sum_{j=1}^d a_j |\psi_j\rangle \delta_{j,k} = a_k |\psi_k\rangle$$

for $k = 1, \dots, d$. In the second step we crucially used the orthonormality of the $|\psi_j\rangle$. Indeed, if A is of the form in Eq. (1.6) but the vectors $|\psi_j\rangle$ in this decomposition are *not* orthogonal, they need not be eigenvectors of A , see Exercise 1.3.

An important class of Hermitian operators are the (orthogonal) projections, which are the Hermitian operators P such that $P^2 = P$. Equivalently, their eigenvalues are in $\{0, 1\}$, see Exercise 1.4. For example, $P = |\psi\rangle \langle \psi|$ is a projection for any unit vector $|\psi\rangle$, as we already discussed in Eq. (1.2). Note that the eigendecomposition (1.6) decomposes any Hermitian operator into a linear combination of projections $|\psi_j\rangle \langle \psi_j|$ onto (pairwise orthogonal) eigenvectors.

The eigendecomposition as written above is not unique. Indeed, while the eigenvalues and eigenspaces are uniquely determined, (1.6) depends on choosing an orthonormal basis of eigenvectors in each eigenspaces. If we want a unique decomposition, we can instead write:

$$A = \sum_{a \in S} a P_a, \quad (1.7)$$

where S is the set of eigenvalues of A and P_a denotes the orthogonal projection onto the eigenspace corresponding to eigenvalue $a \in S$. This decomposition is unique, and it can be obtained from Eq. (1.6) by setting $S := \{a_j\}$ and $P_a := \sum_{j \text{ s.t. } a_j=a} |\psi_j\rangle \langle \psi_j|$.

We now come to a central definition. We say that an operator A is *positive semidefinite (PSD)* if A is Hermitian and all its eigenvalues are nonnegative. Thus A can be written as in Eq. (1.6) with $a_i \geq 0$. Positive semidefinite operators are so important that we will give them their own notation and define

$$\text{PSD}(\mathcal{H}) = \{A \in L(\mathcal{H}) : A \text{ positive semidefinite}\}. \quad (1.8)$$

A positive semidefinite operator such that all $a_i > 0$ is called *positive definite* (PD), and we write $\text{PD}(\mathcal{H})$ for the subset of positive definite operators. Equivalently, $\text{PD}(\mathcal{H})$ consists of those operators in $\text{PSD}(\mathcal{H})$ that are invertible. Both sets $\text{PSD}(\mathcal{H})$ and $\text{PD}(\mathcal{H})$ are convex,² as you can prove in Exercise 1.11.

In general it can be difficult to compute the eigenvalues. To this end, the following criterion is useful to test when an operator is PSD (Exercises 1.5 and 1.7):

²Recall that a set S is *convex* if $px + (1-p)y \in S$ for every $x, y \in S$ and $p \in [0, 1]$.

Lemma 1.4 (When is an operator positive semidefinite?). *For a Hermitian operator $A \in \mathcal{L}(\mathcal{H})$, the following five conditions are equivalent:*

- (a) A is positive semidefinite.
- (b) $A = B^\dagger B$ for an operator $B \in \mathcal{L}(\mathcal{H})$.
- (c) $A = B^\dagger B$ for an operator $B \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ and some Hilbert space \mathcal{K} .
- (d) $\langle \psi | A | \psi \rangle \geq 0$ for every $|\psi\rangle \in \mathcal{H}$.
- (e) $\text{Tr}[AC] \geq 0$ for every $C \in \text{PSD}(\mathcal{H})$.

Careful: There are Hermitian matrices A such that all entries A_{ij} are positive, but A is *not* positive semidefinite! You can find an example in Exercise 1.8.

We introduce one final piece of notation: for two operators $A, B \in \mathcal{L}(\mathcal{H})$, we write

$$A \geq B \quad \text{or} \quad B \leq A \tag{1.9}$$

to denote that $A - B \in \text{PSD}(\mathcal{H})$. This defines a partial order on $\mathcal{L}(\mathcal{H})$ that is sometimes called the *Loewner order*. For example, $A \geq 0$ means simply that A is positive semidefinite, while $A \leq I$ states that $I - A$ is positive semidefinite, i.e., A is Hermitian and has eigenvalues less or equal to one. See Exercise 1.9 for more detail.

Remark 1.5 (Operators vs. numbers). It is useful to think of the Hermitian operators as the operator counterpart of the real numbers \mathbb{R} , and the PSD operators as the operator counterpart of the nonnegative numbers $\mathbb{R}_{\geq 0}$. In fact, this is exactly what you obtain for $\mathcal{H} = \mathbb{C}$. For example, the characterization $A = B^\dagger B$ of PSD operators generalizes the statement that the nonnegative numbers are precisely the absolute values squared of arbitrary complex numbers: $a \in \mathbb{R}_{\geq 0}$ if and only if $a = \bar{b}b = |b|^2$ for some $b \in \mathbb{C}$.

For Hermitian operators, one can construct new operators from old ones by applying an arbitrary function to the eigenvalues while keeping the eigenvectors the same. This is also known as a ‘functional calculus’, and it is formally defined as follows:

Definition 1.6 (Functions of Hermitian operators). Let $f: D \rightarrow \mathbb{R}$ be an arbitrary function where $D \subseteq \mathbb{R}$. For any Hermitian operator A with eigendecomposition $A = \sum_{j=1}^d a_j |\psi_j\rangle\langle\psi_j|$ and all eigenvalues $a_j \in D$, we define³

$$f(A) := \sum_{j=1}^d f(a_j) |\psi_j\rangle\langle\psi_j|.$$

Moreover, if $g: D \rightarrow \mathbb{R}$ is another function as above then $f(A)g(A) = (fg)(A)$, where fg is the pointwise product of f and g (i.e., $(fg)(x) = f(x)g(x)$ for $x \in D$). Similarly, if $g: f(D) \rightarrow \mathbb{R}$ then $g(f(A)) = (g \circ f)(A)$, where $g \circ f$ is the composition (i.e., $(g \circ f)(x) = g(f(x))$ for $x \in D$).

For example, if $f(x) = x^n$ is the n -th power function on \mathbb{R} , then

$$f(A) = A^n = \underbrace{A \cdots A}_{n \text{ times}}$$

³This definition does not depend on the choice of eigendecomposition. Indeed, if we write A in the form (1.7), which does not depend on any choices, then $f(A) = \sum_a f(a)P_a$.

so this agree with the usual definition of the n -th power of an operator. For a more interesting example, take the square-root function $f(x) = \sqrt{x}$ which is only defined on $\mathbb{R}_{\geq 0}$. We can use it to define the *square root* of a PSD operator A by $\sqrt{A} := A^{1/2} := f(A)$. That is, if $A = \sum_{j=1}^d a_j |\psi_j\rangle\langle\psi_j|$ is an eigendecomposition of A then the square root is given by

$$\sqrt{A} := A^{1/2} := \sum_{j=1}^d \sqrt{a_j} |\psi_j\rangle\langle\psi_j|. \quad (1.10)$$

This operator is again PSD and, clearly, $(\sqrt{A})^2 = A$. Moreover, \sqrt{A} is the unique PSD operator with the property that it squares to A . You can show this in Exercise 1.10.

Warning: In general, it is *not* true that $\sqrt{AB} = \sqrt{A}\sqrt{B}$ for A, B PSD. Indeed, AB will in general not even be PSD! (Can you find an example?)

1.3 Quantum states

We will now discuss the state space of quantum systems.

Definition 1.7 (State). A *state*, *quantum state*, *density operator*, or *density ‘matrix’* is by definition a positive semidefinite operator with trace one. We denote the set of states on a Hilbert space \mathcal{H} by

$$\mathcal{D}(\mathcal{H}) = \{\rho \in \text{PSD}(\mathcal{H}) \mid \text{Tr}[\rho] = 1\}.$$

Axiom 1.8 (States). The state space of a quantum system with Hilbert space \mathcal{H} is given by $\mathcal{D}(\mathcal{H})$.

This definition might look surprising to you if you have taken a quantum mechanics course (or a popular quantum computing course). Aren’t quantum states described by unit vectors in Hilbert space? In fact, $\rho = |\psi\rangle\langle\psi|$ is a quantum state for any unit vector $|\psi\rangle \in \mathcal{H}$. Such states are called *pure*, as in the following definition.

Definition 1.9 (Pure and mixed states). A quantum state ρ is called *pure* if it is of the form $\rho = |\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle \in \mathcal{H}$. Quantum states that are not pure are called *mixed*. In particular, we always have a *maximally mixed state*, defined by $\tau = \frac{I}{d}$, where I denotes the identity operator and $d = \dim \mathcal{H}$.

Note that the pure states $\rho = |\psi\rangle\langle\psi|$ are precisely the states of rank one, or equivalently, the states that have a single nonzero eigenvalue (which is then necessarily equal to 1). Note that the pure states are in one-to-one correspondence with unit vectors, up to an overall phase (i.e., $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ give rise to the same pure state). **Indeed, you might know that an overall phase of a state vector is unobservable** – in this sense state vectors are more redundant than density matrices. Mathematically, **the space of pure states is a projective space**. For convenience we will sometimes say things like “the pure state $|\psi\rangle$ ”, even though this is slightly imprecise and we should rather say “the pure state $|\psi\rangle\langle\psi|$ ”. You can explore the difference between pure and mixed states in Exercise 1.12.

What is the meaning of mixed states? Suppose $\{p_j, \rho_j\}$ is an *ensemble* of quantum states – i.e., (p_j) is a probability distribution and the ρ_j are states. Such an ensemble might describe a device that outputs state ρ_j with probability p_j . How should we describe the average output of such a device? Clearly, we should take the average

$$\rho = \sum_j p_j \rho_j. \quad (1.11)$$

Then it is easy to see that ρ is again a quantum state. But note that even if the ρ_j are all pure, ρ will in general be mixed, see Exercise 1.13. Thus, mixed states not only arise naturally, but they are in fact crucial to describe such common situations.

Mixed states also allow us to describe probability distributions using quantum states. Let

$$P(\Sigma) := \left\{ (p_x)_{x \in \Sigma} \in \mathbb{R}_{\geq 0}^{\Sigma} : \sum_{x \in \Sigma} p_x = 1 \right\} \cong \left\{ p : \Sigma \rightarrow \mathbb{R}_{\geq 0} : \sum_{x \in \Sigma} p(x) = 1 \right\} \quad (1.12)$$

denote the set of all probability distributions on a finite set Σ (depending on the context we will think of p as a vector or as a function). Throughout these lecture notes all probability distributions will have finitely many outcomes unless explicitly stated otherwise. Then we make the following definition.

Definition 1.10 (Classical states). Let Σ be a finite set. A quantum state ρ on $\mathcal{H} = \mathbb{C}^{\Sigma}$ is called *classical* if it is of the form

$$\rho = \sum_{x \in \Sigma} p_x |x\rangle\langle x| \quad (1.13)$$

where $(p_x)_{x \in \Sigma} \in P(\Sigma)$ is an arbitrary probability distribution. In other words, the classical states are precisely those that are diagonal with respect to the standard basis.

For example, the classical states of a qubit $\mathcal{H} = \mathbb{C}^2$ are of the form

$$\rho = p_0 |0\rangle\langle 0| + p_1 |1\rangle\langle 1| = \begin{pmatrix} p_0 & 0 \\ 0 & p_1 \end{pmatrix},$$

where $p_0, p_1 \geq 0$ and $p_0 + p_1 = 1$.

Remark 1.11 (Why not restrict to pure states?). There is a more general and somewhat philosophical point that is worth mentioning. In quantum computing, we usually start out with a pure initial state (say, all qubits are initialized in $|0\rangle$ or $|1\rangle$), apply unitary operations, and only at the very end carry out a measurement. This allows one to exclusively work with unit vectors $|\psi\rangle$ rather than with density operators ρ .

In contrast, in quantum information theory we often deal with uncertainty and noise. In this situation, mixed states arise naturally, as we already saw above. Similarly, instead of only dealing with unitary operations, we will use the more general notion of a quantum channel, which can send pure states to mixed states. This will be introduced in Lecture 4.

(In physics language, this is the distinction between ‘closed’ and ‘open’ quantum systems.)

However, it is very important to point out that both formalisms are completely equivalent.

For example, one of the key points of Lecture 2 will be that we can always think of mixed states as describing a part of a larger system that is in a pure state. Similarly, in Lecture 5 we will see that quantum channels can be understood as the effect of unitary operations on a larger system.

To understand the structure of quantum states better, note that they are (by definition) Hermitian operators, so the spectral theorem (Theorem 1.3) applies. Accordingly, we can write any quantum state in the form

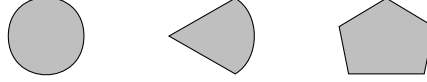
$$\rho = \sum_{j=1}^d p_j |\psi_j\rangle\langle \psi_j|, \quad (1.14)$$

with eigenvalues $p_j \in \mathbb{R}$, orthonormal eigenvectors $|\psi_j\rangle$, and $d = \dim \mathcal{H}$. Since ρ is positive semidefinite, all $p_j \geq 0$, and since $\text{Tr}[\rho] = 1$, $\sum_{j=1}^d p_j = 1$. We record this useful result.

Lemma 1.12. *For any quantum state, the collection of eigenvalues forms a probability distribution.*

If we put $\rho_j = |\psi_j\rangle\langle\psi_j|$, then Eq. (1.14) is precisely of the form of Eq. (1.11). This shows that **any quantum state can be understood as the average of an ensemble of (pairwise orthogonal) pure states**. But careful: In general there are many different ways of writing a given quantum state as an ensemble. In particular, if someone hands you a quantum state in the form of Eq. (1.11), the ρ_j might have nothing to do with the eigendecomposition (since the ρ_j need neither be pure nor orthogonal). You can explore this in Exercise 1.13.

Formally, the fact that for any ensemble $\{p_j, \rho_j\}$ of quantum states the average (1.11) is again a quantum state means that $D(\mathcal{H})$ is convex. The following picture shows three convex sets:



The first has a ‘round’ boundary, while the second also has some ‘flat’ sides, and the third is a ‘polygon’. What does the convex set of quantum states look like? To get more intuition we consider the case of a single quantum bit.

1.4 States of a single quantum bit: Bloch ball

In this section we will study the geometry of $D(\mathbb{C}^2)$ – the state space of a single qubit. We start by observing that the four *Pauli matrices*

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1.15)$$

are linearly independent and form a basis of the real vector space of Hermitian 2×2 -matrices.⁴ Indeed, a 2×2 matrix is Hermitian iff its diagonal entries are real and its top-right entry is the complex conjugate of its bottom-left entry. Note that X, Y, Z are traceless, while $\text{Tr}[I] = 2$. As a consequence, we see that

$$H = \frac{1}{2}(I + xX + yY + zZ) = \frac{1}{2} \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}, \quad \text{where } \vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3,$$

is the most general form of a Hermitian 2×2 -matrix with $\text{Tr}[H] = 1$. The vector \vec{r} is called the *Bloch vector* of H , and so far it is completely arbitrary.

When is H a quantum state? We need to determine that H is PSD, i.e., has nonnegative eigenvalues. Since $\text{Tr}[H] = 1$, its eigenvalues are given by $\{\lambda, 1 - \lambda\}$ for some $\lambda \in \mathbb{R}$. A moments thought shows that $\lambda \geq 0$ and $1 - \lambda \geq 0$ if and only if $\lambda(1 - \lambda) \geq 0$ (since λ and $1 - \lambda$ cannot both be negative). But note that this product can be computed by the determinant of H :

$$\begin{aligned} \lambda(1 - \lambda) &= \det(H) = \frac{1}{4}((1+z)(1-z) - (x+iy)(x-iy)) \\ &= \frac{1}{4}(1 - x^2 - y^2 - z^2) = \frac{1}{4}(1 - \|\vec{r}\|^2). \end{aligned} \quad (1.16)$$

Thus, H is a quantum state if and only if $\|\vec{r}\| \leq 1$. Thus we have shown that the state space of a qubit can be identified with the unit ball in \mathbb{R}^3 . This is known as the *Bloch ball* and it is clearly convex, in agreement with our prior discussion.

⁴The Pauli matrices satisfy the following properties which are extremely useful: They each square to the identity, $X^2 = Y^2 = Z^2 = I$, and we have the ‘cyclic’ identity $XYZ = iI$, which implies $XY = -YX = iZ$ etc.

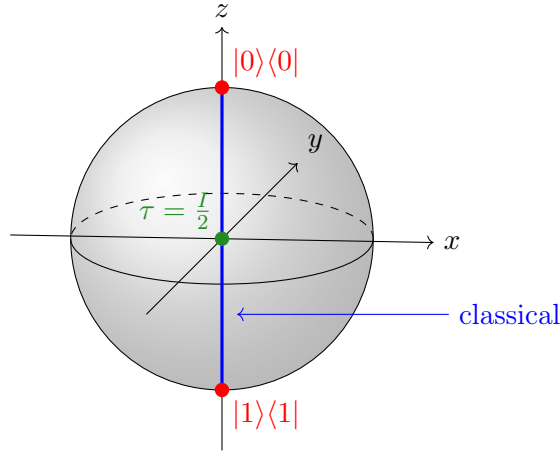
When is H a *pure* state? This is the case precisely when $\lambda(1 - \lambda) = 0$ (one eigenvalue is zero and the other is one). By Eq. (1.16), the latter means that $\|\vec{r}\| = 1$, i.e., \vec{r} is contained in the boundary of the ball – the unit sphere – which is called the *Bloch sphere*. We summarize:

Lemma 1.13 (Bloch ball). *Any qubit state $\rho \in D(\mathbb{C}^2)$ can be written in the form*

$$\rho = \frac{1}{2} (I + r_x X + r_y Y + r_z Z), \quad (1.17)$$

where $\vec{r} = \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}$ is an arbitrary vector of norm $\|\vec{r}\| \leq 1$. Moreover, ρ is pure if and only if $\|\vec{r}\| = 1$.

The following figure visualizes the Bloch ball and some important features that we discuss now:



The north and south poles have Bloch vectors

$$\vec{r}_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{r}_1 = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \quad (1.18)$$

which correspond via Eq. (1.17) to the pure states

$$|0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The blue line segment between north and south pole corresponds precisely to the classical states

$$\rho = p|0\rangle\langle 0| + (1 - p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1 - p \end{pmatrix}. \quad (1.19)$$

In particular, the origin of the Bloch ball corresponds to the maximally mixed qubit state $\tau = I/2$, with Bloch vector $\vec{r} = 0$.

The ‘west and east poles’ of the Bloch ball

$$\vec{r}_+ = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{r}_- = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}. \quad (1.20)$$

must also correspond to pure states. We claim that they correspond to the so-called *Hadamard basis* states, which are defined by

$$|\pm\rangle := \frac{1}{\sqrt{2}} (|0\rangle \pm |1\rangle) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}. \quad (1.21)$$

Indeed,

$$|+\rangle\langle+| = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad |-\rangle\langle-| = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad (1.22)$$

are precisely the states that via Eq. (1.17) correspond to \vec{r}_{\pm} .

Can you figure out the pure states that correspond to the ‘front’ and ‘back poles’ of the Bloch ball?

You can learn more about the Bloch sphere in Exercise 1.14. In particular, you may prove there the useful fact that the components of the Bloch vector \vec{r} can be calculated by

$$x = \text{Tr}[X\rho], \quad y = \text{Tr}[Y\rho], \quad z = \text{Tr}[Z\rho].$$

In Exercise 1.15 you can design a procedure for estimating unknown qubit state by determining its Bloch vector. This is an eminently practical task faced by experimental quantum physicists on a daily basis. Naturally this will also require a way to extract information from a quantum system. We will discuss this next.

1.5 Measurements

The discussion so far has been slightly formal, since we did not yet discuss the rules for getting information out of a quantum system. For this we need the notion of a measurement.

Definition 1.14 (Measurement). A *measurement* or *POVM* (short for positive operator valued measure) on a Hilbert space \mathcal{H} with outcomes in some finite set Ω is a function

$$\mu: \Omega \rightarrow \text{PSD}(\mathcal{H}) \quad \text{such that} \quad \sum_{x \in \Omega} \mu(x) = I. \quad (1.23)$$

If all $\mu(x)$ are orthogonal projections then we say that μ is *projective*.

When we apply a measurement μ to a quantum system in some state ρ , the outcome will be an element $x \in \Omega$. We will often draw pictures such as the following to illustrate this situation:

$$\rho \longrightarrow \boxed{\begin{array}{c} \curvearrowright \\ \mu \end{array}} \Longrightarrow x \in \Omega \quad (1.24)$$

By convention, single lines correspond to quantum systems, while double lines denote classical values.

Importantly, the measurement outcome x will in general be *random* (even if we know μ and ρ precisely). Indeed, quantum mechanics is a *probabilistic* theory. How can we calculate the probability of each measurement outcome? This is known as the *Born’s rule*, which is the content of our next axiom.

Axiom 1.15 (Born's rule). If we measure a quantum system in state $\rho \in \mathcal{D}(\mathcal{H})$ using a measurement μ , then the probability of outcome $x \in \Omega$ is given by *Born's rule*:

$$\Pr(\text{outcome } x) = \text{Tr}[\mu(x)\rho] \quad (1.25)$$

Let us verify that Born's rule in Eq. (1.25) defines a probability distribution. Indeed, $\text{Tr}[\mu(x)\rho] \geq 0$ (since by Lemma 1.4 the trace of a product of two PSD operators is always nonnegative), and

$$\sum_{x \in \Omega} \text{Tr}[\mu(x)\rho] = \text{Tr}\left[\sum_{x \in \Omega} \mu(x)\rho\right] = \text{Tr}[\rho] = 1,$$

where we first use linearity, then Eq. (1.23), and finally that quantum states have trace one. Thus, Born's rule is well-defined.

Remark 1.16 (After the measurement?). You may wonder what happens to the quantum state after the measurement – perhaps you remember from your quantum mechanics course that the state ‘collapses’ into a post-measurement state, or something similar. At this point we do *not* want to make any statement about this. For now we will simply assume that the quantum state is ‘gone’ after the measurement and all that remains is the measurement outcome – as in Figure (1.24).

How can we construct measurements? One way is by using an orthonormal basis. For a qubit, the standard basis of \mathbb{C}^2 is $\{|0\rangle, |1\rangle\}$ and the corresponding *standard basis measurement* is defined by

$$\mu_{\text{Std}}: \{0, 1\} \rightarrow \text{PSD}(\mathbb{C}^2), \quad x \mapsto |x\rangle\langle x|. \quad (1.26)$$

This clearly defines a projective measurement, since the $|x\rangle\langle x|$ are projections and

$$|0\rangle\langle 0| + |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Another basis of \mathbb{C}^2 is the Hadamard basis $\{|+\rangle, |-\rangle\}$ defined in Eq. (1.21). The corresponding measurement is called the *Hadamard basis measurement*. It is given by

$$\mu_{\text{Had}}: \{0, 1\} \rightarrow \text{PSD}(\mathbb{C}^2), \quad \mu_{\text{Had}}(0) = |+\rangle\langle +|, \quad \mu_{\text{Had}}(1) = |-\rangle\langle -|. \quad (1.27)$$

This also defines a projective measurement, as follows from Eq. (1.22).

In fact, by Eq. (1.3), the same procedure works for any quantum system and any orthonormal basis of that system. Basis measurements are so useful that they get their own definition.

Definition 1.17 (Basis measurement). For an arbitrary orthonormal basis $\{|\psi_x\rangle\}_{x \in \Omega}$ of \mathcal{H} , the corresponding *basis measurement* is the projective measurement

$$\mu: \Omega \rightarrow \text{PSD}(\mathcal{H}), \quad \mu(x) = |\psi_x\rangle\langle \psi_x|.$$

If we take \mathbb{C}^d or \mathbb{C}^Σ with its standard basis, so $\mu(x) = |x\rangle\langle x|$, this is called the *standard basis measurement*.

For a basis measurement, Born's rule can be rewritten as follows:

$$\Pr(\text{outcome } x \mid \text{state } \rho) = \text{Tr}[\mu(x)\rho] = \text{Tr}[|\psi_x\rangle\langle \psi_x|\rho] = \langle \psi_x | \rho | \psi_x \rangle, \quad (1.28)$$

where the last equality follows from the cyclic property of the trace ($\text{Tr}[ABC] = \text{Tr}[BCA]$), or by evaluating the trace in the basis $\{|\psi_x\rangle\}$. *In particular, if you perform a basis measurement and $\rho = |\psi_x\rangle\langle\psi_x|$ is one of the basis state then the outcome will be x with certainty!*

For a standard basis measurement, we would write Eq. (1.28) as

$$\text{Pr}(\text{outcome } x \mid \text{state } \rho) = \langle x | \rho | x \rangle. \quad (1.29)$$

If $\rho = |\Psi\rangle\langle\Psi|$ is a pure state, then Eq. (1.28) is the same as $|\langle\psi_x|\Psi\rangle|^2$, while Eq. (1.29) can also be written as $|\langle x|\Psi\rangle|^2 = |\Psi_x|^2$, where Ψ_x is the coefficient of $|x\rangle$ when expanding $|\Psi\rangle = \sum_x \Psi_x |x\rangle$ in the standard basis. (If you are attending Ronald de Wolf's quantum computing course then this formula will look very familiar to you!)

As a concrete example, suppose that we have a qubit in state $\rho = |0\rangle\langle 0|$ and we carry out the standard basis measurement in Eq. (1.26). Then the probability of outcome '0' is given by

$$p_{\text{Std}}(0) = \langle 0 | \rho | 0 \rangle = |\langle 0 | 0 \rangle|^2 = 1,$$

i.e., the measurement yields outcome '0' with certainty (as one might expect). In contrast, if we perform the Hadamard basis measurement in Eq. (1.27) then the probability of outcome '0' (corresponding to the basis vector $|+\rangle$) is given by

$$p_{\text{Had}}(0) = \langle + | \rho | + \rangle = |\langle + | 0 \rangle|^2 = \frac{1}{2},$$

so both outcomes are equally likely. Similarly, if $\rho = |1\rangle\langle 1|$ then the standard basis measurement always yields outcome '1', while the Hadamard basis measurement is again completely random. This shows that the standard and the Hadamard basis are in some way 'complementary' – if our qubit is in a standard basis state then doing a Hadamard basis measurement reveals no information at all.

In Exercise 1.17, you can show an *uncertainty relation*, which establishes a precise quantitative tradeoff between the uncertainty in the two measurement outcomes. In particular, there exists no quantum state for which both outcomes are certain.

Are all measurements projective or even basis measurements? Certainly not! Exercise 1.20 discusses the so-called 'pretty good measurements' as a concrete example of a family of measurements that are in general not projective.

Remark 1.18 (Quantum theory as non-commutative probability theory). It is instructive to think of the formalism of quantum information as a non-commutative generalization of ordinary probability theory. This can be made precise in many ways, for example, using operator algebras. Here is a very concrete way. We saw that probability distributions can be embedded into quantum states by associating to each distribution the corresponding "classical" state (Definition 1.10):

$$P(\Sigma) \rightarrow D(\mathbb{C}^\Sigma), \quad p = (p_x)_{x \in \Sigma} \mapsto \rho = \sum_{x \in \Sigma} p_x |x\rangle\langle x|.$$

We can also ask the converse question: How can we get a probability distribution out of a quantum state? This is exactly achieved by measurements (Definition 1.14), since for any measurement $\mu: \Omega \rightarrow \text{PSD}(\mathcal{H})$ we obtain a map

$$D(\mathcal{H}) \rightarrow P(\Omega), \quad \rho \mapsto p = (p_x)_{x \in \Omega}, \quad \text{where } p(x) = \text{Tr}[\mu(x)\rho].$$

In fact, *any* map $D(\mathcal{H}) \rightarrow P(\Omega)$ that is compatible with convex combinations is necessarily of this form, see Exercise 1.19! This gives a nice and purely mathematical motivation for defining measurements as we did.

Remark 1.19 (Measurements vs. observables). If you ever attended a course in quantum mechanics, you may know the notion of an *observable*, which is another way to think about measurements. In fact, observables correspond precisely to *projective* measurements with outcomes in the reals (i.e., $\Omega \subseteq \mathbb{R}$). In Exercise 1.21 you can explore this further.

1.6 Exercises

- 1.1 **Dirac notation quiz:** In the Dirac notation, every vector is written as a ‘ket’ $|\psi\rangle$ and every linear functional is written as a ‘bra’ $\langle\psi| = |\psi\rangle^\dagger$, where † denotes the adjoint. One can think of kets as column vectors and bras as row vectors. Hence, if $|\psi\rangle$ is a column vector, then $\langle\psi|$ denotes the row vector obtained by taking the *conjugate transpose* of the column vector.

Let $|\psi\rangle$ and $|\phi\rangle$ be vectors in \mathbb{C}^n and A an $n \times n$ matrix. Which of the following expressions are syntactically correct? For those that do, what kind of object do they represent (e.g., numbers, vectors, ...)? Can you write them using ‘ordinary’ notation?

- | | | | |
|-----------------------------------|------------------------------------|-----------------------------------|---|
| (a) $ \psi\rangle + \langle\phi $ | (d) $\langle\psi A$ | (g) $ \psi\rangle\langle\phi A$ | (j) $\langle\psi A \phi\rangle + \langle\psi \phi\rangle$ |
| (b) $ \psi\rangle\langle\phi $ | (e) $\langle\psi A + \langle\psi $ | (h) $ \psi\rangle A \langle\phi $ | (k) $\langle\psi \phi\rangle\langle\psi $ |
| (c) $A\langle\psi $ | (f) $ \psi\rangle\langle\phi + A$ | (i) $\langle\psi A \phi\rangle$ | (l) $\langle\psi \phi\rangle A$ |

- 1.2 **Trace vs. inner product:** Let $A = |\psi\rangle\langle\psi|$, $B = |\phi\rangle\langle\phi|$ for $|\psi\rangle, |\phi\rangle \in \mathcal{H}$. Verify that $\text{Tr}[AB] = |\langle\psi|\phi\rangle|^2$.

1.3 Eigenvalue basics:

- Suppose that $H = \sum a_i |\psi_i\rangle\langle\psi_i|$. Show that when the $|\psi_i\rangle$ are orthogonal then they are eigenvectors of H . Show that when the $|\psi_i\rangle$ are orthonormal then the numbers a_i are eigenvalues of H . Are these assumption necessary?
- Consider the matrix $H = |0\rangle\langle 0| + |+\rangle\langle +|$, where $|+\rangle$ is defined as in Eq. (1.21). Compute its eigenvectors and eigenvalues.
- Compute the eigenvalues and eigenvectors of the matrix

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Hint: You can avoid computing the determinant of a 4 by 4 matrix!

- 1.4 **Projections:** A Hermitian operator P is an (*orthogonal*) *projection* if it satisfies either of the following two properties: (i) any eigenvalue of P is 0 or 1, (ii) $P^2 = P$. Show that these two properties are equivalent.

- 1.5 **Criterion for positive semidefiniteness:** Prove Lemma 1.4.

- 1.6 **PSD factorization:** Find a PSD matrix A , such that $A = B^\dagger B$, for some $B \in L(\mathcal{H}, \mathcal{K})$ with $\dim \mathcal{K} < \dim \mathcal{H}$. Also, find a PSD matrix A , for which such \mathcal{K} and B do not exist. In general, how large does $\dim \mathcal{K}$ have to be?

- 1.7 **Positive semidefinite operators:** For all $Q \in \text{PSD}(\mathcal{H})$, show that:

- If $Q \in \text{PSD}(\mathcal{H})$ and $A \in L(\mathcal{H})$ then $A^\dagger Q A \in \text{PSD}(\mathcal{H})$
- If $Q \in \text{PD}(\mathcal{H})$ then $Q^{-1} \in \text{PD}(\mathcal{H})$.

- 1.8 **Positive entries but not PSD:** Find an example of a Hermitian matrix A with nonnegative entries which is *not* a positive semidefinite matrix.

- 1.9 **Positive semidefinite order:** Given two operators A and B , we write $A \leq B$ if the operator $B - A$ is positive semidefinite. Show that the following three conditions are equivalent:

- (a) $0 \leq A \leq I$.
- (b) A is Hermitian and has eigenvalues in $[0, 1]$.
- (c) $\langle \psi | A | \psi \rangle \in [0, 1]$ for every unit vector $|\psi\rangle \in \mathcal{H}$.

1.10 **Uniqueness of the PSD square root:** Let $A \in \text{PSD}(\mathcal{H})$. Show that if $B \in \text{PSD}(\mathcal{H})$ is such that $B^2 = A$, then $B = \sqrt{A}$. As such, the PSD square root is unique.

1.11 **Convexity:**

- (a) Show that $\text{PSD}(\mathcal{H})$ and $\text{PD}(\mathcal{H})$ are convex and closed under multiplication by $\mathbb{R}_{\geq 0}$ (i.e., *convex cones*).
- (b) Show that $\text{D}(\mathcal{H})$ is convex.
- (c) An *extreme point* of a convex set S is an element $z \in S$ that cannot be written as a proper convex combination (i.e., $z \neq px + (1-p)y$ for any $p \in (0, 1)$ and $x \neq y \in S$). Show that the extreme points of $\text{D}(\mathcal{H})$ are precisely the pure states.

1.12 **Pure states and unit vectors:**

- (a) Consider the following states:

$$\rho_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \quad \rho_3 = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Verify for each state whether it is pure, and if so, find $|\psi_i\rangle$ such that $\rho_i = |\psi_i\rangle\langle\psi_i|$.

- (b) Let $|\psi\rangle = a|0\rangle + b|1\rangle \in \mathbb{C}^2$ with a and b are complex numbers with $|a|^2 + |b|^2 = 1$. Compute the associated density matrix $|\psi\rangle\langle\psi|$.

1.13 **States vs. ensembles:** Consider the ensemble consisting of the qubit states $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ occurring with probabilities $\frac{2}{3}$ and $\frac{1}{3}$, respectively.

- (a) Compute the quantum state ρ corresponding to this ensemble. Is ρ pure or mixed?
- (b) Find an ensemble consisting of *three* distinct pure states (with non-zero probabilities) that corresponds to the same state ρ . Why can such an ensemble not correspond to an eigendecomposition?

1.14 **Bloch sphere:** Recall from Lemma 1.13 that the state ρ of a single qubit can be parameterized by the Bloch vector $\vec{r} \in \mathbb{R}^3$, $\|\vec{r}\| \leq 1$. Namely:

$$\rho = \frac{1}{2}(I + r_x X + r_y Y + r_z Z).$$

- (a) Show that $r_x = \text{Tr}[\rho X]$, $r_y = \text{Tr}[\rho Y]$, and $r_z = \text{Tr}[\rho Z]$.
- (b) Let σ be another qubit state, with Bloch vector \vec{s} . Verify that $\text{Tr}[\rho\sigma] = \frac{1}{2}(1 + \vec{r} \cdot \vec{s})$.
- (c) Let $\{|\psi_i\rangle\}_{i=0,1}$ denote an arbitrary orthonormal basis of \mathbb{C}^2 . Let \vec{r}_i denote the Bloch vector of $|\psi_i\rangle\langle\psi_i|$ for $i \in \{0, 1\}$. Show that $\vec{r}_0 = -\vec{r}_1$.
- (d) Let $\mu: \{0, 1\} \rightarrow \text{PSD}(\mathbb{C}^2)$ denote the basis measurement corresponding to the basis in part (c). Show that the probability of obtaining outcome $i \in \{0, 1\}$ when measuring ρ using μ is given by $\frac{1}{2}(1 + \vec{r} \cdot \vec{r}_i)$. How can you visualize this fact on the Bloch sphere?

Hint: $X^2 = Y^2 = Z^2 = I$, while $XY = iZ$, $YZ = iX$, $ZX = iY$ are traceless.

1.15 **Estimating an unknown qubit state:** Your goal in this problem is to come up with a procedure for estimating an unknown qubit state ρ .

- (a) Consider the following function:

$$\mu: \{x, y, z\} \times \{0, 1\} \rightarrow \text{PSD}(\mathbb{C}^2), \quad \mu(a, b) = \frac{I + (-1)^b \sigma_a}{6},$$


where $\sigma_x = X$, $\sigma_y = Y$, and $\sigma_z = Z$ are the three Pauli matrices. Show that μ is a valid measurement. Is it projective?

- (b) Show that the probabilities of outcomes when measuring ρ using μ are given by

$$p(a, b) = \frac{1 + (-1)^b r_a}{6},$$

where $\vec{r} = \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}$ is the Bloch vector of ρ . *Hint: See the hint in Exercise 1.14.*

- (c) Now suppose an experimentalist prepares the state ρ and carries out the measurement μ not just once but a large number of times. Afterwards they send you a list $\{(a_i, b_i)\}_{i=1}^n$, where n is the number of repetitions and (a_i, b_i) is the measurement outcome in the i -th repetition. Assume that n is very large. How can you obtain a good estimate of the quantum state ρ from this data?

- 1.16  **Practice:** In Exercise 1.15, you discussed how to estimate an unknown qubit state ρ by performing the following measurement on many copies of ρ :

$$\mu: \{x, y, z\} \times \{0, 1\} \rightarrow \text{PSD}(\mathbb{C}^2), \quad \mu(a, b) = \frac{I + (-1)^b \sigma_a}{6},$$

where $\sigma_x = X$, $\sigma_y = Y$, and $\sigma_z = Z$ are the Pauli matrices.

The file `01-measurement-outcomes.txt` on the course homepage contains $N = 100\,000$ measurement outcomes produced in this way (one per row). Give an estimate for the unknown state ρ .

- 1.17 **Uncertainty relation:** Given a measurement $\mu: \{0, 1\} \rightarrow \text{PSD}(\mathcal{H})$ with two outcomes and a state $\rho \in \text{D}(\mathcal{H})$, define the *bias* by

$$\beta(\rho) = |\text{Tr}[\mu(0)\rho] - \text{Tr}[\mu(1)\rho]|.$$

- (a) Show that $\beta \in [0, 1]$, that $\beta = 1$ iff the measurement outcome is certain, and that $\beta = 0$ iff both outcomes are equally likely (for the given measurement and state).

In class, we discussed how to measure a qubit in the standard basis $|0\rangle, |1\rangle$ and in the Hadamard basis $|+\rangle, |-\rangle$. Let β_{Std} and β_{Had} denote the bias for these two measurements.

- (b) Compute $\beta_{\text{Std}}(\rho)$ and $\beta_{\text{Had}}(\rho)$ in terms of the Bloch vector of the qubit state ρ .
(c) Show that $\beta_{\text{Std}}^2(\rho) + \beta_{\text{Had}}^2(\rho) \leq 1$. Why is this called an *uncertainty relation*?

- 1.18 **Functionals:** Let $\lambda: \text{L}(\mathcal{H}) \rightarrow \mathbb{C}$ be a linear function.

- (a) Show that there exists a unique $X \in \text{L}(\mathcal{H})$ such that $\lambda[M] = \text{Tr}[X^\dagger M]$ for all $M \in \text{L}(\mathcal{H})$.
(b) Now assume that $\lambda[M] \geq 0$ for all $M \geq 0$. What does this imply for X ?

- 1.19 **Measurements:** Let \mathcal{H} be a Hilbert space, Ω a finite set, and $M: \text{D}(\mathcal{H}) \rightarrow \text{P}(\Omega)$ an arbitrary map that preserves convex combinations, i.e., $M(\sum_j p_j \rho_j) = \sum_j p_j M(\rho_j)$ for any ensemble $\{p_j, \rho_j\}$ of quantum states. Show that there exists a measurement $\mu: \Omega \rightarrow \text{PSD}(\mathcal{H})$ such that $p = M(\rho)$ is given by $p(x) = \text{Tr}[\mu(x)\rho]$ for all $x \in \Omega$ and $\rho \in \text{D}(\mathcal{H})$.

1.20 **Pretty good measurement:** Let $\rho_1, \dots, \rho_n \in \mathcal{D}(\mathcal{H})$ be quantum states with the property that $I \in \text{span} \{\rho_1, \dots, \rho_n\}$.

- (a) Show that $A := \sum_{j=1}^n \rho_j$ is positive definite.
- (b) Define $\mu: \{1, \dots, n\} \rightarrow \mathcal{L}(\mathcal{H})$ by $\mu(j) = A^{-1/2} \rho_j A^{-1/2}$. Show that μ is a measurement.

The measurement μ is called the ‘pretty good measurement’. See also Exercise 4.3.

1.21 **▲ Observables (for those of you who have taken a quantum mechanics course):** In this problem we discuss the relationship between measurements as defined in Definition 1.14 and the notion of ‘observables’ that you might be familiar with from an introductory quantum mechanics course. An *observable* on a quantum system is by definition a Hermitian operator on the corresponding Hilbert space \mathcal{H} .

- (a) Let $\mu: \Omega \rightarrow \text{PSD}(\mathcal{H})$ be a *projective* measurement with outcomes in the real numbers, i.e., a finite subset $\Omega \subseteq \mathbb{R}$. Show that the following operator is an observable:

$$\mathcal{O} = \sum_{x \in \Omega} x \mu(x) \tag{1.30}$$

In fact, this is always an eigendecomposition, but you need not prove this.

- (b) Argue that, conversely, any observable can be written as in Eq. (1.30) for some suitable μ .
- (c) Now suppose that the system is in state ρ and we perform the measurement μ . Show that the *expectation value* of the measurement outcome is given by $\text{Tr}[\rho \mathcal{O}]$.

For a pure state $\rho = |\psi\rangle\langle\psi|$, this can also be written as $\langle\psi|\mathcal{O}|\psi\rangle$. Do you recognize these formulas from your quantum mechanics class?

- (d) Consider an arbitrary qubit observable $\mathcal{O} = tI + s_x X + s_y Y + s_z Z$. Compute its expectation value in a state with Bloch vector \vec{r} .

Lecture 2

Joint systems, reduced states, purifications

Last week, we mathematically defined quantum states and measurements, and we discussed that probabilities of measurement outcomes are computed by Born's rule [Eq. (1.25)]. We saw that states are described by 'density operators' – positive semidefinite operators with unit trace. A basic distinction is between *pure states* $\rho = |\psi\rangle\langle\psi|$, which correspond to unit vectors in Hilbert space (up to overall phase), and *mixed states*, which cannot be written in this way. We discussed that one motivation for mixed states is that they allow us to model ensembles.

Today, we will see another use for mixed states. If we have a composite system that consists of two or more subsystems, then, even if the overall state is pure, the subsystems are typically described by mixed states (see Eq. (2.13)). This phenomenon is closely related to the notion of *entanglement*, which will be discussed in more detail in Lecture 3.

2.1 Joint or composite systems

Axiom 2.1 (Composing systems). For a quantum system composed of n subsystems, with Hilbert spaces $\mathcal{H}_1, \dots, \mathcal{H}_n$, the overall Hilbert space is given by the tensor product $\mathcal{H} = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n$.

For example, a quantum system comprised of n qubits is described by the Hilbert space

$$\mathcal{H} = \underbrace{\mathbb{C}^2 \otimes \dots \otimes \mathbb{C}^2}_{n \text{ factors}} = (\mathbb{C}^2)^{\otimes n}.$$

This space has a natural product basis

$$|x_1, \dots, x_n\rangle = |x_1\rangle \otimes \dots \otimes |x_n\rangle,$$

indexed by bitstrings $x = (x_1, \dots, x_n) \in \{0, 1\}^n$. We often leave out the commas and write, e.g.,

$$|010\rangle = |0\rangle \otimes |1\rangle \otimes |0\rangle.$$

More generally, if $\mathcal{H}_i = \mathbb{C}^{\Sigma_i}$ for $i = 1, \dots, n$, then

$$\mathcal{H} = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n = \mathbb{C}^{\Sigma_1} \otimes \dots \otimes \mathbb{C}^{\Sigma_n}$$

has a natural product basis $|x_1, \dots, x_n\rangle = |x_1\rangle \otimes \dots \otimes |x_n\rangle$ indexed by tuples $x \in \Sigma = \Sigma_1 \times \dots \times \Sigma_n$. In this way, $\mathcal{H} \cong \mathbb{C}^\Sigma$ are isomorphic.

What are possible states of a joint system? Here is a first class of states:

Definition 2.2 (Product states and correlated states). A state $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n)$ is called a *product state* if

$$\rho = \rho_1 \otimes \cdots \otimes \rho_n. \quad (2.1)$$

for $\rho_i \in \mathcal{D}(\mathcal{H}_i)$, $i = 1, \dots, n$. A state that is not a product state is called *correlated*.

You can think of product states as the quantum generalization of joint probability distributions where the random variables are *independent*, which means that the probability distribution factors as $p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n)$. See Exercise 2.1.

Not all states are product states. Here is an example of a (maximally classically) correlated two-qubit state:

$$\rho = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11| = \frac{1}{2}|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| \otimes |1\rangle\langle 1| = \frac{1}{2} \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \\ & & & 1 \end{pmatrix}. \quad (2.2)$$

To see the middle equality, remember that $|00\rangle = |0\rangle \otimes |0\rangle$, so $|00\rangle\langle 00| = (|0\rangle \otimes |0\rangle)(\langle 0| \otimes \langle 0|) = |0\rangle\langle 0| \otimes |0\rangle\langle 0|$ etc. The right-hand side matrix is with respect to the product basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. Note that Eq. (2.2) is a classical state – corresponding to a probability distribution of two bits which are both equal to 0 or both equal to 1, with 50% probability each. Thus the notion of correlations has nothing to do with quantum mechanics per se.

Remark 2.3 (Tensor product of operators). In Eq. (2.1) we used the tensor product of operators. Let us recall its definition. If $X \in \mathcal{L}(\mathcal{H}_1, \mathcal{K}_1)$ and $Y \in \mathcal{L}(\mathcal{H}_2, \mathcal{K}_2)$ are linear operators, then their tensor product $X \otimes Y$ is a linear operator in $\mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{K}_1 \otimes \mathcal{K}_2)$ defined as follows:

$$(X \otimes Y)(|\psi\rangle \otimes |\phi\rangle) := X|\psi\rangle \otimes Y|\phi\rangle \quad \forall |\psi\rangle \in \mathcal{H}_1, |\phi\rangle \in \mathcal{H}_2. \quad (2.3)$$

Note that this definition is *not* circular – we define the tensor product of operators in terms of the tensor product of vectors. In particular, the matrix entries of $X \otimes Y$ with respect to product bases are given by

$$\langle a, b | X \otimes Y | c, d \rangle = \langle a | X | c \rangle \langle b | Y | d \rangle.$$

Thus, if we think of operators as matrices then $X \otimes Y$ is simply given by the Kronecker product of X and Y .

An important special case is when one of the Hilbert spaces is one-dimensional. E.g., suppose that $\mathcal{H}_2 = \mathbb{C}$. In this case, any vector $|\eta\rangle \in \mathcal{K}_2$ can be identified with an operator $Y \in \mathcal{L}(\mathbb{C}, \mathcal{K}_2)$ (in coordinates: a column vector is a matrix with a single column). Thus, we can think of $X \otimes |\eta\rangle$ as the operator in $\mathcal{L}(\mathcal{H}_1, \mathcal{K}_1 \otimes \mathcal{K}_2)$ that acts as

$$(X \otimes |\eta\rangle)|\psi\rangle = X|\psi\rangle \otimes |\eta\rangle \quad \forall |\psi\rangle \in \mathcal{H}_1. \quad (2.4)$$

(If also $\mathcal{H}_1 = \mathbb{C}$ then we simply recover the tensor product of vectors.)

Similarly, if $\mathcal{K}_2 = \mathbb{C}$ then $Y \in \mathcal{L}(\mathcal{H}_2, \mathbb{C})$ is nothing but a dual vector $\langle \eta | \in \mathcal{H}_2^*$ (in coordinates, a matrix with a single row is the same as a row vector), so we can think of $X \otimes \langle \chi |$ as an operator in $\mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{K}_1)$ that acts as

$$(X \otimes \langle \chi |)(|\psi\rangle \otimes |\phi\rangle) = X|\psi\rangle \underbrace{\langle \chi | \phi \rangle}_{\in \mathbb{C}} = \langle \chi | \phi \rangle X|\psi\rangle \quad \forall |\psi\rangle \in \mathcal{H}_1, |\phi\rangle \in \mathcal{H}_2. \quad (2.5)$$

Since $\langle \chi | \phi \rangle$ is a number, it does not matter if we write it on the left or on the right (by linearity). (If also $\mathcal{K}_1 = \mathbb{C}$ then we recover the tensor product of dual vectors.)

If all this seems confusing to you, you can simply take Eqs. (2.4) and (2.5) as the definition of the tensor product of an operator and a vector (or dual vector).

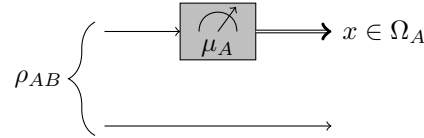
Remark 2.4. Product states are rather special. Indeed, a simple dimension counting argument shows that generic states are correlated. Note that the space of Hermitian operators on a d -dimensional Hilbert space has real dimension d^2 , likewise the space of PSD operators, so the space of density operators has dimension $d^2 - 1$ (the condition that $\text{Tr } \rho = 1$ reduces the dimension by one). For simplicity, suppose that each \mathcal{H}_i is d -dimensional, so that $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ has dimension d^n . Then the space of density operators on \mathcal{H} has dimension $d^{2n} - 1$, which grows exponentially with n . In contrast, the space of product states has dimension $n(d^2 - 1)$, which grows only linearly with n (i.e., much slower).

When writing tensor products of vectors and operators, it can be confusing to remember which tensor factors we are referring to. To simplify our life, we will henceforth adopt a notation that is ubiquitous in the quantum information literature.

Definition 2.5 (Subscripts for subsystems). From now on we will always use *subscripts* to indicate which subsystem some mathematical object refers to. Thus, we write $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ for the Hilbert space of a quantum system comprised of two subsystems A and B , $|\Psi_{AB}\rangle$ for vectors in \mathcal{H}_{AB} , ρ_{AB} for states in $\mathcal{D}(\mathcal{H}_{AB})$, X_B for linear operators on \mathcal{H}_B , and so forth.

2.2 Measurements on subsystems

Now suppose μ_A is a measurement on subsystem A (Definition 1.14), as in the following picture.



How can we calculate the probability of measurement outcomes when the overall system is in state ρ_{AB} ? The answer is given by the following input from physics:

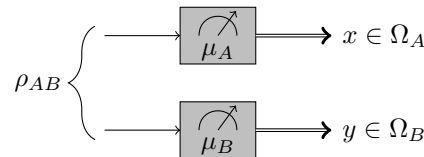
Axiom 2.6 (Measurement on a subsystem, part I). If the joint system is in state ρ_{AB} and we apply a measurement $\mu_A: \Omega_A \rightarrow \text{PSD}(\mathcal{H}_A)$ on A , the probability of outcomes is calculated as follows:

$$\Pr(\text{outcome } x) = \text{Tr}[\rho_{AB}(\mu_A(x) \otimes I_B)]. \quad (2.6)$$

Note that Eq. (2.6) is precisely Born's rule [Eq. (1.25)] for the following measurement on AB :

$$\mu_A \otimes I_B: \Omega \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B), \quad x \mapsto \mu_A(x) \otimes I_B.$$

What if we measure *both* on subsystem A and on subsystem B , as in the following figure?



In this case, we must use the following measurement on AB :

$$\mu_A \otimes \mu_B: \Omega_A \times \Omega_B \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B), \quad (x, y) \mapsto \mu_A(x) \otimes \mu_B(y).$$

Axiom 2.7 (Joint measurement). If the joint system is in state ρ_{AB} and we apply measurement $\mu_A: \Omega_A \rightarrow \text{PSD}(\mathcal{H}_A)$ on A and $\mu_B: \Omega_B \rightarrow \text{PSD}(\mathcal{H}_B)$ on B , then the joint probability of outcomes $x \in \Omega_A$ and $y \in \Omega_B$ is calculated as follows:

$$\Pr(\text{outcomes } x \text{ and } y) = \text{Tr}[\rho_{AB}(\mu_A(x) \otimes \mu_B(y))]. \quad (2.7)$$

It is clear how to generalize these rules to more than two subsystems.

2.3 Partial trace and reduced states

If $p(x, y)$ is a joint probability distribution of two random variables, then we know that the distribution of the first random variable is given by the *marginal* distribution $p(x) = \sum_y p(x, y)$. We are looking for the quantum counterpart of this definition.

Now suppose that we are given a quantum state ρ_{AB} on a quantum system AB composed of two subsystems A and B , with overall Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Which state ρ_A should we use to describe the state of subsystem A alone? In other words, what is the quantum version of a marginal distribution?

Clearly, we would like the state ρ_A to reproduce the statistics of all possible measurements on A (but contain no information about B). In view of Axiom 2.6, this means it should satisfy

$$\text{Tr}[\rho_{AB}(\mu_A(x) \otimes I_B)] \stackrel{!}{=} \text{Tr}[\rho_A \mu_A(x)] \quad (2.8)$$

for all possible Ω , measurements $\mu_A: \Omega \rightarrow \text{PSD}(\mathcal{H}_A)$, and outcomes $x \in \Omega$? How can we find such a ρ_A ? We first give the solution and then verify that it does the job.

Definition 2.8 (Partial trace). The *partial trace* over B is the linear map

$$\text{Tr}_B: \text{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \rightarrow \text{L}(\mathcal{H}_A)$$

defined as follows: For every $M_{AB} \in \text{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$,

$$\text{Tr}_B[M_{AB}] := \sum_b (I_A \otimes \langle b|) M_{AB} (I_A \otimes |b\rangle), \quad (2.9)$$

where $|b\rangle$ is an arbitrary orthonormal basis of \mathcal{H}_B (the result is independent of the choice of basis). The operator $\text{Tr}_B[M_{AB}] \in \text{L}(\mathcal{H}_A)$ is called the *partial trace of M_{AB} over B* .

See Eqs. (2.4) and (2.5) to remind yourself of the meaning of $I_A \otimes \langle b|$ and $I_A \otimes |b\rangle$. Concretely, the matrix entries of $\text{Tr}_B[M_{AB}]$ with respect to an arbitrary orthonormal basis $|a\rangle$ of \mathcal{H}_A are:

$$\langle a| \text{Tr}_B[M_{AB}] |a'\rangle = \sum_b \langle a, b| M_{AB} |a', b\rangle. \quad (2.10)$$

Note that the partial trace not only sends operators to operators – but it is itself a linear operator! Such maps are often called *superoperators*.

The following lemma justifies the terminology *partial trace*:

Lemma 2.9. Let $M_{AB} = X_A \otimes Y_B$, where $X_A \in \mathcal{L}(\mathcal{H}_A)$ and $Y_B \in \mathcal{L}(\mathcal{H}_B)$. Then,

$$\text{Tr}_B[X_A \otimes Y_B] = X_A \text{Tr}[Y_B] = \text{Tr}[Y_B] X_A.$$

Proof. Use Eq. (2.9) to see that

$$\text{Tr}_B[X_A \otimes Y_B] = \sum_b (I_A \otimes \langle b|) (X_A \otimes Y_B) (I_A \otimes |b\rangle) = \sum_b X_A \langle b|Y_B|b\rangle = X_A \text{Tr}[Y_B].$$

Since $\text{Tr}[Y_B]$ is a number, we can also write this as $\text{Tr}[Y_B] X_A$. \square

We now list some important properties of the partial trace.

Lemma 2.10. For all $X_A \in \mathcal{L}(\mathcal{H}_A)$ and $M_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we have:

- (a) $\text{Tr}[M_{AB}(X_A \otimes I_B)] = \text{Tr}[\text{Tr}_B[M_{AB}] X_A]$.
- (b) $\text{Tr}[M_{AB}] = \text{Tr}[\text{Tr}_B[M_{AB}]]$.
- (c) If M_{AB} is positive semidefinite, then so is $\text{Tr}_B[M_{AB}]$.

Proof. (a) Let us give a careful proof of this crucial identity:

$$\begin{aligned} \text{Tr}[M_{AB}(X_A \otimes I_B)] &= \sum_{a,b} (\langle a| \otimes \langle b|) M_{AB} (X_A \otimes I_B) (|a\rangle \otimes |b\rangle) \\ &= \sum_{a,b} (\langle a| \otimes \langle b|) M_{AB} (X_A |a\rangle \otimes |b\rangle) \\ &= \sum_{a,b} \langle a| (I_A \otimes \langle b|) M_{AB} (I_A \otimes |b\rangle) X_A |a\rangle \\ &= \sum_a \langle a| \sum_b (I_A \otimes \langle b|) M_{AB} (I_A \otimes |b\rangle) X_A |a\rangle \\ &= \sum_a \langle a| \text{Tr}_B[M_{AB}] X_A |a\rangle = \text{Tr}[\text{Tr}_B[M_{AB}] X_A]. \end{aligned}$$

Here, we first evaluate the trace in an arbitrary product basis, next we use Eq. (2.3), then Eqs. (2.4) and (2.5), and after moving the sum over b inside we recognize the definition of the partial trace from Eq. (2.9).

- (b) This follows directly from (a) by choosing $X_A = I_A$ (the identity operator).
- (c) To see this, note that if $X_A \in \text{PSD}(\mathcal{H}_A)$ then $X_A \otimes I_B \in \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$, so

$$\text{Tr}[\text{Tr}_B[M_{AB}] X_A] = \text{Tr}[M_{AB}(X_A \otimes I_B)] \geq 0.$$

The equality is (a), and the inequality holds since M_{AB} is PSD using the criterion in Lemma 1.4 (e). This in turn implies that $\text{Tr}_B[M_{AB}]$ is PSD (by the same criterion). \square

The following lemma gives some further useful properties. You can prove it in Exercise 2.4.

Lemma 2.11. For all $X_A \in \mathcal{L}(\mathcal{H}_A)$, $Y_B \in \mathcal{L}(\mathcal{H}_B)$, $M_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $N_{BC} \in \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_C)$, and $O_{ABC} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, we have:

- (a) $\text{Tr}_B[(X_A \otimes I_B)M_{AB}] = X_A \text{Tr}_B[M_{AB}]$ and $\text{Tr}_B[M_{AB}(X_A \otimes I_B)] = \text{Tr}_B[M_{AB}] X_A$.
- (b) $\text{Tr}_B[(I_A \otimes Y_B)M_{AB}] = \text{Tr}_B[M_{AB}(I_A \otimes Y_B)]$.
- (c) $\text{Tr}_{AB}[O_{ABC}] = \text{Tr}_A[\text{Tr}_B[O_{ABC}]] = \text{Tr}_B[\text{Tr}_A[O_{ABC}]]$.

$$(d) \text{Tr}_C[X_A \otimes N_{BC}] = X_A \otimes \text{Tr}_C[N_{BC}].$$

From Lemma 2.10 we recognize that the partial trace solves our problem. Simply define

$$\rho_A := \text{Tr}_B[\rho_{AB}].$$

Then the desired property (2.8) is a direct consequence of part (a) of the lemma (choose $M_{AB} = \rho_{AB}$ and $X_A = \mu_A(x)$), while parts (b) and (c) imply that ρ_A is a state, i.e., $\rho_A \in \mathcal{D}(\mathcal{H}_A)$. This calls for its own definition and notation:

Definition 2.12 (Reduced states). Given a state ρ_{AB} on AB , we define its *reduced state* on subsystem A by $\rho_A := \text{Tr}_B[\rho_{AB}]$. Similarly, we define the reduced state on subsystem B by $\rho_B := \text{Tr}_A[\rho_{AB}]$.

We use the same notation for three or more subsystems. For example, if ρ_{ABC} is a state on three subsystems ABC , then we denote its reduced states by $\rho_{AB} := \text{Tr}_C[\rho_{ABC}]$, $\rho_{AC} := \text{Tr}_B[\rho_{ABC}]$, $\rho_A := \text{Tr}_{BC}[\rho_{ABC}]$, etc.

It can be tempting to assume that all information about a state ρ_{AB} is contained in its reduced states ρ_A and ρ_B , and that we can always reconstruct ρ_{AB} from ρ_A and ρ_B by computing their tensor product $\rho_A \otimes \rho_B$. However, this is not the case! Such reconstruction generally is not possible – it works if only if ρ_{AB} itself happens to be a product state.

Lemma 2.13. *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then, ρ_{AB} is a product state if and only if $\rho_{AB} = \rho_A \otimes \rho_B$ (i.e., the product of its reduced states).*

Proof. Clearly if $\rho_{AB} = \rho_A \otimes \rho_B$ then it is a product state. Conversely, suppose that ρ_{AB} is a product state, which means $\rho_{AB} = \sigma_A \otimes \omega_B$ for some arbitrary states σ_A and ω_B . Then,

$$\rho_A = \text{Tr}_B[\rho_{AB}] = \text{Tr}_B[\sigma_A \otimes \omega_B] = \sigma_A,$$

where the first equality is the definition of the reduced state and the last equality holds thanks to Lemma 2.9. Similarly, one can verify that $\rho_B = \omega_B$. This confirms that $\rho_{AB} = \rho_A \otimes \rho_B$. \square

Let us discuss a concrete example. For a system of two qubits, $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$, consider the (standard) *maximally entangled state*, which is the pure state

$$\rho_{AB} = |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|, \quad \text{where} \quad |\Phi_{AB}^+\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (2.11)$$

The superscript “+” means nothing in particular, it is just a symbol to indicate this particular vector. (Why is this state called “entangled”? We will discuss this in Lecture 3.) Note that

$$\begin{aligned} \rho_{AB} &= |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+| \\ &= \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|) \\ &= \frac{1}{2}(|0\rangle\langle 0| \otimes \underbrace{|0\rangle\langle 0|}_{\text{Tr}=1} + |1\rangle\langle 0| \otimes \underbrace{|1\rangle\langle 0|}_{\text{Tr}=0} + |0\rangle\langle 1| \otimes \underbrace{|0\rangle\langle 1|}_{\text{Tr}=0} + |1\rangle\langle 1| \otimes \underbrace{|1\rangle\langle 1|}_{\text{Tr}=1}). \end{aligned} \quad (2.12)$$

To compute the reduced state on A , we can simply use linearity and Lemma 2.9 for each of the four terms. Using the traces indicated in (2.12), the result is

$$\rho_A = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \rho_B, \quad (2.13)$$

Correlated		Product
Maximally entangled	Maximally classically correlated	Independent random bits
$\frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ rank = 1	$\frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ rank = 2	$\frac{1}{4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ rank = 4

Table 2.1: Summary of the four states from Eq. (2.12), Eq. (2.2), and Eq. (2.15). While in all three cases the reduced states are maximally mixed, i.e., $\rho_A = \rho_B = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the global state encodes different types of correlations: entanglement, classical correlations, and no correlations. We will come back to these three states again in Example 8.4 of Lecture 8.

where the second equality follows by symmetry between the A and B systems in $|\Phi_{AB}^+\rangle$.

Note that something remarkable has happened: *We started with a pure state ρ_{AB} , but nevertheless its reduced states ρ_A and ρ_B were mixed!* This is an important reason for considering mixed states – they naturally arise when describing the state of a subsystem.

Remark 2.14. It is instructive to write down the maximally entangled state with respect to the product basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$:

$$|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad \rho_{AB} = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} (1 \ 0 \ 0 \ 1) = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \quad (2.14)$$

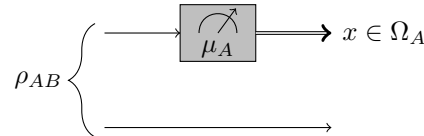
The right-hand side matrix can also be read off directly from Eq. (2.12). Note that ρ_{AB} cannot be rebuilt back from its reduced states ρ_A and ρ_B from Eq. (2.13). Indeed,

$$\rho_A \otimes \rho_B = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}_A \otimes \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}_B = \frac{1}{4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \neq \rho_{AB}. \quad (2.15)$$

This is a feature common to all correlated states, quantum or classical. Note that compared to the classical correlated state (2.2), the maximally entangled state in addition also has 1s in the top right and bottom left corners. This is a crucial difference! For example, the latter is a pure state (has rank one), while Eq. (2.2) is mixed (its rank is two). The product state in Eq. (2.15) is even more mixed since it has rank 4 (see Table 2.1 for a summary).

You can discuss some other examples in Exercises 2.5 and 2.7.

Equipped with the partial trace, we can also address another question that you might have had. Let us return to the situation of Section 2.2, where discussed measurements on part of a quantum system, as in the following figure:



In Axiom 2.6, we specified the probabilities of measurement outcomes in such a situation. But what state should we assign to B after the measurement? This is given in the following rule.

Axiom 2.15 (Measurement on a subsystem, part II). Suppose a quantum system is in state ρ_{AB} , and we apply a measurement $\mu_A: \Omega_A \rightarrow \text{PSD}(\mathcal{H}_A)$ on A and obtain outcome $x \in \Omega_A$. Then the state of the remaining system B after the measurement is given by

$$\rho_{B,x} = \frac{\text{Tr}_A[\rho_{AB}(\mu_A(x) \otimes I_B)]}{\text{Tr}[\rho_{AB}(\mu_A(x) \otimes I_B)]}. \quad (2.16)$$

This state is called the *post-measurement state* on B corresponding to this outcome.

Observe that the denominator in Eq. (2.16) is precisely the probability of outcome x according to Axiom 2.6. It is also the trace of the numerator. Therefore, if this probability is nonzero, Eq. (2.16) is a well-defined quantum state.

It is a pleasant exercise to verify that if one first measures subsystem A and then subsystem B , the joint probability of outcomes is given exactly by Axiom 2.7, see Exercise 2.8.

Remark 2.16. If we average the state of B over all possible measurement outcomes, we obtain

$$\sum_{x \in \Omega_A} p_x \rho_{B,x} = \sum_{x \in \Omega_A} \text{Tr}_A[\rho_{AB}(\mu_A(x) \otimes I_B)] = \text{Tr}_A[\rho_{AB}] = \rho_B.$$

This is completely reasonable: if someone carries out a measurement on system A , but you only have access to system B and receive no information about the outcome, your description of system B should not change (otherwise this could be used to signal instantaneously, at a speed faster than the speed of light).

2.4 Purifications

It is natural to ask whether we can also go the other way around. Suppose we start with a mixed state σ_A – can we always find a pure state on a larger system so that σ_A is its reduced state? Indeed, this can always be done.

Lemma 2.17 (Existence of purifications). *Let $\sigma_A \in \text{D}(\mathcal{H}_A)$ be a state and \mathcal{H}_B a Hilbert space of dimension $\dim \mathcal{H}_B \geq \text{rank } \sigma_A$. Then there exists a vector $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ such that*

$$\text{Tr}_B[|\Psi_{AB}\rangle\langle\Psi_{AB}|] = \sigma_A. \quad (2.17)$$

Any vector $|\Psi_{AB}\rangle$ or pure state $|\Psi_{AB}\rangle\langle\Psi_{AB}|$ with this property is called a purification of σ_A .

Note that any vector $|\Psi_{AB}\rangle$ that satisfies (2.17) is automatically a unit vector, By Lemma 2.10 (b).

$$\|\Psi_{AB}\|^2 = \text{Tr}[|\Psi_{AB}\rangle\langle\Psi_{AB}|] = \text{Tr}[\text{Tr}_B[|\Psi_{AB}\rangle\langle\Psi_{AB}|]] = \text{Tr}[\sigma_A] = 1.$$

Proof. Using an eigendecomposition (1.6), we can write

$$\sigma_A = \sum_{i=1}^r p_i |e_i\rangle\langle e_i|, \quad (2.18)$$

where $r := \text{rank } \sigma_A$, $\{p_i\}_{i=1}^r$ is the probability distribution formed by the *nonzero* eigenvalues of σ_A ($p_i > 0$ and $\sum_{i=1}^r p_i = \text{Tr } \sigma_A = 1$), and $|e_i\rangle$ are corresponding orthonormal eigenvectors.

Since $\dim \mathcal{H}_B \geq r$, we can choose orthonormal $|f_1\rangle, \dots, |f_r\rangle \in \mathcal{H}_B$. We claim that the following vector is a purification of σ_A :

$$|\Psi_{AB}\rangle := \sum_{i=1}^r \sqrt{p_i} |e_i\rangle \otimes |f_i\rangle. \quad (2.19)$$

Indeed,

$$\begin{aligned} \text{Tr}_B[|\Psi_{AB}\rangle\langle\Psi_{AB}|] &= \text{Tr}_B\left[\sum_{i,j=1}^r \sqrt{p_i p_j} |e_i\rangle\langle e_j| \otimes |f_i\rangle\langle f_j|\right] \\ &= \sum_{i,j=1}^r \sqrt{p_i p_j} |e_i\rangle\langle e_j| \underbrace{\text{Tr}[|f_i\rangle\langle f_j|]}_{=\delta_{i,j}} = \sum_{i=1}^r p_i |e_i\rangle\langle e_i| = \sigma_A \end{aligned}$$

by virtually the same calculation that we used to deduce Eq. (2.13) from Eq. (2.11). \square

The construction in the proof of Lemma 2.17 is quite important and also works in a more general situation. Suppose that we are given a quantum state $\sigma_A = \sum_{i=1}^k p_i |\psi_i\rangle\langle\psi_i|$, where the $|\psi_i\rangle$ are still unit vectors but *not* assumed to be pairwise orthogonal (so k can now be larger than the rank of σ_A). Then it is still true that for any orthonormal set of vectors $|f_1\rangle, \dots, |f_k\rangle \in \mathcal{H}_B$,

$$|\Psi_{AB}\rangle := \sum_{i=1}^k \sqrt{p_i} |\psi_i\rangle \otimes |f_i\rangle$$

is a purification of σ_A . This follows by the same calculation as in the proof of Lemma 2.17 (we never used that the $|e_i\rangle$ were orthogonal). You can practice this construction in Exercise 2.9.

Are purifications unique? In the proof of Lemma 2.17 we chose an arbitrary orthonormal basis of \mathcal{H}_B , so clearly they are not unique. However, this is the only source of ambiguity:

Lemma 2.18 (Uniqueness of purifications). *Let $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, $|\Phi_{AC}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_C$ be two purifications of $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$. If $\dim \mathcal{H}_B \leq \dim \mathcal{H}_C$, then there is an isometry $V_{B \rightarrow C}: \mathcal{H}_B \rightarrow \mathcal{H}_C$ such that*

$$|\Phi_{AC}\rangle = (I_A \otimes V_{B \rightarrow C})|\Psi_{AB}\rangle.$$

In particular, if $\dim \mathcal{H}_B = \dim \mathcal{H}_C$ then the two purifications are related by a unitary on the purifying system!

Recall that an operator $V \in \mathcal{L}(\mathcal{H}, \mathcal{K})$ is called an *isometry* if $V^\dagger V = I_{\mathcal{H}}$. Isometries preserve inner products, so they map orthonormal sets to orthonormal sets. This implies that $\dim \mathcal{H} \leq \dim \mathcal{K}$. We denote the set of all isometries from \mathcal{H} to \mathcal{K} by

$$\mathcal{U}(\mathcal{H}, \mathcal{K}) := \{V \in \mathcal{L}(\mathcal{H}, \mathcal{K}) : V^\dagger V = I_{\mathcal{H}}\}. \quad (2.20)$$

If $\dim \mathcal{H} = \dim \mathcal{K}$ then any isometry U is a *unitary*, which means that it also satisfies $UU^\dagger = I_{\mathcal{K}}$. Equivalently, an operator is unitary if its adjoint is its inverse. We denote the set of unitary operators on $\mathcal{H} = \mathcal{K}$ by

$$\mathcal{U}(\mathcal{H}) := \mathcal{U}(\mathcal{H}, \mathcal{H}) = \{U \in \mathcal{L}(\mathcal{H}) : U^\dagger U = I_{\mathcal{H}}, UU^\dagger = I_{\mathcal{K}}\}. \quad (2.21)$$

You can prove Lemma 2.18 in Exercise 2.13.

There is a particularly convenient way to construct a purification.

Definition 2.19 (Standard purification). For any state $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$ on $\mathcal{H}_A = \mathbb{C}^d$, the *standard purification* is defined as

$$|\Psi_{AB}^{\text{std}}\rangle := \left(\sqrt{\sigma_A} \otimes I_B\right) \sum_x |x\rangle \otimes |x\rangle, \quad (2.22)$$

where $\mathcal{H}_B := \mathcal{H}_A$ and $\{|x\rangle\}$ denotes the standard basis of \mathcal{H}_A .

The square root $\sqrt{\sigma_A}$ in Eq. (2.22) is the PSD operator defined by taking the square roots of the eigenvalues of σ_A , while keeping the eigenvectors the same. See Eq. (1.10) for more details and see Definition 1.6 for how to define general functions of Hermitian operators.

To see that Eq. (2.22) defines a purification, simply compute the partial trace:

$$\begin{aligned} \text{Tr}_B \left[|\Psi_{AB}^{\text{std}}\rangle \langle \Psi_{AB}^{\text{std}}| \right] &= \sum_{x,y} \text{Tr}_B \left[\left(\sqrt{\sigma_A} \otimes I_B \right) \left(|x\rangle \langle y| \otimes |x\rangle \langle y| \right) \left(\sqrt{\sigma_A} \otimes I_B \right) \right] \\ &= \sqrt{\sigma_A} \sum_{x,y} \text{Tr}_B \left[|x\rangle \langle y| \otimes |x\rangle \langle y| \right] \sqrt{\sigma_A} \\ &= \sqrt{\sigma_A} \sum_{x,y} |x\rangle \langle y| \underbrace{\text{Tr} \left[|x\rangle \langle y| \right]}_{=\delta_{x,y}} \sqrt{\sigma_A} = \sqrt{\sigma_A} \underbrace{\sum_x |x\rangle \langle x|}_{=I_A} \sqrt{\sigma_A} = \sigma_A. \end{aligned}$$

To go from the first to the second line, use Lemma 2.11 (a).

2.5 Schmidt decomposition

States of the form (2.19) are quite pleasant to work with, since both sets $\{|e_i\rangle\}$ and $\{|f_j\rangle\}$ consist of orthonormal vectors. For example, it is easy to calculate their reduced states. In fact, any *bipartite* pure state (i.e., pure state of two systems) can be written in this form – this is called the Schmidt decomposition.

Theorem 2.20 (Schmidt decomposition). Any $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written as

$$|\Psi_{AB}\rangle = \sum_{i=1}^r s_i |e_i\rangle \otimes |f_i\rangle,$$

where the $s_i > 0$, the $|e_i\rangle \in \mathcal{H}_A$ are orthonormal, and the $|f_i\rangle \in \mathcal{H}_B$ are orthonormal.

A decomposition of this form is called a Schmidt decomposition of $|\Psi_{AB}\rangle$, r is called the Schmidt rank, and the s_i are called the Schmidt coefficients of $|\Psi_{AB}\rangle$.

Using the Schmidt decomposition, we see as before that the reduced states are given by

$$\rho_A = \sum_{i=1}^r s_i^2 |e_i\rangle \langle e_i|, \quad \rho_B = \sum_{i=1}^r s_i^2 |f_i\rangle \langle f_i|. \quad (2.23)$$

This is a very important fact which has important consequences, such as the following.

Corollary 2.21 (Reduced states of pure states). If $\rho_{AB} = |\Psi_{AB}\rangle \langle \Psi_{AB}|$ is a pure state, then ρ_A and ρ_B have the same rank (namely r) and the same nonzero eigenvalues (namely, the $\{s_i^2\}$).

Corollary 2.22 (When is a pure state a product state?). *Let $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ be a pure state. Then, ρ_A is pure if and only if ρ_B is pure if and only if ρ_{AB} is a product state.*

Proof. Clearly, ρ_A is pure if and only if ρ_B is pure since both have the same rank (Corollary 2.21).

If ρ_A is pure then there is only one nonzero Schmidt coefficient ($s_1 = 1$), so $|\Psi_{AB}\rangle = |e_1\rangle \otimes |f_1\rangle$ and so $\rho_{AB} = |e_1\rangle\langle e_1| \otimes |f_1\rangle\langle f_1|$ is a product state.

Conversely, suppose that ρ_{AB} is a product state, so $\rho_{AB} = \rho_A \otimes \rho_B$ (Lemma 2.13). Since ρ_{AB} is pure, we have $1 = \text{rank } \rho_{AB} = \text{rank } \rho_A \text{rank } \rho_B$. Thus, both ρ_A and ρ_B have rank one, hence are pure states. \square

It is crucially important in Corollary 2.22 that the global state ρ_{AB} is pure. For mixed ρ_{AB} , it still holds that if ρ_A is pure then ρ_{AB} is a product state – you will prove this in Exercise 2.10 – but the converse is patently false. That is, there exist many (mixed) product states ρ_{AB} such that ρ_A or ρ_B are not pure.

The Schmidt decomposition is a mild restatement of the singular value decomposition of operators, which we recall in the following. You can prove the latter using the former in Exercise 2.12. For completeness, we sketch a proof of the singular value decomposition (but you have probably seen this before and it is also somewhat outside the scope of this class).

Theorem 2.23 (Singular value decomposition). *Any operator $M \in L(\mathcal{H}, \mathcal{K})$ has a singular value decomposition (SVD): That is, we can write*

$$M = \sum_{i=1}^r s_i |e_i\rangle\langle g_i|, \quad (2.24)$$

where $r = \text{rank } M$, $s_i > 0$, the $|e_i\rangle$ are orthonormal in \mathcal{K} , and the $|g_i\rangle$ are orthonormal in \mathcal{H} . The numbers s_i are called the singular values of M , and the $|e_i\rangle$ and $|g_i\rangle$ are called left and right singular vectors of M , respectively.

Proof. Consider the operator MM^\dagger , which is always positive semidefinite (part (c) of Lemma 1.4), so it has an eigendecomposition

$$MM^\dagger = \sum_i t_i |e_i\rangle\langle e_i|,$$

where $|e_i\rangle$ is an orthonormal basis in \mathcal{K} . Suppose that $t_1, \dots, t_r > 0$, while $t_i = 0$ for $i > r$. Note that the latter means that $\|M^\dagger|e_i\rangle\|^2 = \langle e_i|MM^\dagger|e_i\rangle = 0$, so $M^\dagger|e_i\rangle = 0$ for all $i > r$. Define $s_i := \sqrt{t_i}$. For $i = 1, \dots, r$, set $|g_i\rangle = \frac{M^\dagger|e_i\rangle}{s_i} \in \mathcal{H}$. Then the $|g_i\rangle$ are orthonormal, since

$$\langle g_i|g_j\rangle = \frac{\langle e_i|MM^\dagger|e_j\rangle}{s_i s_j} = \frac{t_j \langle e_i|e_j\rangle}{s_i s_j} = \frac{t_j}{s_i s_j} \delta_{i,j} = \delta_{i,j}.$$

For $i = 1, \dots, r$, it holds that

$$M|g_i\rangle = \frac{MM^\dagger|e_i\rangle}{s_i} = \frac{t_i|e_i\rangle}{s_i} = s_i|e_i\rangle.$$

This shows that M acts as in Eq. (2.24) for all vectors in the span of $|g_1\rangle, \dots, |g_r\rangle$. It remains to prove that $M|\psi\rangle = 0$ for every $|\psi\rangle$ that is orthogonal to $|g_1\rangle, \dots, |g_r\rangle$. Indeed

$$\langle e_i|M|\psi\rangle = (M^\dagger|e_i\rangle)^\dagger|\psi\rangle = \begin{cases} s_i \langle g_i|\psi\rangle = 0 & \text{if } i = 1, \dots, r, \text{ since then } \langle g_i|\psi\rangle = 0, \\ 0 & \text{if } i > r, \text{ since then } M^\dagger|e_i\rangle = 0. \end{cases}$$

We still need to check that r equals the rank of M . This follows from

$$r = \text{rank } M^\dagger M \leq \text{rank } M \leq r,$$

where we first used that r is the rank of $M^\dagger M$ (the number of nonzero t_i 's), then that the rank of a product is no larger than the rank of the factors, and finally Eq. (2.24), noting that its right-hand side has rank no larger than r . \square

How can we find the singular values in practice?

- We see directly from Eq. (2.24) (but also from the proof) that the singular values $\{s_i\}$ are necessarily the *square roots* of the nonzero eigenvalues of MM^\dagger (equivalently, of $M^\dagger M$). In other words, the singular values are the nonzero eigenvalues of $\sqrt{M^\dagger M}$ (or of $\sqrt{MM^\dagger}$).
- If $M = M^\dagger$, then its singular values are simply the *absolute values* of its nonzero eigenvalues (Exercise 2.11).

2.6 Exercises

Throughout, A, B, C denote quantum systems with Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C$. The sets $\{|a\rangle\}$ and $\{|b\rangle\}$ denote arbitrary orthonormal bases of \mathcal{H}_A and \mathcal{H}_B ; $|a, b\rangle = |a\rangle \otimes |b\rangle$ denotes the product basis of $\mathcal{H}_A \otimes \mathcal{H}_B$.

2.1 Product means independent: For $i \in \{1, \dots, n\}$, let p_i be a probability distribution on some finite set Σ_i and ρ_i the corresponding classical state (see Definition 1.10). Show that $\rho = \rho_1 \otimes \dots \otimes \rho_n$ is the classical state corresponding to the joint distribution $p(x_1, \dots, x_n) = p_1(x_1) \dots p_n(x_n)$ where each $x_i \in \Sigma_i$ is picked independently from the distribution p_i .

2.2 Not product: Show that the maximally correlated state (2.2) and the maximally entangled state (2.14) are indeed correlated, i.e., they are not product states.

2.3 Nayak's bound: Alice wants to communicate m bits to Bob by sending n qubits. She chooses one state $\rho_x \in \mathcal{D}(\mathcal{H})$, where $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, for each possible message $x \in \{0, 1\}^m$ that she may want to send. Bob uses a measurement $\mu: \{0, 1\}^m \rightarrow \text{PSD}(\mathcal{H})$ to decode the message.

- Write down a formula for the probability that Bob successfully decodes Alice's message, assuming the latter is drawn from a known probability distribution $p(x)$ on $\{0, 1\}^m$.
- Show that if the message is drawn *uniformly* at random, then the probability that Bob successfully decodes the bitstring is at most 2^{n-m} .

2.4 Partial trace trickery: Prove Lemma 2.11.

2.5 Reduced states of classical states: Consider the classical state

$$\rho_{XY} = \sum_{x,y} p(x, y) |x, y\rangle\langle x, y|$$

on $\mathcal{H}_X \otimes \mathcal{H}_Y$, where $\mathcal{H}_X = \mathbb{C}^{\Sigma_X}$, $\mathcal{H}_Y = \mathbb{C}^{\Sigma_Y}$, and $p(x, y)$ is an arbitrary probability distribution on $\Sigma_X \times \Sigma_Y$. Compute the reduced states ρ_X and ρ_Y .

2.6 Partial trace of any two-qubit operator: Let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$. Compute the partial traces Tr_A and Tr_B of

$$M_{AB} = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix}.$$

The matrix is written in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.

2.7 Reduced states of a pure state: Compute the reduced states ρ_A and ρ_B of the two-qubit pure state $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ given by $|\Psi_{AB}\rangle = \frac{1}{3}|0,0\rangle + \frac{2}{3}|0,1\rangle + \frac{2}{3}|1,0\rangle$.

2.8 Parallel vs. sequential: Consider a system in state ρ_{AB} . Imagine you first apply a measurement μ_A on A and then a measurement μ_B on B . Use Axioms 2.6 and 2.15 to verify that the joint probability of the two measurement outcomes is given by Axiom 2.7.

2.9 Purification:

- (a) Find a purification $|\psi_{AB}\rangle$ of the single-qubit state $\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +|$.
- (b) Find a purification $|\phi_{ABC}\rangle$ of the two-qubit state $\rho_{AB} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|1+\rangle\langle 1+|$.
- (c) Compute the reduced state of system B for your purification $|\phi_{ABC}\rangle$.

2.10 Extensions of pure states: Let $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C$ be arbitrary Hilbert spaces.

- (a) Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that ρ_A is pure. Show that $\rho_{AB} = \rho_A \otimes \rho_B$.
Hint: In class we proved this when ρ_{AB} is pure. Use a purification to reduce to this case.
- (b) Let $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ such that ρ_{AB} is pure. Show that $\rho_{AC} = \rho_A \otimes \rho_C$ and $\rho_{BC} = \rho_B \otimes \rho_C$.
- (c) Let $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ such that ρ_{AB} and ρ_{AC} are pure. Show that $\rho_{ABC} = \rho_A \otimes \rho_B \otimes \rho_C$.

2.11 Singular values and eigenvalues: Show that if M is Hermitian then its singular values are equal to the *absolute values* of its nonzero eigenvalues.

2.12 Schmidt decomposition from SVD: Let $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ be an arbitrary pure state. Fix orthonormal bases $|a\rangle$ and $|b\rangle$, write $|\Psi_{AB}\rangle = \sum_{a,b} M_{ab}|a,b\rangle$, and define a corresponding operator $M = \sum_{a,b} M_{ab}|a\rangle\langle b|$.


- (a) Verify that $\rho_A = MM^\dagger$ and $\rho_B = M^T \overline{M}$. Here the transpose and the complex conjugate are computed with respect to the fixed bases.
- (b) Let $M = \sum_i s_i |e_i\rangle\langle f_i|$ be a singular value decomposition. Show that $|\Psi_{AB}\rangle = \sum_i s_i |e_i\rangle \otimes |f_i\rangle$ is a Schmidt decomposition. Thus you have proved Theorem 2.20.
- (c) Explain how to find a Schmidt decomposition of the following two-qubit pure state:

$$|\Psi\rangle = \frac{\sqrt{2}+1}{\sqrt{12}}(|00\rangle + |11\rangle) + \frac{\sqrt{2}-1}{\sqrt{12}}(|01\rangle + |10\rangle).$$

2.13 Uniqueness of purifications: Let $\rho_A = \sum_{i=1}^r p_i |e_i\rangle\langle e_i|$, where p_i are the nonzero eigenvalues of ρ_A and $|e_i\rangle$ corresponding orthonormal eigenvectors. If some eigenvalue appears more than once then this decomposition is *not* unique.

- (a) Show that, nevertheless, *any* purification $|\Psi_{AB}\rangle$ of ρ_A has a Schmidt decomposition of the form $|\Psi_{AB}\rangle = \sum_{i=1}^r s_i |e_i\rangle \otimes |f_i\rangle$, with the same $|e_i\rangle$ as above.
Hint: Start with an arbitrary Schmidt decomposition and rewrite it in the desired form.

(b) Prove Lemma 2.18.

2.14  **Practice:** Consider the following 2×2 -matrix M :

$$M = \begin{pmatrix} -0.2422 + 0.07118i & -0.1689 - 0.6927i \\ -0.1108 - 0.19192i & -0.6045 + 0.1161i \end{pmatrix}$$

Let $|\Psi_{AB}\rangle = \sum_{a,b} M_{ab} |ab\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ and let ρ_{AB} denote the corresponding pure state.

- (a) Compute the Schmidt coefficients of $|\Psi_{AB}\rangle$.
- (b) Compute the reduced states ρ_A and ρ_B .

Lecture 3

Entanglement

Today we will discuss the phenomenon of *quantum entanglement*, which is one of the most profound phenomena of quantum information. Recall that last week we defined the maximally entangled state (2.11) of two qubits. Today, we will first give a general definition of entanglement (and maximal entanglement). Then we will discuss three examples that show that the state $|\Phi^+\rangle$ is a remarkable resource that can be used to achieve things that are not possible otherwise.

3.1 Separable and entangled states

Recall that we defined a quantum state to be *correlated* if it is not a product, $\rho_{AB} \neq \rho_A \otimes \rho_B$. This has nothing to do with quantum mechanics per se, since correlations also exist in probability theory. For example, the state (2.2) describes correlations between two bits that are random but always equal to each other.

How can we distinguish classical from quantum correlations? The following definition gives us one way to do so:

Definition 3.1 (Separable and entangled states). Let \mathcal{H}_A and \mathcal{H}_B be two Hilbert spaces. A quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called *separable* or *unentangled* (between systems A and B) if it is a convex combination of product states, i.e., if there is a probability distribution $(p_i)_{i \in I}$ and states $\rho_{A,i} \in \mathcal{D}(\mathcal{H}_A)$, $\rho_{B,i} \in \mathcal{D}(\mathcal{H}_B)$ such that

$$\rho_{AB} = \sum_{i \in I} p_i \rho_{A,i} \otimes \rho_{B,i}. \quad (3.1)$$

We let $\text{SepD}(\mathcal{H}_A : \mathcal{H}_B)$ denote the convex set of separable states on $\mathcal{H}_A \otimes \mathcal{H}_B$.

A state that is not separable is called *entangled*.

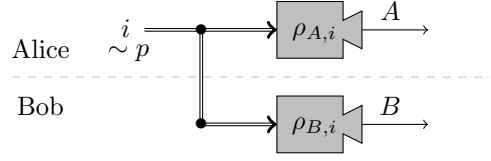
As an example, note that any classical state ρ_{XY} on $\mathbb{C}^{\Sigma_X} \otimes \mathbb{C}^{\Sigma_Y}$ is separable, since we recall from Definition 1.10 that in this case

$$\rho_{XY} = \sum_{x,y} p(x,y) |x,y\rangle\langle x,y| = \sum_{x,y} p(x,y) |x\rangle\langle x| \otimes |y\rangle\langle y|.$$

for some joint distribution $p \in \mathcal{P}(\Sigma_X \times \Sigma_Y)$. This is clearly of the form (3.1). This example justifies the idea that entanglement captures only the non-classical part of correlations.

In fact, the separable states are precisely the states which can be created in the following way: We have two protagonists, Alice and Bob, each in their separate laboratory. Alice draws a random index $i \in I$ according to a probability distribution p , and sends this index i over to Bob.

Finally, Alice prepares some quantum state $\rho_{A,i}$ in her laboratory, while Bob creates some other state $\rho_{B,i}$. Clearly, their joint state at the end of this process is described by Eq. (3.1). We can visualize this protocol in the following diagram:



Note that before the final step of the protocol, the two parties are classically correlated, since they share the random variable i . As the final step is purely local, we should not think of it as creating any additional correlations. Accordingly, the entangled states are those that *cannot* be prepared by only using classical correlations. This gives a satisfying motivation for Definition 3.1.

Remark 3.2. For the notion of entanglement to make sense in the first place, you need a system consisting of (at least) two distinguished subsystems, say A and B . In other words, it does not make sense to talk about the entanglement of a state on a general Hilbert space \mathcal{H} – we must always refer to a tensor product factorization $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$.

What do entangled states look like? For pure states, the situation simplifies considerably.

Lemma 3.3. *Let $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}| \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a pure state. Then ρ_{AB} is separable if and only if it is a product state.*

Proof. Clearly, any product state is separable. For the converse, note that any separable state ρ_{AB} can be written in the form

$$\rho_{AB} = \sum_{i \in I} p_i |\psi_{A,i}\rangle\langle\psi_{A,i}| \otimes |\phi_{B,i}\rangle\langle\phi_{B,i}| = \sum_{i \in I} p_i (|\psi_{A,i}\rangle \otimes |\phi_{B,i}\rangle)(\langle\psi_{A,i}| \otimes \langle\phi_{B,i}|)$$

with $p_i > 0$ and unit vectors $|\psi_{A,i}\rangle \in \mathcal{H}_A$ and $|\phi_{B,i}\rangle \in \mathcal{H}_B$. (Indeed, if any of the $\rho_{A,i}$ or $\rho_{B,i}$ are mixed we can expand them into a convex combination of pure states by using the spectral theorem, see Eq. (1.14), and we can always leave out terms with $p_i = 0$.) Since $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$, it follows that

$$\begin{aligned} 1 &= \langle\Psi_{AB}|\rho_{AB}|\Psi_{AB}\rangle = \sum_{i \in I} p_i \langle\Psi_{AB}|(|\psi_{A,i}\rangle \otimes |\phi_{B,i}\rangle)(\langle\psi_{A,i}| \otimes \langle\phi_{B,i}|)|\Psi_{AB}\rangle \\ &= \sum_{i \in I} p_i |\langle\Psi_{AB}|\psi_{A,i} \otimes \psi_{B,i}\rangle|^2. \end{aligned}$$

Using the Cauchy-Schwartz inequality,¹ this implies that, for all $i \in I$,

$$|\langle\Psi_{AB}|\psi_{A,i} \otimes \psi_{B,i}\rangle| = 1$$

and hence that $|\Psi_{AB}\rangle$ is proportional to $|\psi_{A,i}\rangle \otimes |\psi_{B,i}\rangle$. Since these are all unit vectors, it follows that $\rho_{AB} = |\psi_{A,i}\rangle\langle\psi_{A,i}| \otimes |\psi_{B,i}\rangle\langle\psi_{B,i}|$ for all $i \in I$. Thus, ρ_{AB} is a product state. \square

Thus we can use the criteria from Corollary 2.22 to test if a pure state is entangled or not.

¹Recall that the *Cauchy-Schwartz inequality* states that $|\langle\alpha|\beta\rangle| \leq \|\alpha\| \|\beta\|$, with equality if and only if $|\alpha\rangle$ and $|\beta\rangle$ are proportional.

Corollary 3.4. *Let $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}| \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a pure state. Then ρ_{AB} is separable if and only if ρ_A is pure if and only if ρ_B is pure.*

For example, we can now quickly confirm that the maximally entangled two-qubit state (2.11),

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

is entangled, since we know that ρ_A is not pure. In fact, $\rho_A = \frac{I}{2}$ is maximally mixed. We now give a general definition of maximally entangled states:

Definition 3.5 (Maximally entangled states). Let \mathcal{H}_A and \mathcal{H}_B be two Hilbert spaces. A state ρ_{AB} is called *maximally entangled* if \mathcal{H}_A and \mathcal{H}_B have the same dimension d and ρ_{AB} is of the form

$$\rho_{AB} = |\Phi_{AB}\rangle\langle\Phi_{AB}|, \quad |\Phi_{AB}\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |e_i\rangle \otimes |f_i\rangle, \quad (3.2)$$

where $\{|e_i\rangle\}_{i=1}^d$ and $\{|f_i\rangle\}_{i=1}^d$ are orthonormal bases of \mathcal{H}_A and \mathcal{H}_B , respectively. If $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^d$ or \mathbb{C}^Σ and we take the standard basis, we always have the (*standard*) *maximally entangled states*

$$|\Phi_{AB}^+\rangle := \frac{1}{\sqrt{d}} \sum_x |x, x\rangle = \frac{1}{\sqrt{d}} \sum_x |x\rangle \otimes |x\rangle. \quad (3.3)$$

We will often leave out “standard” and simply call $|\Phi_{AB}^+\rangle$ “the” maximally entangled state. Analogously to the two-qubit example we have the following observation, which you can verify in Exercise 3.5:

Lemma 3.6. *Let \mathcal{H}_A and \mathcal{H}_B be two Hilbert spaces and ρ_{AB} a state. Then ρ_{AB} is maximally entangled if and only if ρ_{AB} is pure and both ρ_A and ρ_B are maximally mixed.*

Since a pure state ρ_{AB} is separable if and only if ρ_A is pure (Corollary 3.4), and since the maximally mixed state is as far from being pure as possible, Lemma 3.6 gives some first justification the term “maximally entangled”. We will return to this point and give some stronger reasons in later lectures.

The maximally entangled states satisfy a useful symmetry, as you can verify in Exercise 3.6:

Lemma 3.7. *Let $|\Phi^+\rangle$ denote the standard maximally entangled state (3.3) for $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}$, where $\mathcal{H} = \mathbb{C}^d$ or \mathbb{C}^Σ . Then:*

- (a) Transpose trick: $(M \otimes I)|\Phi^+\rangle = (I \otimes M^\top)|\Phi^+\rangle$ for all $M \in \mathcal{L}(\mathcal{H})$.
- (b) $(U \otimes \bar{U})|\Phi^+\rangle = |\Phi^+\rangle$ for all unitaries $U \in \mathcal{U}(\mathcal{H})$.
- (c) $\langle\Phi^+|\bar{M} \otimes N|\Phi^+\rangle = \frac{1}{d} \text{Tr}[M^\dagger N]$ for all $M, N \in \mathcal{L}(\mathcal{H})$.

The transpose and complex conjugate are taken with respect to the standard basis of \mathcal{H} .

These definitions raise many interesting and deep questions:

- *What is so special about $|\Phi_{AB}^+\rangle$ or other entangled states?* We will spend the remainder of today’s lecture on this topic.
- *How can we detect if a given state is entangled?* For pure states we saw that this is quite easy, but for mixed states this is generally very hard (in fact, NP-hard). In Exercise 3.7 you can explore a useful criterion for mixed states.

- How much entanglement is there in a given state ρ_{AB} ?
- How can we manipulate entanglement?

We will discuss the last two questions in subsequent lectures (Lectures 10 and 12).

3.2 Bell states and superdense coding

For two qubits, the maximally entangled state is the first of the four *Bell states*:

$$\begin{aligned} |\Phi^+\rangle &= |\Phi^{(00)}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |\Phi^{(10)}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |\Phi^{(01)}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), & |\Phi^{(11)}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned} \quad (3.4)$$

These four vectors form an orthonormal basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$, called the *Bell basis*.

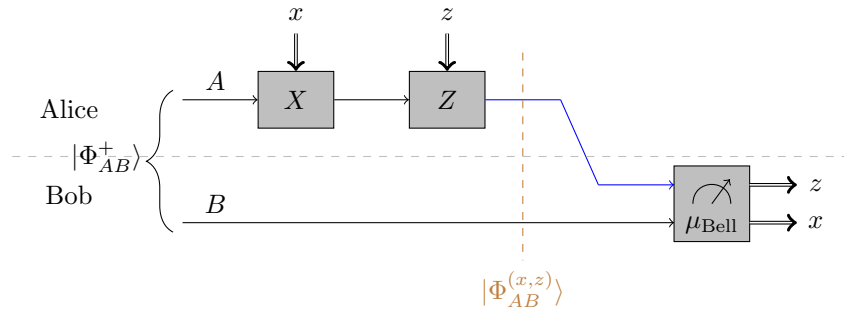
They are all maximally entangled in the sense of Definition 3.5. In particular, their reduced states on a single qubit are all the same: $\rho_A = \frac{I}{2} = \rho_B$. Accordingly the four Bell states are *locally indistinguishable* – if two parties Alice and Bob each have one qubit of a Bell state, neither Alice or Bob can tell the four states apart by a measurement of their qubit.

For this reason it is perhaps surprising that the Bell states enjoy the following *local conversation* property: For all $x, z \in \{0, 1\}$,

$$|\Phi^{(x,z)}\rangle = (Z^z X^x \otimes I)|\Phi^+\rangle = (I \otimes X^x Z^z)|\Phi^+\rangle, \quad (3.5)$$

where X and Z are two of the Pauli matrices (1.15). This is easy to verify, and you can do so in Exercise 3.10. Thus, if Alice and Bob each possess one qubit of a Bell state, any of them can apply a suitable combination of Pauli X or Z matrices to convert their joint state to any of the other four Bell states.

We can use this for a first application of entanglement. Suppose that Alice wants to communicate n bits to Bob (without any error), but she is only allowed to send a single qubit. We know from Exercise 2.3 that this is possibly only for $n = 1$. Interestingly, entanglement allows them to do better. Consider the following protocol, which is known as *superdense coding*.²



Let us describe the protocol more formally: Alice and Bob start out in a maximally entangled state $|\Phi_{AB}^+\rangle$, where system A belongs to Alice and system B belongs to Bob. Now suppose Alice wants to communicate two bits $x, z \in \{0, 1\}$. First, if $x = 1$ then she applies a Pauli X operator on her system. Next, if $z = 1$ then she applies a Pauli Z operator on her system.³ Finally, Alice

²Confusingly, in quantum protocols and circuits time goes from left to right, so the order of operations is reversed compared to symbolic expressions.

³Note that X and Z are unitary operators. For any state ρ and unitary U , the operator $U\rho U^\dagger$ is again a state. If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then $U\rho U^\dagger$ is the pure state corresponding to the unit vector $U|\psi\rangle$. It turns out that the operation $\rho \mapsto U\rho U^\dagger$ can be implemented physically – this is what we mean by the boxes labeled X and Z in the picture. In Lecture 4 we will discuss in detail what are the allowed operations in quantum theory.

sends over her qubit to Bob. Now both qubits belong to Bob, so he can perform a Bell basis measurement, that is,

$$\mu_{\text{Bell}}: \{0, 1\}^2 \rightarrow \text{PSD}(\mathbb{C}^2 \otimes \mathbb{C}^2), \quad (x, z) \mapsto |\Phi_{AB}^{(x,z)}\rangle\langle\Phi_{AB}^{(x,z)}|. \quad (3.6)$$

We claim that Bob’s measurement outcome is precisely Alice’s message. Indeed, by Eq. (3.5), the joint state of the system when Alice sends over her qubit is precisely the Bell state $|\Phi^{(x,z)}\rangle$, as indicated in the picture. Thus the Bell basis measurement can perfectly identify the state.

The superdense coding protocol allows us to *communicate two bits by sending a single qubit*. Of course, the protocol also consumes one maximally entangled qubit state, so there is no contradiction to Exercise 2.3. This gives us some first insight into the power of entanglement as a resource for communication.

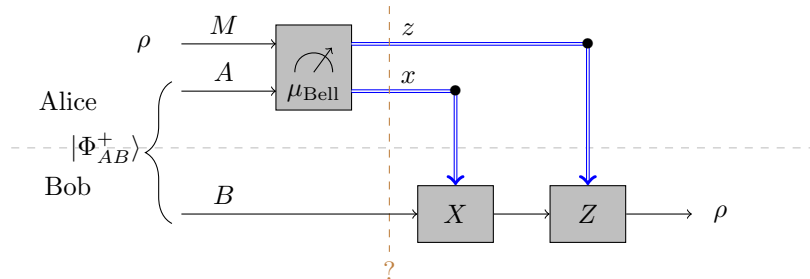
3.3 Teleportation

We now discuss another interesting protocol known as *teleportation*. Teleportation is dual to superdense coding since it achieves the opposite task: it lets you *transfer one qubit by sending two bits*.

Wait, does this even make sense? This is clearly impossible, since there are many more quantum states than bit strings. Indeed, there uncountably many quantum states of any given dimension (think of the Bloch ball for a single qubit), while there are only finitely many bitstrings of any given length.

Remark 3.8. If Alice wants to classically send a qubit state ρ to Bob, the only thing she can do is to send him a “recipe” for preparing this state. For example, she could send him the four matrix entries of ρ with respect to the standard basis. However, since the matrix entries are complex numbers, they can only be specified to a finite precision by sending bitstrings, so this would describe ρ only approximately. And even if they went through this trouble, the state reconstructed by Bob would not preserve the correlations Alice’s state might have had with another system. For example, if $\rho_A = \text{Tr}_R[|\Psi_{AR}\rangle\langle\Psi_{AR}|]$ where R is some reference system that is not accessible to Bob, the state he reconstructs would not be correlated with R .

In Section 3.2, we saw that a similar “no go” result could be overcome by using entanglement. Thus we can ask if our task here can in fact be achieved if Alice and Bob share some entanglement, say a maximally entangled state. Surprisingly, in this scenario it is indeed possible to perfectly transmit a quantum state by sending only classical information. The protocol that achieves this is known as the *teleportation* protocol. We first describe the protocol and then argue its correctness:



As visualized in the picture,⁴ Alice and Bob start out in a state $\rho_M \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|$, where ρ_M is

⁴In the picture, we labeled the initial state of system AB by $|\Phi_{AB}^+\rangle$ instead of the more correct $|\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|$, in line with the very customary but somewhat imprecise way of referring to unit vectors as pure “states”. We also left out the tensor product sign “ \otimes ” between this state and ρ_M . We did so only for brevity to make the pictures not too overloaded – the mathematical formulas given in the text are always precise.

some arbitrary qubit state (which can be unknown to Alice). First, Alice performs a Bell basis measurement (3.6) on both her qubits (M and A). The measurement outcome are two bits x and z , which she sends over to Bob. Finally, if $x = 1$ then Bob applies a Pauli X operator, and if $z = 1$ he applies a Pauli Z operator.

We claim that after the protocol has completed, the state of Bob's qubit is ρ – that is, exactly the same state that started out in Alice's M qubit! To verify this claim, we wish to calculate the post-measurement state on Bob's qubit for any possible measurement outcome (x, z) . We claim:⁵

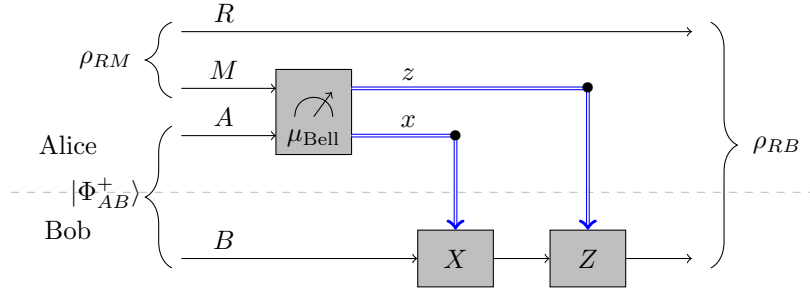
$$\text{Tr}_{MA}[(\mu_{\text{Bell},MA}(x, z) \otimes I_B)(\rho_M \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|)] = \frac{1}{4}X^x Z^z \rho_B Z^z X^x, \quad (3.7)$$

where ρ_B denote the same state as ρ_M , except that we consider it as a state on system B (recall that $\mathcal{H}_M = \mathbb{C}^2 = \mathcal{H}_B$, so this makes sense)! Comparing with Axiom 2.15, we see that each outcome (x, z) occurs with probability $\frac{1}{4}$, and that the corresponding post-measurement state is

$$X^x Z^z \rho_B Z^z X^x.$$

Using $X^2 = Z^2 = I$, we see that Bob's operations precisely undo the Pauli operators (cf. Footnote 3), and hence the final output state is ρ_B . Thus the teleportation protocol indeed behaves correctly.

In fact, the teleportation protocol not only transmits Alice's state to Bob, but it also *preserves any correlations* it might have had with some other system. In other words, if we start with $\rho_{RM} \otimes |\Psi_{AB}^+\rangle\langle\Psi_{AB}^+|$, where R is some arbitrary other quantum system and ρ_{RM} an arbitrary quantum state, and we apply the teleportation protocol to MAB as above, then the resulting joint state on systems R and B will be ρ_{RB} . This is visualized in the following figure:



This follows from the following equation, which generalizes Eq. (3.7) and which you will show in Exercise 3.11, together with the same reasoning as above:

$$\begin{aligned} & \text{Tr}_{MA}[(I_R \otimes \mu_{\text{Bell},MA}(x, z) \otimes I_B)(\rho_{RM} \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|)] \\ &= \frac{1}{4}(I \otimes X^x Z^z) \rho_{RB} (I \otimes Z^z X^x), \end{aligned} \quad (3.8)$$

where ρ_{RB} denotes the same state as ρ_{RM} , except that we consider it as a state on system RB (completely analogously to above).

Remark 3.9. Quantum teleportation is analogous to the classical *one-time pad*, a protocol for transmitting a private probabilistic bit from Alice to Bob by using only public communication and a secret shared random bit.

⁵Note that the two tensor product symbols “ \otimes ” in the left-hand side of Eq. (3.7) do *not* refer to the same tensor factors. The first one refers to $(\mathcal{H}_M \otimes \mathcal{H}_A) \otimes \mathcal{H}_B$, while the second refers to $\mathcal{H}_M \otimes (\mathcal{H}_A \otimes \mathcal{H}_B)$. Our use of subscripts make this completely unambiguous.

The maximally entangled state of two qubits is also called *EPR pair*, for Einstein, Podolsky, and Rosen who wrote a famous paper about it, or also *ebit*, for “entangled pair of qubits”. Teleportation and superdense coding can be summarized by the following two *resource inequalities*:

$$\begin{aligned} \text{teleportation:} \quad & \text{ebit} + 2[c \rightarrow c] \geq [q \rightarrow q], \\ \text{superdense coding:} \quad & \text{ebit} + [q \rightarrow q] \geq 2[c \rightarrow c], \end{aligned}$$

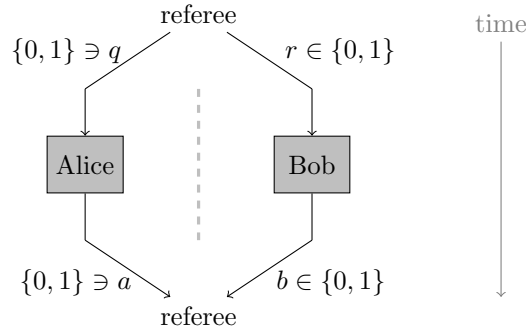
where $[c \rightarrow c]$ denotes one bit of classical communication and $[q \rightarrow q]$ denotes one qubit of quantum communication. You can read the inequality sign as “is at least as good as” or “can be used to implement”.

3.4 CHSH game (optional)

We will now discuss another way by which we can compare classical and quantum correlations. Namely, we will play a so-called nonlocal game!

In a *nonlocal game*, we imagine that a number of players play against a referee. The referee hands them questions and the players attempt to reply with answers that win them the game. The players’ goal is to maximize their chances of winning. Before the game starts, the players meet and agree upon a joint strategy – but then they are placed far apart from each other and cannot communicate with each other while the game is being played (this can be ensured by the laws of special relativity). The point is the following: *Since the players are constrained by the laws of physics, we can design games where players utilizing quantum theory may have an advantage.*

The *CHSH game* is a famous example of a nonlocal game that was invented by Clauser, Horne, Shimony, and Holt. It involves two players – Alice and Bob. As questions, they each receive a bit (q for Alice and r for Bob), and their answers are likewise bits (denoted a for Alice and b for Bob). The following figure illustrates the setup:



The win the game according to the following rules:

q	r	winning condition
0	0	$a = b$
0	1	$a = b$
1	0	$a = b$
1	1	$a \neq b$

The winning condition can be succinctly stated as follows:

$$a \oplus b = qr, \tag{3.9}$$

where “ \oplus ” means addition modulo 2 (also known as the XOR operation).⁶ We will always assume that the referee picks the questions (q, r) with equal probability $\frac{1}{4}$. Then we can speak of the *winning probability* of a given strategy.

Classical strategies

It is easy to see that the CHSH game does not have a perfect winning strategy in a world that is “local” and “realistic”. Here, “local” means that each player’s answer only depends on its immediate surroundings, and “realistic” means that the player’s strategy must assign a definite answer to any possible question – before that question is being asked. In other words, in a local and realistic world we assume that

$$a = f(q), \quad b = g(r)$$

for functions $f, g: \{0, 1\} \rightarrow \{0, 1\}$. When we say that the players may jointly agree on a strategy before the game is being played, we mean they may select these answer functions f, g in a correlated way.⁷ In mathematical terms, the functions f and g are allowed to be correlated random variables.

If the players’ strategy can be described by classical mechanics then the above would provide an adequate model. Thus, strategies of this form are usually referred to as *classical strategies* or also a *local hidden variable models*.⁸

Suppose now for sake of finding a contradiction that Alice and Bob can win the CHSH game perfectly using such a classical strategy. Then,

$$\begin{aligned} 1 &= 0 \oplus 0 \oplus 0 \oplus 1 \\ &= (f(0) \oplus g(0)) \oplus (f(0) \oplus g(1)) \oplus (f(1) \oplus g(0)) \oplus (f(1) \oplus g(1)) \\ &= 0. \end{aligned}$$

The second equality follows from Eq. (3.9), since we assumed that the strategy is perfect, and the last equality holds because each term $f(q), g(r)$ appears twice, but $x \oplus x = 0$ for any $x \in \{0, 1\}$. This contradiction shows that there is no perfect classical winning strategy for the CHSH game.

More quantitatively, if we imagine that the referee picks the questions (q, r) with equal probability $\frac{1}{4}$, then the winning probability is bounded by

$$p_{\text{win, classical}} \leq \frac{3}{4} \tag{3.10}$$

for any classical strategy, since the players must get at least one of the four possible answers wrong! This winning probability can be achieved, e.g., by the trivial strategy $f(q) = g(r) = 0$ for all $q, r \in \{0, 1\}$.

Remark 3.10. Eq. (3.10) can be thought of as a so-called *Bell inequality*. Perhaps you have seen Bell inequalities before in a different form. If so, do you see the connection?

⁶That is, $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$.

⁷For example, when the players meet before the game is being played, they could flip a coin, resulting in some random $\lambda \in \{0, 1\}$, and agree on the strategy given by $f(q) = q + \lambda \bmod 2$ and $g(r) = r + \lambda \bmod 2$.

⁸Where does this name come from? In our example above, we can think of λ as a “hidden variable”. Once its value is revealed, the behavior of the parties is purely deterministic. This is just like the behavior of a mechanical system, which can look very complex or even chaotic, but is in fact deterministic given the initial conditions at some point in time.

Quantum strategies

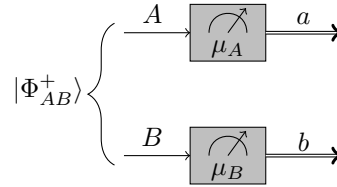
In a quantum strategy, we imagine that the two players are described by quantum theory. Thus they start out by sharing an arbitrary joint state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, where \mathcal{H}_A describes a quantum system in Alice's possession and \mathcal{H}_B describes a quantum system in Bob's possession. Upon receiving her question $q \in \{0, 1\}$, Alice will perform a measurement $\mu_{A,q}: \{0, 1\} \rightarrow \text{PSD}(\mathcal{H}_A)$ and uses the measurement outcome as her answer a . Likewise, when Bob receives his question $r \in \{0, 1\}$, he performs a measurement $\mu_{B,r}: \{0, 1\} \rightarrow \text{PSD}(\mathcal{H}_B)$ and uses the measurement outcome as his answer b .

It is not hard to see that any classical strategy is also a quantum strategy. But the crucial question is: Can players following a quantum strategy do better and “beat” the classical bound (3.10)? Remarkably, this is indeed possible to achieve

$$p_{\text{win,quantum}} = \frac{1}{2} \left(1 + \frac{1}{\sqrt{2}} \right) \approx 0.85, \quad (3.11)$$

and the key is to use entanglement!

We will now discuss how this can be achieved. Suppose that Alice and Bob share the maximally entangled state of two qubits, $|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Let us study what happens if they perform arbitrary basis measurements on their qubits, like in the following picture:



For a single qubit, we can parameterize any basis measurement by a unit vector $\vec{r} \in \mathbb{R}^3$. Indeed, any such vector is the Bloch vector of a pure state $|\psi_0\rangle\langle\psi_0|$, see Lemma 1.13, and hence determines a basis measurement

$$\mu: \{0, 1\} \rightarrow \text{PSD}(\mathbb{C}^2), \quad \begin{cases} \mu(0) = |\psi_0\rangle\langle\psi_0| = \frac{1}{2}(I + r_x X + r_y Y + r_z Z), \\ \mu(1) = I - |\psi_0\rangle\langle\psi_0| = \frac{1}{2}(I - r_x X - r_y Y - r_z Z). \end{cases} \quad (3.12)$$

Now suppose Alice's basis measurement is described by a unit vector \vec{r} , and Bob's basis measurement is described by a unit vector \vec{s} . What is the joint probability $p(a, b)$ of Alice's measurement outcome a and Bob's measurement outcome b ? For $a = b = 0$,

$$\begin{aligned} p(0, 0) &= \frac{1}{4} \langle \Phi^+ | (I + r_x X + r_y Y + r_z Z) \otimes (I + s_x X + s_y Y + s_z Z) | \Phi^+ \rangle \\ &= \frac{1}{4} (1 + r_x s_x - r_y s_y + r_z s_z), \end{aligned} \quad (3.13)$$

where the first step is Axiom 2.7 and the second step follows from a short calculation using the transpose trick (Lemma 3.7) and the algebraic properties of the Pauli operators, as you can verify in Exercise 3.13. In view of Eq. (3.12), $p(1, 1)$ is calculated by substituting $\vec{r} \mapsto -\vec{r}$ and $\vec{s} \mapsto -\vec{s}$ in Eq. (3.13); since this leaves the result invariant it is clear that $p(1, 1) = p(0, 0)$. Thus, the probability that the two measurement outcomes a, b agree is given by:

$$\Pr_{\vec{r}, \vec{s}}(a = b) = p(0, 0) + p(1, 1) = \frac{1}{2} (1 + r_x s_x - r_y s_y + r_z s_z), \quad (3.14)$$

where we use the subscript to indicate the ‘measurement axes’ \vec{r} and \vec{s} . This means that the measurement outcomes can be strongly correlated or anticorrelated, depending on the alignment

of the measurement axes! See also Exercise 3.14. This becomes even more clear if we restrict to the x - z -plane (i.e., $r_y = s_y = 0$), so that the formula simplifies to

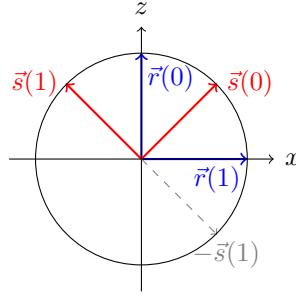
$$\Pr_{\vec{r}, \vec{s}}(a = b) = \frac{1}{2} (1 + \vec{r} \cdot \vec{s}) = \frac{1}{2} (1 + \cos \theta), \quad (3.15)$$

where θ is the angle between \vec{r} and \vec{s} .

How can we use this? We wish to find measurement axes $\vec{r}(q)$ for Alice (one for each question $q \in \{0, 1\}$ that she receives) and $\vec{s}(r)$ for Bob (one for each question $r \in \{0, 1\}$ he receives) such that the winning probability

$$\begin{aligned} p_{\text{win}} &= \frac{1}{4} \Pr_{\vec{r}(0), \vec{s}(0)}(a = b) + \frac{1}{4} \Pr_{\vec{r}(0), \vec{s}(1)}(a = b) + \frac{1}{4} \Pr_{\vec{r}(1), \vec{s}(0)}(a = b) + \frac{1}{4} \Pr_{\vec{r}(1), \vec{s}(1)}(a \neq b) \\ &= \frac{1}{2} + \frac{1}{8} (\vec{r}(0) \cdot \vec{s}(0) + \vec{r}(0) \cdot \vec{s}(1) + \vec{r}(1) \cdot \vec{s}(0) - \vec{r}(1) \cdot \vec{s}(1)) \end{aligned}$$

is as large as possible. The second formula holds assuming we restrict to the x - z -plane. Thus, we want to choose the axes such that the angle between $\vec{r}(q)$ and $\vec{s}(r)$ is minimized for all q and r – except when $q = r = 1$, in which case we want to maximize the angle. The following picture shows a particular symmetric choice of axes:



Clearly, the three inner products are $\frac{1}{\sqrt{2}}$, while the last is $-\frac{1}{\sqrt{2}}$. Thus, this strategy achieves

$$p_{\text{win}} = \frac{1}{2} + \frac{1}{8} \cdot 4 \cdot \frac{1}{\sqrt{2}} = \frac{1}{2} \left(1 + \frac{1}{\sqrt{2}} \right),$$

which is exactly what we wanted to show, see Eq. (3.11). This shows that in a precise and quantitative sense, quantum theory enables stronger correlations than what is possible using a classical (hidden variable) theory!

It is natural to ask if we can do even better. Is it perhaps even possible to win the game with certainty using a quantum strategy? It turns out that this is not the case: The *Tsirelson bound* asserts that, for any quantum strategy:

$$p_{\text{win, quantum}} \leq \frac{1}{2} \left(1 + \frac{1}{\sqrt{2}} \right)$$

Thus the strategy that we derived above is optimal. Moreover, it is essentially unique – any other strategy that saturates the bound coincides with our strategy, except for a choice of bases (roughly speaking). This is known as the *rigidity property* of the CHSH game.

3.5 Exercises

3.1 Pure or mixed, entangled or separable: Let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$ and consider the following two-qubit state given with respect to the basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$:

$$\rho_{AB} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

- (a) Is ρ_{AB} pure or mixed? Justify your answer.
- (b) Is ρ_{AB} entangled or separable? Justify your answer.

3.2 Can correlations be shared?

- (a) Let $p \in \mathcal{P}(\Sigma_X \times \Sigma_Y)$ be a joint distribution of two random variables X, Y . Show that there always exists a joint distribution $q \in \mathcal{P}(\Sigma_X \times \Sigma_Y \times \Sigma_{Y'})$ of three random variables X, Y, Y' such that $\Sigma_{Y'} = \Sigma_Y$ and *both* the marginal distribution of X, Y as well as the marginal distribution of X, Y' are equal to p .
- (b) Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be an entangled pure state. Show that it is *impossible* to find a state $\sigma_{ABB'} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{B'})$ such that $\mathcal{H}_{B'} = \mathcal{H}_B$ and $\sigma_{AB} = \sigma_{AB'} = \rho_{AB}$.
Hint: Exercise 2.10.

This exercise shows that pure state entanglement is *monogamous* in a very strong sense. We will come back to this idea in Lecture 13.

3.3 Two-qubit entanglement: For any pure state $|\Psi\rangle = \Psi_{00}|00\rangle + \Psi_{01}|01\rangle + \Psi_{10}|10\rangle + \Psi_{11}|11\rangle$ of two qubits, define the following determinant-like quantity:

$$\Delta := \Psi_{00}\Psi_{11} - \Psi_{01}\Psi_{10}.$$

Show that $\Delta = 0$ if and only if $|\Psi\rangle$ is a product state.

3.4 Maximally entangled states:

- (a) Let $|\Phi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a maximally entangled state, and $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ an arbitrary pure state. Show that $|\Psi_{AB}\rangle$ is maximally entangled if and only if there exist unitaries $U_A \in \mathcal{U}(\mathcal{H}_A)$ and $V_B \in \mathcal{U}(\mathcal{H}_B)$ such that $(U_A \otimes V_B)|\Phi_{AB}\rangle = |\Psi_{AB}\rangle$.
- (b) Let $|\Phi_{A_1B_1}\rangle \in \mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}$ and $|\Phi'_{A_2B_2}\rangle \in \mathcal{H}_{A_2} \otimes \mathcal{H}_{B_2}$ be two maximally entangled states. Show that $|\Phi_{A_1B_1}\rangle \otimes |\Phi'_{A_2B_2}\rangle$ is maximally entangled between systems A_1A_2 and B_1B_2 .⁹

3.5 Maximal entanglement criterion: Prove Lemma 3.6.

3.6 Maximally entangled state tricks: Prove Lemma 3.7.

3.7 Partial transpose test I: Given a linear operator $M_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and a choice of orthonormal basis of \mathcal{H}_A , we define the *partial transpose* $M_{AB}^{\top_A}$ of M_{AB} on system A as the following operator in $\mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$: If $M_{AB} = M_A \otimes M_B$, we set

$$M_{AB}^{\top_A} := M_A^{\top} \otimes M_B,$$

⁹By this we mean that it is maximally entangled with respect to the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A = \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$ and $\mathcal{H}_B = \mathcal{H}_{B_1} \otimes \mathcal{H}_{B_2}$.

where the transpose is taken with respect to chosen basis (just for the ordinary transpose, the result will depend on this choice). Since any M_{AB} can be written as linear combination of operators of the form $M_A \otimes M_B$, we can extend the definition to all of $L(\mathcal{H}_A \otimes \mathcal{H}_B)$ by linearity. One can similarly define the partial transpose on system B .

(a) Show that if ρ_{AB} is a separable state then $\rho_{AB}^{\top_A}$ is PSD.

Thus we obtain an important criterion for entanglement, which is known as the *partial transpose test*: if $\rho_{AB}^{\top_A}$ has a negative eigenvalue, then the state ρ_{AB} must be entangled.

(b) Use this to show that the maximally entangled state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is entangled.

It is important to realize that the partial transpose test only gives a *sufficient* criterion for entanglement. In other words, there are entangled states with PSD partial transpose (except in dimensions 2×2 , 2×3 , and 3×2).

3.8 Partial transpose test II: Let $\rho_0 = |\Phi^+\rangle\langle\Phi^+|$ denote the maximally entangled state on $\mathbb{C}^d \otimes \mathbb{C}^d$, and let $\rho_1 = \frac{I \otimes I - |\Phi^+\rangle\langle\Phi^+|}{d^2 - 1}$.

(a) Show that $\rho(t) := (1 - t)\rho_0 + t\rho_1$ is a quantum state for all $t \in [0, 1]$.

(b) Show that $\rho_0^{\top_A} = \frac{1}{d}F$, where $F \in L(\mathbb{C}^d \otimes \mathbb{C}^d)$ is the *swap operator* defined by

$$F(|x\rangle \otimes |y\rangle) = |y\rangle \otimes |x\rangle,$$

for all $x, y \in \{0, \dots, d - 1\}$.

(c) For what range of t does the partial transpose test show that $\rho(t)$ is entangled?

3.9 Practice In this problem you can experiment with the partial transpose test. In the files `03-state-A.txt`, `03-state-B.txt` and `03-state-C.txt`, you will find three density matrices of the following dimensions:

$$\alpha \in D(\mathbb{C}^2 \otimes \mathbb{C}^2), \quad \beta \in D(\mathbb{C}^2 \otimes \mathbb{C}^3), \quad \gamma \in D(\mathbb{C}^2 \otimes \mathbb{C}^4).$$

For each of the three states, compute the partial transpose, and output the smallest eigenvalue of the resulting matrix. For each state, output whether the state is entangled, or separable, or whether the partial transpose test was inconclusive.

Hint: You may use that the partial transpose test is necessary and sufficient when the total dimension is at most 6.

3.10 Local conversion of Bell states: Verify Eq. (3.5).

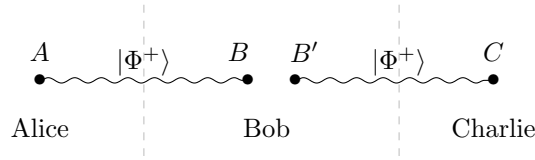
3.11 Teleportation: Here you can fill in the gaps left in the lecture notes in the analysis of teleportation, where we asserted Eq. (3.7) and its generalization Eq. (3.8) without proof.

(a) Let $I_{B \rightarrow M}$ denote the identity operator $\mathcal{H}_B \rightarrow \mathcal{H}_M$ (this make sense as $\mathcal{H}_M = \mathcal{H}_B = \mathbb{C}^2$). Show that $(I_M \otimes \langle\Phi_{AB}^+|)(|\Phi_{MA}^+\rangle \otimes I_B) = \frac{1}{2}I_{B \rightarrow M}$.

(b) Deduce that $(I_M \otimes \langle\Phi_{AB}^+|)(|\Phi_{MA}^{(x,z)}\rangle \otimes I_B) = \frac{1}{2}I_{B \rightarrow M}Z^zX^x$ for any $x, z \in \{0, 1\}$.

(c) Verify Eq. (3.8).

3.12 Entanglement swapping: Consider the following scenario involving three friends Alice, Bob, and Charlie: Alice has one qubit A , Bob has two qubits B and B' , and Charlie has one qubit C . The joint state of the four qubits is $|\Phi_{AB}^+\rangle \otimes |\Phi_{B'C}^+\rangle$, as in the following picture:



- (a) Compute the reduced state of subsystem AC (i.e., the qubits shared by Alice and Charlie). Is this state correlated or not? If it is correlated, is it entangled?
- (b) Explain in one or two sentences how the three friends can use teleportation to create a maximally entangled state between Alice and Charlie.

3.13 Transpose trick: For any $\vec{r}, \vec{s} \in \mathbb{R}^3$, show that

$$\langle \Phi^+ | (I + r_x X + r_y Y + r_z Z) \otimes (I + s_x X + s_y Y + s_z Z) | \Phi^+ \rangle = 1 + r_x s_x - r_y s_y + r_z s_z.$$

3.14 Measuring a maximally state: Suppose that Alice and Bob each have one qubit, and their joint state is the maximally entangled state $|\Phi^+\rangle$. Use Eq. (3.14) to compute the probability that their measurement outcomes coincide in each of the following situations:

- (a) Both perform a standard basis measurement (1.26).
- (b) Both perform a Hadamard basis measurement (1.27).
- (c) Alice performs a standard basis measurement and Bob a Hadamard basis measurement.

Lecture 4

Trace distance and fidelity, classical and quantum channels

Today, we are concerned with two separate topics. First, we discuss ways of quantifying the similarity of two quantum states. For this we introduce two quantities: the *trace distance*, which is defined in terms of generalization of the ℓ^1 -norm of vectors, and the *fidelity*, which generalizes the overlap $|\langle\phi|\psi\rangle|$ of pure states. Both are useful and we discuss their most important properties.

We then switch gears and spend the remainder of the lecture to work towards the definition of a *quantum channel*. Roughly speaking, channels describe the most general way by which we can modify (or ‘process’) quantum states. We first discuss the classical situation and then turn towards the quantum case – motivating and defining quantum channels as *completely positive* and *trace-preserving superoperators*. We will continue the discussion of channels next week. At that point, we will have fully developed the basic formalism of quantum information theory.

4.1 Norms of operators

Since quantum states are operators, we can in principle use any norm or metric on $L(\mathcal{H})$ to define a distance measure. What are useful norms to consider?

There is a general procedure to define the norm of an operator by considering the ℓ^p -norm of its singular values (Theorem 2.23). Recall that the ℓ^p -norm of a vector $x \in \mathbb{C}^n$ is defined by

$$\|x\|_p := \begin{cases} \left(\sum_{j=1}^n |x_j|^p\right)^{1/p} & \text{if } p \in [1, \infty), \\ \max_{j=1}^n |x_j| & \text{if } p = \infty. \end{cases} \quad (4.1)$$

Note that $\|x\|_\infty = \lim_{p \rightarrow \infty} \|x\|_p$. We can define corresponding norms of operators as follows:

Definition 4.1 (Schatten p -norm, trace norm, Frobenius norm, operator norm). For any $p \in [1, \infty]$, we define the *Schatten p -norm* of an operator $M \in L(\mathcal{H}, \mathcal{K})$ by

$$\|M\|_p := \left\| \begin{pmatrix} s_1 \\ \vdots \\ s_r \end{pmatrix} \right\|_p,$$

where $s_1, \dots, s_r > 0$ are the singular values of M .

The cases $p \in \{1, 2, \infty\}$ are particularly important and have special names: $\|M\|_1$ is called the *trace norm* (or nuclear norm), $\|M\|_2$ is the *Frobenius norm* (or Hilbert-Schmidt norm), and $\|M\|_\infty$ is the *operator norm* (or spectral norm).

The Schatten p -norms are norms on $L(\mathcal{H}, \mathcal{K})$ for any $p \in [1, \infty]$. This is not obvious for general p , but easy to verify for $p \in \{1, 2, \infty\}$ by using properties discussed below. If M is Hermitian, then the singular values are the absolute values of the nonzero eigenvalues (Exercise 2.11), so $\|M\|_p$ is the ℓ^p -norm of the eigenvalues.

Let us write down explicit formulas for the three important special cases: The *trace norm* of an operator $M \in L(\mathcal{H}, \mathcal{K})$ is given by

$$\|M\|_1 = \sum_{i=1}^r s_i = \text{Tr} \sqrt{M^\dagger M} = \text{Tr} \sqrt{M M^\dagger}. \quad (4.2)$$

To see the right-hand side expressions, recall that the singular values are the nonzero eigenvalues of the operator $\sqrt{M^\dagger M}$ or $\sqrt{M M^\dagger}$. If M is PSD then $\|M\|_1 = \text{Tr}[M]$, since $\sqrt{M^\dagger M} = \sqrt{M^2} = M$.

Similarly, the *Frobenius norm* of an operator $M \in L(\mathcal{H}, \mathcal{K})$ is given by

$$\|M\|_2 = \left(\sum_{i=1}^r s_i^2 \right)^{1/2} = \left(\text{Tr} M^\dagger M \right)^{1/2} = \left(\text{Tr} M M^\dagger \right)^{1/2}. \quad (4.3)$$

Just like the ℓ^2 -norm of vectors, the Frobenius norm is induced by an inner product – namely the so-called *Hilbert-Schmidt inner product* on $L(\mathcal{H}, \mathcal{K})$, which is defined by

$$\langle M, N \rangle := \text{Tr}[M^\dagger N] \quad \forall M, N \in L(\mathcal{H}, \mathcal{K}). \quad (4.4)$$

Thus, $L(\mathcal{H}, \mathcal{K})$ is itself a Hilbert space if we use this inner product. In fact, under the natural isomorphism $L(\mathcal{H}, \mathcal{K}) \cong \mathcal{H}^* \otimes \mathcal{K}$, the Frobenius norm and Hilbert-Schmidt inner product simply corresponds to the norm and inner product of the Hilbert space $\mathcal{H}^* \otimes \mathcal{K}$. Concretely, this means the following: For an operator M , denote by $M_{ab} = \langle e_a | M | f_b \rangle$ its matrix elements with respect to arbitrary fixed orthonormal bases of $|e_a\rangle$ of \mathcal{K} and $|f_b\rangle$ of \mathcal{H} . Then, Eqs. (4.3) and (4.4) become

$$\|M\|_2 = \sum_{a,b} |M_{ab}|^2 \quad \text{and} \quad \langle M, N \rangle = \sum_{a,b} \overline{M_{ab}} N_{ab}.$$

This shows that the Frobenius norm and Hilbert-Schmidt inner product are simply the ℓ^2 -norm and inner product of the matrix entries, thought of as vectors in \mathbb{C}^{d^2} .

Finally, we can write the *operator norm* as follows:

$$\|M\|_\infty = \max_{i=1}^r s_i = \max_{|\phi\rangle \in \mathcal{H}, \|\phi\| \leq 1} \|M|\phi\rangle\|. \quad (4.5)$$

The right-hand side expression follows since its square is the maximal eigenvalue of $M^\dagger M$. It shows that $\|\cdot\|_\infty$ is the operator norm induced by the norms of \mathcal{H} and \mathcal{K} , which justifies its name.

Lemma 4.2. *The Schatten norms satisfy the following properties for all $M \in L(\mathcal{H}, \mathcal{K})$:*

- (a) *Invariance under taking adjoints, conjugation, and transposition (the latter with respect to any orthonormal basis):* $\|M\|_p = \|M^\dagger\|_p = \|\overline{M}\|_p = \|M^\top\|_p$.
- (b) *Invariance under isometries:* $\|M\|_p = \|V M W^\dagger\|_p$ for all $V \in U(\mathcal{K}, \mathcal{K}')$, $W \in U(\mathcal{H}, \mathcal{H}')$. In particular, they are invariant under left and right multiplication by unitaries.
- (c) *They are monotonically decreasing in p . In particular:* $\|M\|_1 \geq \|M\|_2 \geq \|M\|_\infty$.

Proof. Parts (a) and (b) follows directly from the corresponding properties of the singular values. Part (c) is a direct consequence of the same property for the ordinary ℓ^p -norms. \square

The Schatten norms also satisfy a version of the *Hölder inequality*: For $\frac{1}{p} + \frac{1}{q} = 1$,

$$|\mathrm{Tr}[M^\dagger N]| \leq \|M\|_p \|N\|_q \quad \forall M, N \in \mathcal{L}(\mathcal{H}, \mathcal{K}).$$

In fact, if $\frac{1}{p} + \frac{1}{q} = 1$ then the norm $\|\cdot\|_q$ is dual to $\|\cdot\|_p$.¹ This means that

$$\|M\|_q = \max_{\substack{N \in \mathcal{L}(\mathcal{H}, \mathcal{K}), \\ \|N\|_p \leq 1}} |\mathrm{Tr}[M^\dagger N]| \quad \forall M \in \mathcal{L}(\mathcal{H}, \mathcal{K}).$$

We record and prove two important special cases to give you a flavor of the reasoning:

Lemma 4.3 (Cauchy-Schwarz and Hölder). *For any $M \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, we have*

$$\|M\|_2 = \max_{\substack{N \in \mathcal{L}(\mathcal{H}, \mathcal{K}), \\ \|N\|_2 \leq 1}} |\mathrm{Tr}[M^\dagger N]|, \quad (4.6)$$

$$\|M\|_1 = \max_{\substack{N \in \mathcal{L}(\mathcal{H}, \mathcal{K}), \\ \|N\|_\infty \leq 1}} |\mathrm{Tr}[M^\dagger N]|. \quad (4.7)$$

In particular, we have the following Cauchy-Schwarz and Hölder inequality: For $M, N \in \mathcal{L}(\mathcal{H}, \mathcal{K})$,

$$|\mathrm{Tr}[M^\dagger N]| \leq \|M\|_2 \|N\|_2, \quad (4.8)$$

$$|\mathrm{Tr}[M^\dagger N]| \leq \|M\|_1 \|N\|_\infty. \quad (4.9)$$

Proof. The Cauchy-Schwarz inequality holds for any inner product, so in particular for (4.4). Thus we obtain Eq. (4.8), which also shows ‘ \geq ’ in Eq. (4.6). For the other direction choose $N = \frac{M}{\|M\|_2}$.

To prove Eq. (4.7) and hence Eq. (4.9), let $M = \sum_i s_i |e_i\rangle\langle g_i|$ be a singular value decomposition of M . Then:

$$|\mathrm{Tr}[M^\dagger N]| = \left| \mathrm{Tr} \left[\sum_i s_i |g_i\rangle\langle e_i| N \right] \right| = \left| \sum_i s_i \langle e_i | N | g_i \rangle \right| \leq \sum_i s_i \underbrace{|\langle e_i | N | g_i \rangle|}_{\leq \|N\|_\infty} \leq \|M\|_1 \|N\|_\infty,$$

where we estimated the underbraced inner product by using the Cauchy-Schwarz inequality for vectors in \mathcal{K} and then Eq. (4.5). For $N = \sum_i |e_i\rangle\langle g_i|$, the above inequalities hold with equality. \square

The case that $\mathcal{H} = \mathcal{K}$ is so important that we re-state Eq. (4.7) with a slight extension.

Lemma 4.4. *For all $M \in \mathcal{L}(\mathcal{H})$, we have*

$$\|M\|_1 = \max_{\substack{N \in \mathcal{L}(\mathcal{H}), \\ \|N\|_\infty \leq 1}} |\mathrm{Tr}[MN]| = \max_{U \in \mathcal{U}(\mathcal{H})} |\mathrm{Tr}[MU]| \geq |\mathrm{Tr}[M]|.$$

Proof. The first equality is just Eq. (4.7) for $\mathcal{H} = \mathcal{K}$ and M substituted by M^\dagger (which has the same trace norm). For the second equality, note that ‘ \geq ’ holds since $\|U\|_\infty = 1$ for any unitary $U \in \mathcal{U}(\mathcal{H})$, while for ‘ \leq ’ we note that, since $\mathcal{H} = \mathcal{K}$, we can extend the map N from the proof of Eq. (4.7) to a unitary. The final inequality is obvious: simply choose $U = I$. \square

We can also prove other variants of the Hölder inequalities, such as the following. It generalizes and strengthens Eq. (4.9), since by Lemma 4.4 the trace norm is never smaller than the trace.

¹In finite dimensions, this is true for all $p \in [1, \infty]$.

Lemma 4.5. For all $M \in \mathcal{L}(\mathcal{K}, \mathcal{L})$, $N \in \mathcal{L}(\mathcal{H}, \mathcal{K})$, we have $\|MN\|_1 \leq \|M\|_1 \|N\|_\infty$.

Proof. By Eq. (4.7),

$$\|MN\|_1 = \max_{\substack{X \in \mathcal{L}(\mathcal{H}, \mathcal{L}), \\ \|X\|_\infty \leq 1}} |\text{Tr}[(MN)^\dagger X]| = \max_{\substack{X \in \mathcal{L}(\mathcal{H}, \mathcal{L}), \\ \|X\|_\infty \leq 1}} |\text{Tr}[MN X^\dagger]|.$$

To estimate the right-hand side, observe that

$$|\text{Tr}[MN X^\dagger]| \leq \|M\|_1 \|N X^\dagger\|_\infty \leq \|M\|_1 \|N\|_\infty \|X\|_\infty \leq \|M\|_1 \|N\|_\infty,$$

where we first use (4.9), then submultiplicativity of the operator norm, and finally $\|X\|_\infty \leq 1$. \square

Finally, we note that the Schatten p -norms are all *submultiplicative*, which means that

$$\|MN\|_p \leq \|M\|_p \|N\|_p \quad \forall M \in \mathcal{L}(\mathcal{K}, \mathcal{L}), N \in \mathcal{L}(\mathcal{H}, \mathcal{K}).$$

For $p \in \{1, 2, \infty\}$ this is easy to verify directly using the properties established above.

4.2 Trace distance and fidelity

We can use the norms from Section 4.1 to define distance measures between quantum states. One particular important such distance measure is the *trace distance*, which is simply one half times the metric induced by the trace norm. In particular, the trace distance is a metric.

Definition 4.6 (Trace distance). The (*normalized*) *trace distance* between states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ is

$$T(\rho, \sigma) := \frac{1}{2} \|\rho - \sigma\|_1.$$

In Exercise 4.2, you can prove *Helstrom's theorem*, which gives an important *operational interpretation* to the trace distance. Namely, it shows that the optimal probability of distinguishing two quantum states ρ and σ is $\frac{1}{2} + \frac{1}{2}T(\rho, \sigma)$ assuming you are given one of the two with equal probability. This follows readily from the following lemma, which you get to prove in Exercise 4.1.

Lemma 4.7 (Variational characterization). For any two states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$,

$$T(\rho, \sigma) = \max_{0 \leq Q \leq I} \text{Tr}[Q(\rho - \sigma)].$$

Moreover, the maximum is achieved by a projection Q .

We now list some useful properties of the trace distance:

Lemma 4.8. (a) $T(\rho, \sigma) \in [0, 1]$ for $\rho, \sigma \in \mathcal{D}(\mathcal{H})$. Moreover, $T(\rho, \sigma) = 0$ if and only if $\rho = \sigma$.
 (b) Invariance under isometries: $T(\rho, \sigma) = T(V\rho V^\dagger, V\sigma V^\dagger)$ for $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and $V \in \mathcal{U}(\mathcal{H}, \mathcal{K})$.
 (c) Monotonicity: $T(\rho_A, \sigma_A) \leq T(\rho_{AB}, \sigma_{AB})$ for all states $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$.
 (d) Joint convexity: $T(\sum_{x \in \Sigma} p_x \rho_x, \sum_{x \in \Sigma} p_x \sigma_x) \leq \sum_{x \in \Sigma} p_x T(\rho_x, \sigma_x)$, where $p \in \mathcal{P}(\Sigma)$ is an arbitrary probability distribution and ρ_x, σ_x are arbitrary states for $x \in \Sigma$.

You can prove this in Exercise 4.4. Property (c) is quite intuitive, since it means that two states can only get closer if we discard a subsystem.

For pure states $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$, the trace distance is directly related to the *overlap* $|\langle\phi|\psi\rangle|$ of the corresponding vectors (Exercise 4.5):

$$T(\rho, \sigma) = \sqrt{1 - |\langle\phi|\psi\rangle|^2} \quad (4.10)$$

Is there also a useful general definition of an ‘overlap’ of mixed states? This leads us to our second definition, which is the following.

Definition 4.9 (Fidelity). The *fidelity* between two states $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ is defined as

$$F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \text{Tr}[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}] = \text{Tr}[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}].$$

For the second and third equality, see Eq. (4.2). Clearly, F is symmetric, i.e., $F(\rho, \sigma) = F(\sigma, \rho)$.

If $\rho = |\psi\rangle\langle\psi|$ is pure then $\sqrt{\rho} = \rho$, so

$$F(\rho, \sigma) = \text{Tr}[\underbrace{\sqrt{|\psi\rangle\langle\psi|\sigma|\psi\rangle\langle\psi|}}_{\geq 0}] = \sqrt{\langle\psi|\sigma|\psi\rangle} \underbrace{\text{Tr}[\sqrt{|\psi\rangle\langle\psi|}]}_{=\text{Tr}|\psi\rangle\langle\psi|=1} = \sqrt{\langle\psi|\sigma|\psi\rangle}. \quad (4.11)$$

If both states are pure, $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$, then

$$F(\rho, \sigma) = \sqrt{\langle\psi|\phi\rangle\langle\phi|\psi\rangle} = |\langle\psi|\phi\rangle|. \quad (4.12)$$

Thus, the fidelity indeed generalizes the overlap of pure states.

The fidelity is a *similarity measure* rather than a distance measure, i.e., it is *maximized* if the two states are the same.² This follows from the first item in the following lemma.

Lemma 4.10. (a) $F(\rho, \sigma) \in [0, 1]$ for $\rho, \sigma \in \mathcal{D}(\mathcal{H})$. Moreover, $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$.
(b) Invariance under isometries: $F(\rho, \sigma) = F(V\rho V^\dagger, V\sigma V^\dagger)$ for $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and $V \in \mathcal{U}(\mathcal{H}, \mathcal{K})$.
(c) Monotonicity: $F(\rho_A, \sigma_A) \geq F(\rho_{AB}, \sigma_{AB})$ for all states $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$.
(d) Joint concavity: $F(\sum_{x \in \Sigma} p_x \rho_x, \sum_{x \in \Sigma} p_x \sigma_x) \geq \sum_{x \in \Sigma} p_x F(\rho_x, \sigma_x)$, where $p \in \mathcal{P}(\Sigma)$ is an arbitrary probability distribution and ρ_x, σ_x are arbitrary states for $x \in \Sigma$.

You can prove the lemma in Exercise 4.6. Part (b) is easy to see directly; for the rest it is useful to use Uhlmann’s theorem, which we discuss below. Note that the inequality in (c) and (d) go the opposite way than for the trace distance. This is intuitive, since the trace distance is a distance measure, while the fidelity is a similarity measure.

Remark 4.11 (Why so complicated?). Why don’t we simply use $\sqrt{\text{Tr}[\rho\sigma]}$ to generalize the overlap? The problem is that this quantity does not attain its maximum for when $\rho = \sigma$. Indeed, $\text{Tr}[\rho^2]$ can be any number in $[1/d, 1]$, where $d = \dim \mathcal{H}$, so the above is not a good definition.

Remark 4.12 (Tricky conventions). Around half of the quantum information community defines the fidelity as the *square* of our $F(\rho, \sigma)$. This is good to keep in mind when consulting the literature (including textbooks).

We now come to a central result, Uhlmann’s theorem, which gives a nice interpretation of the fidelity. Namely, the fidelity is simply the maximal overlap between two purifications!

²Thus, the fidelity is not a metric. However, $P(\rho, \sigma) := \sqrt{1 - F^2(\rho, \sigma)}$ is a metric, called the *purified distance*.

Theorem 4.13 (Uhlmann). *Let $\rho_A, \sigma_A \in \mathcal{D}(\mathcal{H}_A)$ be states and let \mathcal{H}_B be a Hilbert space such that both states admit purifications on $\mathcal{H}_A \otimes \mathcal{H}_B$.³ Then,*

$$F(\rho_A, \sigma_A) = \max \left\{ |\langle \Psi_{AB} | \Phi_{AB} \rangle| : |\Psi_{AB}\rangle, |\Phi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \text{ purifications of } \rho_A, \sigma_A \right\}.$$

Since any two purifications on the same Hilbert space are related by a unitary (Lemma 2.18), Theorem 4.13 can equivalently be stated as follows:

$$F(\rho_A, \sigma_A) = \max_{U_B \in \mathcal{U}(\mathcal{H}_B)} |\langle \Psi_{AB}^{\text{fix}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{fix}} \rangle|, \quad (4.13)$$

where $|\Psi_{AB}^{\text{fix}}\rangle$ and $|\Phi_{AB}^{\text{fix}}\rangle$ are arbitrary fixed purifications of ρ_A and σ_A , respectively.

We first give a proof under the simplifying assumption that $\mathcal{H}_A = \mathcal{H}_B$. See below for a general proof, which is slightly more technical.

Proof of Theorem 4.13 (if $\mathcal{H}_A = \mathcal{H}_B$). Then, without loss of generality, $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^d$, so we can use the standard purifications of ρ_A and σ_A , respectively (Definition 2.19). That is:

$$|\Psi_{AB}^{\text{std}}\rangle := \left(\sqrt{\rho_A} \otimes I_B \right) \sum_x |x\rangle \otimes |x\rangle, \quad |\Phi_{AB}^{\text{std}}\rangle := \left(\sqrt{\sigma_A} \otimes I_B \right) \sum_x |x\rangle \otimes |x\rangle. \quad (4.14)$$

We will now prove Eq. (4.13), using these as the ‘fixed’ purifications. For $U_B \in \mathcal{U}(\mathcal{H}_B)$,

$$\begin{aligned} |\langle \Psi_{AB}^{\text{std}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{std}} \rangle| &= \sum_{x,y} (\langle x| \otimes \langle x|) (\sqrt{\rho_A} \sqrt{\sigma_A} \otimes U_B) (|y\rangle \otimes |y\rangle) \\ &= \sum_{x,y} \langle x | \sqrt{\rho_A} \sqrt{\sigma_A} | y \rangle \langle x | U | y \rangle \\ &= \sum_{x,y} \langle x | \sqrt{\rho_A} \sqrt{\sigma_A} | y \rangle \langle y | U^\top | x \rangle \\ &= \sum_x \langle x | \sqrt{\rho_A} \sqrt{\sigma_A} U_A^\top | x \rangle = \text{Tr} \left[\sqrt{\rho_A} \sqrt{\sigma_A} U_A^\top \right]. \end{aligned}$$

In the first step we inserted Eq. (4.14), next we used Eq. (2.3), then we perform the transpose, and finally we use Eqs. (1.3) and (1.4). (The unitaries $U_B = U = U_A$ are all the same objects; the subscripts are just notation to help us emphasize on which system the operator acts.) By maximizing the left and right hand side of this equality, we obtain

$$\begin{aligned} \max_{U_B \in \mathcal{U}(\mathcal{H}_B)} |\langle \Psi_{AB}^{\text{std}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{std}} \rangle| &= \max_{U_A \in \mathcal{U}(\mathcal{H}_A)} \text{Tr} \left[\sqrt{\rho_A} \sqrt{\sigma_A} U_A^\top \right] \\ &= \max_{U_A \in \mathcal{U}(\mathcal{H}_A)} \text{Tr} \left[\sqrt{\rho_A} \sqrt{\sigma_A} U_A \right] \\ &= \|\sqrt{\rho_A} \sqrt{\sigma_A}\|_1 = F(\rho_A, \sigma_A). \end{aligned}$$

The second step uses that $U \mapsto U^\top$ is a bijection of the set of unitaries and the third equality is precisely the second characterization in Lemma 4.4. Thus we have proved Eq. (4.13), and thereby the theorem. \square

We now show how to adapt the preceding proof in the general case that \mathcal{H}_A and \mathcal{H}_B are not necessarily the same. The key difference is that we can no longer use the standard purification.

³Recall from Lemma 2.17 that this means that $\dim \mathcal{H}_B \geq \max\{\text{rank } \rho_A, \text{rank } \sigma_A\}$.

Proof of Theorem 4.13 (general case). Let $\mathcal{H}_R = \mathbb{C}^r$ be an auxiliary system of dimension $r = \max\{\text{rank}(\rho_A), \text{rank}(\sigma_A)\}$. We consider the following purifications:

$$|\Psi_{AB}^{\text{fix}}\rangle := \left(\sqrt{\rho_A} V_{R \rightarrow A} \otimes X_{R \rightarrow A} \right) \sum_{x=1}^r |x\rangle \otimes |x\rangle, \quad (4.15)$$

$$|\Phi_{AB}^{\text{fix}}\rangle := \left(\sqrt{\sigma_A} W_{R \rightarrow A} \otimes X_{R \rightarrow A} \right) \sum_{x=1}^r |x\rangle \otimes |x\rangle. \quad (4.16)$$

The operator $V_{R \rightarrow A}$ is an isometry that maps the first $\text{rank}(\rho_A)$ many standard basis vectors $|x\rangle$ to orthonormal eigenvectors of ρ_A corresponding to the nonzero eigenvalues. Likewise, the operator $W_{R \rightarrow A}$ is an isometry that maps the first $\text{rank}(\sigma_A)$ many standard basis vectors $|x\rangle$ to orthonormal eigenvectors of σ_A corresponding to the nonzero eigenvalues. (This is possible since $\dim \mathcal{H}_A \geq r$.) Finally, the operator $X_{R \rightarrow B}$ is an arbitrary isometry. (Such isometries exist we assumed that both ρ_A and σ_A have purifications to $\mathcal{H}_A \otimes \mathcal{H}_B$, so $\dim \mathcal{H}_B \geq r$.) It is easy to verify that Eq. (4.15) defines purifications of ρ_A and σ_A .

We now proceed as before and consider the right-hand side of Eq. (4.13), but now using Eq. (4.15) as the fixed purifications. Now, abbreviating $V = V_{R \rightarrow A}$, $W = W_{R \rightarrow A}$, $X = X_{R \rightarrow B}$,

$$\begin{aligned} |\langle \Psi_{AB}^{\text{fix}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{fix}} \rangle| &= \sum_{x,y=1}^r (\langle x| \otimes \langle y|) (V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W \otimes X^\dagger U_B X) (|y\rangle \otimes |y\rangle) \\ &= \sum_{x,y=1}^r \langle x| V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W |y\rangle \langle x| X^\dagger U_B X |y\rangle \\ &= \sum_{x,y=1}^r \langle x| V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W |y\rangle \langle y| (X^\dagger U_B X)^\top |x\rangle \\ &= \text{Tr}[V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W (X^\dagger U_B X)^\top] \end{aligned}$$

What kind of object are $X^\dagger U_B X$ and its transpose? This is an operator on $\mathcal{H}_R = \mathbb{C}^r$ which we can think of as the restriction of the unitary U_B to a subspace (namely the image $\text{im}(X)$ of the isometry X). As such, it is clear that $\|X^\dagger U_B X\|_\infty \leq 1$, which can also be seen formally by using submultiplicativity and the fact that unitaries and (more generally isometries) have operator norm at most one. This means that

$$\max_{U_B \in \text{U}(\mathcal{H}_B)} |\langle \Psi_{AB}^{\text{fix}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{fix}} \rangle| \leq \max_{\substack{Y \in \text{L}(\mathcal{H}_R), \\ \|Y\|_\infty \leq 1}} \text{Tr}[V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W Y] = \|V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W\|_1 \quad (4.17)$$

using the first characterization in Lemma 4.4. On the other hand, we can write any unitary matrix in $\text{U}(\mathcal{H}_R)$ as $X^\dagger U_B X$ for some unitary $U_B \in \text{U}(\mathcal{H}_B)$ (simply choose U_B to be a direct sum of the desired unitary on $\mathcal{H}_R \cong \text{im}(X)$ and the identity on the orthogonal complement). Hence:

$$\max_{U_B \in \text{U}(\mathcal{H}_B)} |\langle \Psi_{AB}^{\text{fix}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{fix}} \rangle| \geq \max_{Z \in \text{U}(\mathcal{H}_R)} \text{Tr}[V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W Z] = \|V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W\|_1 \quad (4.18)$$

Combining Eqs. (4.17) and (4.18), we find that

$$\max_{U_B \in \text{U}(\mathcal{H}_B)} |\langle \Psi_{AB}^{\text{fix}} | (I_A \otimes U_B) | \Phi_{AB}^{\text{fix}} \rangle| = \|V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W\|_1. \quad (4.19)$$

We are almost done – but we still have to get rid of V and W on the right-hand side. To do so, note that

$$\|\sqrt{\rho_A} \sqrt{\sigma_A}\|_1 = \|V V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W W^\dagger\|_1 \leq \|V^\dagger \sqrt{\rho_A} \sqrt{\sigma_A} W\|_1 \leq \|\sqrt{\rho_A} \sqrt{\sigma_A}\|_1 \quad (4.20)$$

The equality holds, because VV^\dagger projects onto the support of ρ_A , hence of $\sqrt{\rho_A}$, so $VV^\dagger\sqrt{\rho_A} = \sqrt{\rho_A}$; and likewise for WW^\dagger and $\sqrt{\sigma_A}$. The inequalities follow from Lemma 4.5, since the isometries V and W and their adjoints satisfy have operator norm bounded by one. (In the fact, the first inequality is an equation thanks to Lemma 4.2 (b).) Since the left and the right hand side of Eq. (4.20) are the same, it follows that we must have equality throughout, so that

$$\|\sqrt{\rho_A}\sqrt{\sigma_A}\|_1 = \|V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}W\|_1.$$

In view of Eq. (4.19) this concludes the proof. Phew! \square

Finally, we mention that the trace distance and fidelity are related by the so-called *Fuchs-van de Graaf inequalities*: For all $\rho, \sigma \in \mathcal{D}(\mathcal{H})$,

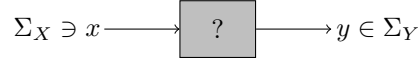
$$1 - F(\rho, \sigma) \leq T(\rho, \sigma) \leq \sqrt{1 - F^2(\rho, \sigma)}. \quad (4.21)$$

You can prove the upper bound in Exercise 4.7. For pure states, we see from comparing Eqs. (4.10) and (4.12) that the upper bound is an equality.

4.3 Channels in probability theory

So far, the only way to manipulate a quantum state has been to measure it – but apart from Footnote 3 we have not discussed at all how quantum states can evolve or be manipulated. In the remainder of today's lecture we will start developing the mathematical formalism that describes the most general way by which quantum states can be manipulated. Before considering the quantum situation, it is instructive to consider the classical situation.

Suppose we are given a 'box' such that we can input a value $x \in \Sigma_X$ and receive as output some value $y \in \Sigma_Y$, as in the following figure:



How should we describe this mathematically? Let us imagine that the box has no memory, i.e., it acts the same way if we use it repeatedly. Then the most straightforward description might be to assume that there exists a function,

$$f: \Sigma_X \rightarrow \Sigma_Y$$

such that $y = f(x)$ for every input x . This is an excellent description if we have engineered the box ourselves to perform a given operation deterministically. But what about if there is some random process happening inside the box? (Or perhaps we have some uncertainty about the inner workings of the box?) In this case, it is natural to allow the output to be *random*, i.e., described by a probability distribution. Mathematically, this means that we should describe the box by a function

$$p(y|x) \quad \text{such that} \quad \begin{cases} p(y|x) \geq 0 & \forall x, y \\ \sum_y p(y|x) = 1 & \forall x. \end{cases} \quad (4.22)$$

The right-hand side condition means that $p(y|x)$ is a probability distribution in y for every fixed x . The interpretation is that if the input to the box is x , then the output y is random, with probabilities given by the $p(y|x)$. That is,

$$p(y|x) = \Pr(\text{output } y | \text{input } x).$$

For this reason we might call $p(y|x)$ a *conditional probability distribution* (but note that we do *not* presuppose the existence of a joint distribution). In information theory, $p(y|x)$ is called a (*memoryless*) *channel*. We will mostly use these two terms. Other terms are (*column*) *stochastic matrix*, *transition operator*, or *Markov operator*.

Remark 4.14 (Functions as channels). Note that given a function $f: \Sigma_X \rightarrow \Sigma_Y$, we can always define

$$p(y|x) = \begin{cases} 1 & \text{if } y = f(x), \\ 0 & \text{otherwise.} \end{cases}$$

Then, if x is the input then $y = f(x)$ is the output with certainty. This shows that channels are a generalization of deterministic functions.

What if the input is also random, say, given by some probability distribution $p(x)$? In this case, the joint probability of input and output is given by $p(x, y) = p(y|x)p(x)$, so the distribution of the output is the marginal distribution of y ,

$$p(y) = \sum_x p(y|x)p(x). \quad (4.23)$$

Here we used the slightly terrible (but concise and rather standard) convention of writing $p(x)$ and $p(y)$ for the input and output distribution, respectively, only distinguishing them by the symbol used for the argument. It would be more precise to use subscripts – writing, say, p_X and p_Y for the input and output distribution, and $P_{Y|X}$ for the channel. Then, Eq. (4.23) reads

$$p_Y(y) = \sum_x P_{Y|X}(y|x)p_X(x). \quad (4.24)$$

Note that this is precisely the formula for matrix-by-vector multiplication – provided we think of the probability distributions $p_X \in \mathbb{R}^{\Sigma_X}$ and $p_Y \in \mathbb{R}^{\Sigma_Y}$ as vectors, and of the channel as a matrix $P_{Y|X} \in \mathbb{R}^{\Sigma_Y \times \Sigma_X}$ (the entry in row y and column x is $P_{Y|X}(y|x)$). The conditions in Eq. (4.22) mean that all entries are nonnegative and that the entries in each column sum to one – such matrices are called (*column*) *stochastic*. Then, the formula Eq. (4.24) for computing the output distribution given a channel and input distribution can be succinctly written as follows:

$$p_Y = P_{Y|X} p_X$$

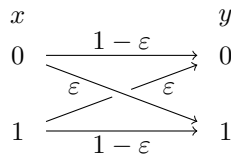
The mapping $p_X \mapsto p_Y$ is evidently linear (since it is implemented by left multiplication with the channel matrix $P_{Y|X}$). Conversely, any linear mapping that sends probability distributions to probability distributions is of this form for some channel $P_{Y|X}$.

Let us discuss two families of channels that are very important in classical information theory.

- (a) A *binary symmetric channel* is a channel which flips a bit with some probability $\varepsilon \in [0, 1]$. That is, $\Sigma_X = \Sigma_Y = \{0, 1\}$ and

$$\begin{aligned} p(0|0) &= p(1|1) = 1 - \varepsilon, \\ p(1|0) &= p(0|1) = \varepsilon \end{aligned}$$

We can visualize this as follows:

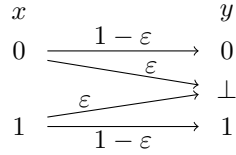


Note that output y does not contain any information about whether the bit has been flipped. This is perhaps the most straightforward way of modeling an unreliable digital information transmission line.

- (b) A *binary erasure channel* is a channel where the input bit is lost (‘erased’) with some probability $\varepsilon \in [0, 1]$. Mathematically, $\Sigma_X = \{0, 1\}$, $\Sigma_Y = \{0, 1, \perp\}$ and

$$\begin{aligned} p(0|0) &= p(1|1) = 1 - \varepsilon, \\ p(\perp|0) &= p(\perp|1) = \varepsilon. \end{aligned}$$

That is, the output is either equal to the input (it never gets flipped), or a new symbol \perp (‘perp’) that indicates that the bit has been lost. This is illustrated in the following picture:



You could for example use this to describe a situation where you send a (physical or digital) packet from a sender to a receiver which sometimes gets lost.

From these examples we see that the formalism of channels can not only describe arbitrary deterministic functions, but it is also very well suited to describing ‘uncertain’ or ‘noisy’ behavior. To understand how to communicate reliably in the presense of uncertainty and noise is one of the central goals of information theory.

4.4 Quantum channels

We now discuss how the preceding gets modified in *quantum* information theory. As before, we would like to model a ‘box’ – but now the box should map quantum states to quantum states:

$$D(\mathcal{H}_A) \ni \rho_A \longrightarrow \boxed{?} \longrightarrow \rho_B \in D(\mathcal{H}_B)$$

Since quantum states are operators, this should be described by a map

$$\Phi: L(\mathcal{H}_A) \rightarrow L(\mathcal{H}_B).$$

What additional properties should such a map satisfy? First of all, we want to demand that Φ is *linear*. This ensures that if $\{p_i, \rho_i\}$ is an ensemble of input states then⁴

$$T\left[\sum_i p_i \rho_i\right] = \sum_i p_i T[\rho_i].$$

The fact that Φ should be linear can be succinctly written as follows:

$$\Phi \in L(L(\mathcal{H}_A), L(\mathcal{H}_B)). \quad (4.25)$$

Thus, Φ is an operator that maps operators to operators! Such maps are called often called *superoperators*, and we will follow this terminology. We will visualize superoperators by pictures such as the following:

⁴You might wonder why we do not rather model Φ by a map $D(\mathcal{H}_A) \rightarrow D(\mathcal{H}_B)$ that preserves convex combinations. The reason is that any such map has a unique extension to linear map from $L(\mathcal{H}_A)$ to $L(\mathcal{H}_B)$, and linear maps are easier to work with.



As for states and operators, we will often use subscripts to indicate the labels of systems. Thus, we will write $\Phi_{A \rightarrow B}$ for a superoperator as in Eq. (4.25) and

$$M_B = \Phi_{A \rightarrow B}[M_A]$$

to apply a superoperator to some $M_A \in \mathcal{L}(\mathcal{H}_A)$, the result of which is an operator $M_B \in \mathcal{L}(\mathcal{H}_B)$. We will consistently use square brackets [...] to apply superoperators to operators.

We still have to discuss which conditions we should impose to Φ to be a quantum channel, but let us first discuss some generalities.

- First, we always have an *identity* superoperator, denoted

$$\mathcal{I}_A: \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_A), \quad \mathcal{I}_A[M_A] = M_A \quad \forall M_A \in \mathcal{L}(\mathcal{H}_A). \quad (4.26)$$

This naturally describes the situation where our box does not change the input at all – or where there is no box. Accordingly, we will visualize \mathcal{I}_A as follows:



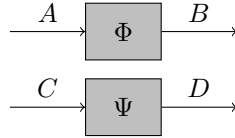
- Second, given two superoperators $\Phi_{A \rightarrow B}$, $\Psi_{C \rightarrow D}$, we can always form their *tensor product*. This is the superoperator

$$\Phi_{A \rightarrow B} \otimes \Psi_{C \rightarrow D} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_C), \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_D)),$$

which is defined as follows on tensor product operators,

$$(\Phi_{A \rightarrow B} \otimes \Psi_{C \rightarrow D})[M_A \otimes N_C] := \Phi_{A \rightarrow B}[M_A] \otimes \Psi_{C \rightarrow D}[N_C], \quad (4.27)$$

extended by linearity – in precise analogy to how we defined the tensor product of operators in terms of the tensor product of vectors (Remark 2.3). The tensor product of two superoperators naturally describes the situation of two boxes where the first is applied to one subsystem and the second to the other, as in the following picture:



What conditions do we want to impose on Φ to legitimately call it a ‘quantum channel’? Clearly, we would like Φ to map quantum states to quantum states:

$$\rho_A \in \mathcal{D}(\mathcal{H}_A) \quad \Rightarrow \quad \Phi_{A \rightarrow B}[\rho_A] \in \mathcal{D}(\mathcal{H}_B)$$

We can equivalently split this up into two conditions and ask that Φ is both

- (a) *Positive*, meaning it maps PSD operators to PSD operators: $\Phi[M_A] \geq 0$ for all $M_A \geq 0$,
- (b) *Trace-preserving*: $\text{Tr}[\Phi[M_A]] = \text{Tr}[M_A]$ for all M_A .

Let us try to come up with maps that satisfy these properties:

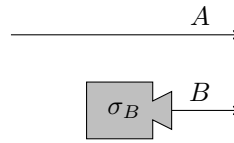
- *Unitaries and isometries:* $\Phi_{A \rightarrow A}[\rho_A] = U\rho_A U^\dagger$ for a fixed unitary $U \in \mathcal{U}(\mathcal{H}_A)$. More generally, we can take $\Phi_{A \rightarrow B}[\rho_A] = V\rho_A V^\dagger$ for an isometry $V \in \mathcal{U}(\mathcal{H}_A, \mathcal{H}_B)$. In our pictures, we will often denote these superoperators simply by “ U ” or “ V ”:

$$\begin{array}{ccc} \xrightarrow{A} & \boxed{U} & \xrightarrow{A} \end{array} \quad \begin{array}{ccc} \xrightarrow{A} & \boxed{V} & \xrightarrow{B} \end{array} \quad (4.28)$$

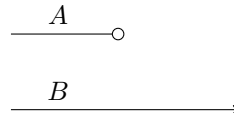
Careful: It is important to keep in mind that these pictures refer not to U and V , but to the corresponding channels Φ .

We already saw examples of unitary superoperators in Lecture 3 when we applied Pauli operators (which are unitaries) in the superdense coding and teleportation protocols. See footnote 3 on p. 40.

- *Add state:* $\Phi_{A \rightarrow AB}[\rho_A] = \rho_A \otimes \sigma_B$ for a fixed state σ_B . This superoperator corresponds to a source that emits an additional quantum system in state σ_B – as in the figure:



- *Partial trace:* $\Phi_{AB \rightarrow A} = \text{Tr}_A$. Indeed, the partial trace is a superoperator that maps states to states, as we discussed above Definition 2.12. This corresponds to the situation where we simply discard a subsystem A:



It is very instructive to note that we can write

$$\text{Tr}_A = \text{Tr} \otimes \mathcal{I}_B,$$

as can be seen by comparing Eq. (4.27) and Lemma 2.9. This shows that our method for drawing pictures makes sense. In fact, it can be developed into a well-defined graphical calculus.

- *Measurement:* We can also represent measurements by superoperators. If $\mu_A: \Omega \rightarrow \text{PSD}(\mathcal{H}_A)$ is an arbitrary measurement on A then we can define

$$\Phi_{A \rightarrow X}[\rho_A] = \sum_{x \in \Omega} \text{Tr}[\mu_A(x)\rho_A] |x\rangle\langle x| \quad \forall \rho_A, \quad (4.29)$$

where X is a new system with Hilbert space $\mathcal{H}_X = \mathbb{C}^\Omega$. By Born’s rule, $\text{Tr}[\mu_A(x)\rho_A]$ is the probability of outcome x using the measurement μ_A . Thus, for any state ρ_A , the output state $\Phi_{A \rightarrow X}[\rho_A]$ is a classical state (Definition 1.10) that describes the probabilities of measurement outcomes. We use the following picture to denote this superoperator:

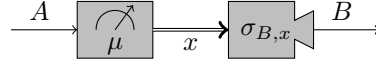


(Recall that, by convention, single lines correspond to quantum systems, while double lines denote classical data.)

- *Measure and prepare*: This superoperator is defined by

$$\Phi_{A \rightarrow B}[\rho_A] = \sum_{x \in \Omega} \text{Tr}[\rho_A \mu_A(x)] \sigma_{B,x},$$

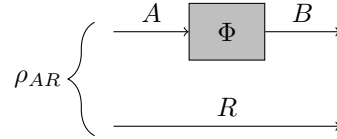
where $\mu_A: \Omega \rightarrow \text{PSD}(\mathcal{H}_A)$ is an arbitrary measurement on A and $\sigma_{B,x}$ a state on B for each possible outcome $x \in \Omega$. This superoperator corresponds to performing a measurement and then preparing a state according to the measurement outcome:



Note that we recover the previous example by taking $\sigma_{B,x} = |x\rangle\langle x|$.

This is encouraging – we found many superoperators that seem reasonable and satisfy conditions (a) and (b), that is, they map states to states.

However, we will now see that there is a *problem* – these two conditions are not sufficient to give a good definition of quantum channels. The reason is that there are superoperators Φ such that the two conditions hold for Φ but fail for $\Phi \otimes \mathcal{I}_R$, i.e., there exists a system R and state ρ_{AR} such that $(\Phi_{A \rightarrow B} \otimes \mathcal{I}_R)[\rho_{AR}]$ is *not* a state!



This is clearly nonsensical, since we want to interpret $\Phi_{A \rightarrow B} \otimes \mathcal{I}_R$ as applying Φ on the A system while leaving the R -system untouched.

For an example of such a superoperator, consider the *transpose map* that sends an operator to its transpose (in some fixed basis):

$$\mathcal{T}[M] = M^T. \quad (4.30)$$

For concreteness, let us consider the qubit case, i.e., $\mathcal{T}: \text{L}(\mathbb{C}^2) \rightarrow \text{L}(\mathbb{C}^2)$.

- It is clear that \mathcal{T} sends states to states, i.e., is positive and trace-preserving. Indeed, the transpose of a PSD operator is PSD, and the trace is likewise invariant under transposition.
- Consider the maximally entangled state of two qubits:

$$\begin{aligned} \rho_{AR} &= |\Phi_{AR}^+\rangle\langle\Phi_{AR}^+| = \frac{1}{2} (|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|) \\ &= \frac{1}{2} (|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|) \\ &= \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

If we apply the transpose channel on the A -subsystem (this is called a *partial transpose*, in analogy to the *partial trace*), we obtain

$$(\mathcal{T} \otimes \mathcal{I}_R)[\rho_{AR}] = \frac{1}{2} \left(|0\rangle\langle 0|^T \otimes |0\rangle\langle 0| + \underbrace{|0\rangle\langle 1|^T}_{\text{}} \otimes |0\rangle\langle 1| + \underbrace{|1\rangle\langle 0|^T}_{\text{}} \otimes |1\rangle\langle 0| + |1\rangle\langle 1|^T \otimes |1\rangle\langle 1| \right)$$

$$\begin{aligned}
&= \frac{1}{2} \left(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \underbrace{|1\rangle\langle 0| \otimes |0\rangle\langle 1|}_{\text{}} + \underbrace{|0\rangle\langle 1| \otimes |1\rangle\langle 0|}_{\text{}} + |1\rangle\langle 1| \otimes |1\rangle\langle 1| \right) \\
&= \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}$$

We see immediately from the matrix representation that ρ_{BR} is not a state. Indeed, while the trace is still one, the right-hand side matrix has an eigenvector $(0, 1, -1, 0)$ with negative eigenvalue $-1/2$. You have already seen this calculation in case you solved Exercise 3.7.

Thus we recognize that ‘positivity’ alone is not enough, we need to demand the stronger condition that even when we tensor with an identity channel we obtain a ‘positive’ map. This property is called ‘complete positivity’. The problem identified above turns out to be the only issue – so we arrive at the following definition of a quantum channel.⁵

Definition 4.15 (Quantum channel). A superoperator $\Phi_{A \rightarrow B} \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$ is called a (*quantum*) *channel* if it is

- (a) *Completely positive*: For all \mathcal{H}_R and $M_{AR} \geq 0$, it holds that $(\Phi_{A \rightarrow B} \otimes \mathcal{I}_R)[M_{AR}] \geq 0$,
- (b) *Trace-preserving*: $\text{Tr}[\Phi_{A \rightarrow B}[M_A]] = \text{Tr}[M_A]$ for all M_A .

We write $\text{CP}(\mathcal{H}_A, \mathcal{H}_B)$ and $\text{C}(\mathcal{H}_A, \mathcal{H}_B)$ for the sets of all completely positive maps $\Phi_{A \rightarrow B}$ and quantum channels, respectively, and we set $\text{CP}(\mathcal{H}_A) := \text{CP}(\mathcal{H}_A, \mathcal{H}_A)$ and $\text{C}(\mathcal{H}_A) := \text{C}(\mathcal{H}_A, \mathcal{H}_A)$.

What are some examples of quantum channels? Clearly, the superoperator \mathcal{I}_A defined in Eq. (4.26) is a quantum channel according to this definition, so we will call it the *identity channel*. In fact, *all* examples given above – except for the transpose map – are quantum channels. You can show this in Exercises 4.10 and 4.11, which also gives some further examples.

We can also build new channels from old ones. For one, the set of quantum channels is a convex set. That is, if $(p_i)_{i \in I}$ is a probability distribution and $\Phi_{A \rightarrow B, i} \in \text{C}(\mathcal{H}_A, \mathcal{H}_B)$ are channels for $i \in I$, then $\sum_{i \in I} p_i \Phi_{A \rightarrow B, i}$ is again a quantum channel. This follows easily from the definition.

Moreover, if $\Phi_{A \rightarrow B}$ is a channel then so is $\Phi_{A \rightarrow B} \otimes \mathcal{I}_R$ for any R . This holds almost by definition, and it implies that channels can be composed in parallel and sequentially (Exercise 4.12):

Lemma 4.16. *If $\Phi_{A \rightarrow B}$ and $\Psi_{B \rightarrow C}$ are channels, then so is $\Psi_{B \rightarrow C} \circ \Phi_{A \rightarrow B}$. If $\Phi_{A \rightarrow B}$ and $\Xi_{C \rightarrow D}$ are channels, then so is $\Phi_{A \rightarrow B} \otimes \Xi_{C \rightarrow D}$.*

Perhaps you still feel a bit uneasy with this definition – could there be another problem that we might have missed in our analysis? Next week we will see that this is not so. Indeed, we will find that any quantum channel according to the above definition can be written as a three-step procedure: first add a system in a fixed state, then apply a unitary, and finally trace over a system. Since quantum physics tells us that these three building blocks are all ‘physical’, this justifies the mathematical definition. See the discussion surrounding Axiom 5.6 for more details.

Remark 4.17 (Complete positivity for classical channels?). In classical information theory the above problem does not appear. Indeed, if $p(y|x)$ is a conditional probability distribution then so is $p(yz'|xz) = p(y|x)\delta_{z,z'}$ (the latter is the same as tensoring the transition matrix with I_Z). Thus, in a classical world, ‘complete positivity’ is automatic.

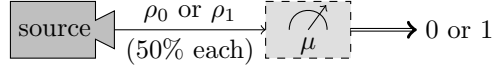
⁵Note that the second condition is unchanged. Indeed, unlike for positivity, it holds automatically that if $\Phi_{A \rightarrow B}$ is trace-preserving then so is $\Phi_{A \rightarrow B} \otimes \mathcal{I}_R$ for any system R .

4.5 Exercises

4.1 **Variational characterization of the trace distance:** Prove Lemma 4.7.

Hint: Consider the spectral decomposition of $\rho - \sigma$.

4.2 **Helstrom's theorem:** Suppose that we have a source that emits states $\rho_0, \rho_1 \in \mathcal{D}(\mathcal{H})$ with 50% probability each. Your goal is to determine a measurement $\mu: \{0, 1\} \rightarrow \text{PSD}(\mathcal{H})$ that identifies the correct state as best as possible, as in the following picture:



By convention, outcome ‘0’ corresponds to state ρ_0 , while outcome ‘1’ corresponds to state ρ_1 . Thus, the probability of success using μ is given by

$$p_{\text{success}} = \frac{1}{2} \text{Tr}[\rho_0 \mu(0)] + \frac{1}{2} \text{Tr}[\rho_1 \mu(1)].$$

Show that *maximal* probability of success over all possible measurements is $\frac{1}{2} + \frac{1}{2} T(\rho_0, \rho_1)$ and can be achieved by a *projective* measurement. This result is known as *Helstrom's theorem*.

4.3 **Practice:** Here you can verify that the measurement from Exercise 1.20 is ‘pretty good’ at distinguishing the states $\rho = |0\rangle\langle 0|$ and $\sigma(t) = (1-t)|0\rangle\langle 0| + t|1\rangle\langle 1|$. Assuming both states are equally likely, plot the following two quantities as functions of $t \in (0, 1]$:

- (a) The optimal probability of distinguishing ρ and $\sigma(t)$ according to Helstrom's theorem.
- (b) The probability of distinguishing ρ and $\sigma(t)$ by using the measurement from Exercise 1.20.

4.4 **Properties of the trace distance:** Verify Lemma 4.8.

4.5 **Trace distance between pure states:** Prove Eq. (4.10). *Hint: $\rho - \sigma$ has rank ≤ 2 .*

4.6 **Properties of the fidelity:** Prove Lemma 4.10.

Hints: Use Uhlmann's theorem for (a), (c), (d). For (c) & (d), construct suitable purifications.

4.7 **Fuchs-van de Graaf:** Use Uhlmann's theorem to prove the upper bound in (4.21).

4.8 **Fidelity inequalities:**

- (a) Show that $|\langle \psi_1 | \phi \rangle|^2 + |\langle \psi_2 | \phi \rangle|^2 \leq 1 + |\langle \psi_1 | \psi_2 \rangle|$ for all vector vectors $|\psi_1\rangle, |\psi_2\rangle, |\phi\rangle \in \mathcal{H}$.
Hint: Upper bound the left-hand side by the largest eigenvalue of some rank-2 matrix.
- (b) Show that $F(\rho_1, \sigma)^2 + F(\rho_2, \sigma)^2 \leq 1 + F(\rho_1, \rho_2)$ for all states $\rho_1, \rho_2, \sigma \in \mathcal{D}(\mathcal{H})$.
- (c) Show the following ‘triangle inequality’: If $F(\alpha, \beta) \geq 1 - \delta$ and $F(\beta, \gamma) \geq 1 - \delta$ for any three states $\alpha, \beta, \gamma \in \mathcal{D}(\mathcal{H})$, then $F(\alpha, \gamma) \geq 1 - 4\delta$.

4.9 **Gentle measurement lemma:** This useful technical result states that if $\rho \in \mathcal{D}(\mathcal{H})$ is a state and $0 \leq Q \leq I$ an operator such that $\text{Tr}[Q\rho] \geq 1 - \varepsilon$, then the following inequalities hold:

$$F\left(\rho, \frac{\sqrt{Q}\rho\sqrt{Q}}{\text{Tr}[Q\rho]}\right) \geq \sqrt{1 - \varepsilon} \quad \text{and} \quad T\left(\rho, \frac{\sqrt{Q}\rho\sqrt{Q}}{\text{Tr}[Q\rho]}\right) \leq \sqrt{\varepsilon} \quad (4.31)$$

- (a) Prove that $\text{Tr} \sqrt{\sqrt{\rho}\sqrt{Q}\rho\sqrt{Q}\sqrt{\rho}} = \text{Tr}[\sqrt{Q}\rho]$ and $\sqrt{Q} \geq Q$.

- (b) Prove the first inequality in Eq. (4.31) using part (a), and deduce the second inequality from the first by using a result from an earlier exercise.
- 4.10 **Quantum channels I:** Show that the following superoperators Φ are channels by directly verifying that they are trace-preserving and completely positive.
- (a) *Isometries:* $\Phi[M] = VMV^\dagger$ for an isometry V .
 - (b) *Add state:* $\Phi[M_A] = M_A \otimes \sigma_B$ for a state σ_B .
 - (c) *Partial trace:* $\Phi[M_{AB}] = \text{Tr}_B[M_{AB}]$.
 - (d) *Classical channel:* $\Phi[M] = \sum_{x,y} p(y|x) \langle x|M|x \rangle |y\rangle\langle y|$, where $p(y|x)$ is a conditional probability distribution (i.e., $p(y|x)$ is a probability distribution in y for each fixed x).
- 4.11 **Quantum channels II:** Show that the following superoperators Φ are channels by directly verifying that they are completely positive and trace-preserving:
- (a) *Mixture of isometries:* $\Phi[M] = \sum_{i=1}^n p_i V_i M V_i^\dagger$, where $(p_i)_{i=1}^n$ is an arbitrary probability distribution and U_1, \dots, U_n arbitrary unitaries.
 - (b) *State replacement:* $\Phi[M] = \text{Tr}[M] \sigma$, where σ is an arbitrary state.
 - (c) *Measure and prepare:* $\Phi[M] = \sum_{x \in \Sigma} \langle x|M|x \rangle \sigma_x$, where $|x\rangle$ denotes the standard basis of \mathbb{C}^Σ and σ_x is an arbitrary state for each $x \in \Sigma$.
- 4.12 **Composing channels:** Prove Lemma 4.16.
- 4.13 **Transpose:** Show that $M_{AB}^{\text{T}_A} = (\mathcal{T}_A \otimes \mathcal{I}_B)[M_{AB}]$ for all $M_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Here, $M_{AB}^{\text{T}_A}$ denotes the partial transpose as defined in Exercise 3.7, and \mathcal{T}_A is the transpose superoperator from Eq. (4.30) (for the same choice of orthonormal basis of \mathcal{H}_A).
- 4.14 **No cloning:** We say that a channel $\Phi \in \mathcal{C}(\mathcal{H}, \mathcal{H} \otimes \mathcal{H})$ *clones* a state $\rho \in \mathcal{D}(\mathcal{H})$ if $\Phi[\rho] = \rho \otimes \rho$. For simplicity, let $\mathcal{H} = \mathbb{C}^2$ be a qubit.
- (a) Show that there exists no channel that clones all *classical* states ρ .
 - (b) Show that there exists no channel that clones all *pure* states ρ .
 - (c) Which states are both pure *and* classical? Find a channel Φ that clones all of them.

Hint: For (a) and (b), use that channels are linear to arrive at a contradiction.

Lecture 5

Structure of quantum channels

Last week, we defined the notion of a *quantum channel* as a completely positive and trace-preserving superoperator (see Definition 4.15). Today we will discuss several characterizations of quantum channels. Those characterizations will give us better mathematical insight into the notion of complete positivity, serve as important tools for what follows, and give us a more satisfying explanation why last week's definition is a sensible one.

5.1 Superoperators and complete positivity

Let $\Phi_{A \rightarrow B} \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$ be a superoperator. It is easy to check when Φ is trace-preserving, but how can we check complete positivity?

We start with a warning. Since $L(\mathcal{H}_A) \cong \mathcal{H}_A \otimes \mathcal{H}_A^*$, we can always think of $\Phi_{A \rightarrow B}$ as an operator in $L(\mathcal{H}_A \otimes \mathcal{H}_A^*, \mathcal{H}_B \otimes \mathcal{H}_B^*)$. Now, despite the similarity of words, it is important to keep in mind that ‘positivity’ or ‘complete positivity’ of Φ does *not* mean that $\Phi_{A \rightarrow B}$ is a PSD operator. Indeed, the latter statement does not even make sense in general, since $\mathcal{H}_A \otimes \mathcal{H}_A^*$ and $\mathcal{H}_B \otimes \mathcal{H}_B^*$ are not necessarily even the same spaces. Instead, our main tool will be to associate with every superoperator an operator in $L(\mathcal{H}_A \otimes \mathcal{H}_B)$ – such operators have the possibility of being PSD, and we will see that this precisely characterizes when Φ is completely positive. We use the following definition:

Definition 5.1 (Choi operator). We define the *Choi operator* associated with a superoperator $\Phi_{A \rightarrow B}$ as

$$J_{AB}^\Phi := \sum_{x,y} |x\rangle\langle y| \otimes \Phi_{A \rightarrow B}[|x\rangle\langle y|] \in L(\mathcal{H}_A \otimes \mathcal{H}_B), \quad (5.1)$$

where $|x\rangle$ denotes an arbitrary orthonormal basis of \mathcal{H}_A .

You can think of J_{AB}^Φ as a block matrix where the block at coordinates (x, y) contains the output $\Phi_{A \rightarrow B}[|x\rangle\langle y|]$ of the channel on input $|x\rangle\langle y|$. This is indeed a complete description of Φ since by linearity you can recover the output of Φ on any input.

We can also write

$$J_{AB}^\Phi = \sum_{x,y} (\mathcal{I}_A \otimes \Phi_{A \rightarrow B})[|x, x\rangle\langle y, y|], \quad (5.2)$$

which makes it clear that J_{AB}^Φ is the result of applying $\Phi_{A \rightarrow B}$ to half of an *unnormalized* maximally entangled state $\sum_x |x, x\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A$, compare Definition 3.5. Note that the latter

state depends on a choice of basis of \mathcal{H}_A , just like the Choi operator. The following figure illustrates Eq. (5.2):

$$\sum_x |xx\rangle \left\{ \begin{array}{l} \xrightarrow{A} \\ \xrightarrow{A} \boxed{\Phi} \xrightarrow{B} \end{array} \right. = J_{AB}^\Phi$$

For example, taking $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^\Sigma$ and the standard basis, the so-called *completely dephasing channel*

$$\Delta[\rho] = \sum_{x \in \Sigma} \langle x | \rho | x \rangle |x\rangle \langle x| \quad (5.3)$$

has the following Choi operator:

$$J_{AB}^\Delta = \sum_x |x\rangle \langle x| \otimes |x\rangle \langle x|, \quad (5.4)$$

an unnormalized maximally correlated state. You can verify this and more in Exercise 5.1.

In fact, the mapping $\Phi \mapsto J^\Phi$ defines an isomorphism, known as the *Choi-Jamiołkowski isomorphism*:

Lemma 5.2 (Choi-Jamiołkowski isomorphism). *The following map is an isomorphism,*

$$\mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B)) \rightarrow \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B), \quad \Phi_{A \rightarrow B} \mapsto J_{AB}^\Phi,$$

with inverse given by

$$\Phi_{A \rightarrow B}[M_A] = \text{Tr}_A[(M_A^\top \otimes I_B) J_{AB}^\Phi] \quad \forall M_A \in \mathcal{L}(\mathcal{H}_A), \quad (5.5)$$

where we take the transpose in the same basis as used to in the definition of the Choi operator.

Proof. The mapping is clearly linear and both spaces have the same dimension, so we only need to show how the channel can be recovered from the Choi operator. For this we prove Eq. (5.5) by a direct calculation:

$$\begin{aligned} \text{Tr}_A[(M_A^\top \otimes I_B) J_{AB}^\Phi] &= \sum_{x,y} \text{Tr}_A[(M_A^\top \otimes I_B)(|x\rangle \langle y| \otimes \Phi_{A \rightarrow B}[|x\rangle \langle y|])] \\ &= \sum_{x,y} \text{Tr}_A[M_A^\top |x\rangle \langle y| \otimes \Phi_{A \rightarrow B}[|x\rangle \langle y|]] \\ &= \sum_{x,y} \underbrace{\text{Tr}[M_A^\top |x\rangle \langle y|]}_{=\langle y|M_A^\top|x\rangle=\langle x|M_A|y\rangle} \Phi_{A \rightarrow B}[|x\rangle \langle y|] \\ &= \sum_{x,y} \Phi_{A \rightarrow B}[|x\rangle \langle x| M_A |y\rangle \langle y|] = \Phi_{A \rightarrow B}[M_A]. \end{aligned}$$

□

It is a nice exercise to verify that this formula indeed recovers Eq. (5.3) from Eq. (5.4).

We now state the central theorem that gives four equivalent ways of characterizing when a superoperator is completely positive.

Theorem 5.3 (When is a superoperator completely positive?). *For a superoperator $\Phi_{A \rightarrow B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$, the following statements are equivalent:*

- (a) $\Phi_{A \rightarrow B}$ is completely positive (i.e., for all \mathcal{H}_R and $M_{AR} \geq 0$ it holds that $(\Phi_{A \rightarrow B} \otimes \mathcal{I}_R)[M_{AR}] \geq 0$).
- (b) $\Phi_{A \rightarrow B} \otimes \mathcal{I}_{A'}$ is positive where $\mathcal{H}_{A'} \cong \mathcal{H}_A$ (i.e., for all $M_{AA'} \geq 0$ it holds that $(\Phi_{A \rightarrow B} \otimes \mathcal{I}_{A'})[M_{AA'}] \geq 0$).
- (c) $J_{AB}^\Phi \geq 0$, i.e., the Choi operator of $\Phi_{A \rightarrow B}$ is positive semidefinite.
- (d) Kraus representation: There exist operators $X_1, \dots, X_r \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ such that

$$\Phi[M] = \sum_{i=1}^r X_i M X_i^\dagger \quad (5.6)$$

for all $M \in \mathcal{L}(\mathcal{H}_A)$.

- (e) Stinespring representation: There exists \mathcal{H}_E and $V \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_E)$ such that

$$\Phi[M] = \text{Tr}_E[V M V^\dagger] \quad (5.7)$$

for all $M \in \mathcal{L}(\mathcal{H}_A)$.

Moreover, r in (d) and $\dim \mathcal{H}_E$ in (e) can be chosen as $\text{rank}(J_{AB}^\Phi) \leq \dim \mathcal{H}_A \dim \mathcal{H}_B$ (or larger).

Proof. The implications (a) \Rightarrow (b) \Rightarrow (c) are immediate. For the implication (d) \Rightarrow (e), simply define $\mathcal{H}_E = \mathbb{C}^r$ and $V := \sum_{i=1}^r X_i \otimes |i\rangle$ and verify that Eq. (5.7) reduces to Eq. (5.6). (We can also go the other way around and obtain Kraus operators from V by setting $X_i := (I_B \otimes \langle i|)V$, showing that (e) \Rightarrow (d).) The implication (e) \Rightarrow (a) is also easy – both $M \mapsto V M V^\dagger$ and Tr_E are completely positive (see Exercise 4.10, complete positivity of part (a) did not rely on the fact that U was unitary), hence so is their composition.

It remains to prove that (c) \Rightarrow (d) with $r = \text{rank } J_{AB}^\Phi$. Since J_{AB}^Φ is PSD, we can use a spectral decomposition (see Theorem 1.3) to write¹

$$J_{AB}^\Phi = \sum_{i=1}^r |v_i\rangle\langle v_i| \quad (5.8)$$

for suitable vectors $|v_i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ that need not be normalized. We can construct operators $X_i \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$ from them similarly as in Exercise 2.12. Simply define

$$X_i := \sum_{a,b} \langle a, b | v_i \rangle |b\rangle\langle a| \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_B). \quad (5.9)$$

Then, using Eq. (5.5),

$$\begin{aligned} \Phi[M] &= \text{Tr}_A[(M_A^\top \otimes I_B) J_{AB}^\Phi] = \sum_i \text{Tr}_A[(M_A^\top \otimes I_B) |v_i\rangle\langle v_i|] \\ &= \sum_i \sum_{a,b} \sum_{a',b'} \langle a, b | v_i \rangle \langle v_i | a', b' \rangle \text{Tr}_A[(M_A^\top \otimes I_B) |a, b\rangle\langle a', b'|] \\ &= \sum_i \sum_{a,b} \sum_{a',b'} \langle a, b | v_i \rangle \langle v_i | a', b' \rangle \text{Tr}_A[M_A^\top |a\rangle\langle a'| \otimes |b\rangle\langle b'|] \end{aligned}$$

¹To get this form, restrict the decomposition of J_{AB}^Φ to the positive eigenvalues $\lambda_i > 0$ and absorb their square root $\sqrt{\lambda_i}$ into the normalization of the eigenvectors $|v_i\rangle$.

$$\begin{aligned}
&= \sum_i \sum_{a,b} \sum_{a',b'} \langle a, b | v_i \rangle \langle v_i | a', b' \rangle \underbrace{\text{Tr}[M_A^\top | a \rangle \langle a' |]}_{=\langle a' | M_A^\top | a \rangle = \langle a | M_A | a' \rangle} | b \rangle \langle b' | \\
&= \sum_i \sum_{a,b} \sum_{a',b'} \langle a, b | v_i \rangle | b \rangle \langle a | M_A | a' \rangle \langle b' | \langle v_i | a', b' \rangle \\
&= \sum_i X_i M_A X_i^\dagger,
\end{aligned}$$

which concludes the proof. \square

Theorem 5.3 is rather remarkable. Criterion (b) shows that complete positivity, which a priori involves an auxiliary Hilbert space \mathcal{H}_R of unbounded dimension, to a single $\mathcal{H}_R \cong \mathcal{H}_A$. And criterion (c) shows that we do not even have to check that $\Phi_{A \rightarrow B} \otimes \mathcal{I}_{A'}$ sends every PSD operator to a PSD operator – it suffices to check this condition just for an (unnormalized, if we wish) maximally entangled state. You can practice this technique in Exercise 5.10.

Criteria (d) and (e) are also very useful in practice, since many quantum channels are naturally given in this form. Indeed, Exercises 4.10 and 4.11 simplify tremendously using Theorem 5.3!

Remark 5.4 (Beyond completely positive maps). For superoperators that are not completely positive, we can still find weak forms of Kraus and Stinespring representations. Namely, any superoperator can be written in the form $\Phi[M] = \sum_i X_i M Y_i^\dagger$ (where, in general, $X_i \neq Y_i$) or $\Phi[M] = V M W^\dagger$ (where, in general, $V \neq W$). This can be proved as above using the singular value decomposition of the Choi operator (which need no longer be PSD) instead of the eigendecomposition in Eq. (5.8). As these representations are much less useful we did not discuss this in class.

5.2 Characterizing quantum channels

With Theorem 5.3 in hand, it is straightforward to characterize quantum channels since we only need to determine when a completely positive map is trace-preserving. This is achieved by the following lemma.

Lemma 5.5 (When is a completely positive superoperator trace-preserving?). *For a completely positive superoperator $\Phi_{A \rightarrow B}$, the following statements are equivalent:*

- (a) $\Phi_{A \rightarrow B}$ is trace-preserving (hence a quantum channel).
- (b) Choi operator: $\text{Tr}_B[J_{AB}^\Phi] = I_A$.
- (c) Kraus representation: $\sum_i X_i^\dagger X_i = I_A$ for one/every Kraus representation.
- (d) Stinespring representation: $V^\dagger V = I_A$ for one/every Stinespring representation. That is, V is an isometry.

In fact, the equivalence between (a) and (b) holds for arbitrary superoperators (completely positive or not).

Proof. We will use the fact, which follows from Exercise 1.18, that for $X, Y \in \mathcal{L}(\mathcal{H})$ it holds that

$$\text{Tr}(XM) = \text{Tr}(YM) \text{ for all } M \in \mathcal{L}(\mathcal{H}) \Leftrightarrow X = Y. \quad (5.10)$$

By Eq. (5.5) for any $M_A \in \mathcal{L}(\mathcal{H}_A)$

$$\text{Tr}[\Phi_{A \rightarrow B}(M_A)] = \text{Tr}[\text{Tr}_A[(M_A^\top \otimes I_B) J_{AB}^\Phi]] = \text{Tr}[M_A^\top \text{Tr}_B[J_{AB}^\Phi]]$$

from which it clearly follows using Eq. (5.10) that (b) and (a) are equivalent since $\text{Tr}(M_A^\top) = \text{Tr}(M_A)$. Next, consider a Kraus representation of the channel, and use the cyclicity of the trace to see that for all $M_A \in \mathcal{L}(\mathcal{H}_A)$

$$\text{Tr}[\Phi_{A \rightarrow B}(M_A)] = \text{Tr}\left[\sum_i X_i M_A X_i^\dagger\right] = \text{Tr}\left[M_A \left(\sum_i X_i^\dagger X_i\right)\right].$$

So, again using Eq. (5.10), we see that (a) and (c) are equivalent. Finally, for a Stinespring representation, again using the cyclicity of the trace we see that

$$\text{Tr}[\Phi_{A \rightarrow B}(M_A)] = \text{Tr}[\text{Tr}_E[V M_A V^\dagger]] = \text{Tr}[M_A V^\dagger V]$$

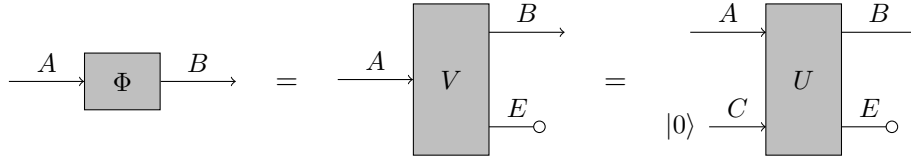
allowing us to conclude the equivalence of (a) and (d). \square

It is worth stating again that if we put half of a normalized maximally entangled state into a channel then we get a quantum state, which is nothing but the Choi operator normalized to be a quantum state. This state is also known as the *Choi state* of $\Phi_{A \rightarrow B}$, and it is given by

$$\frac{1}{d_A} J_{AB}^\Phi = (\mathcal{I}_A \otimes \Phi_{A \rightarrow B})[|\Phi_{AA}^+\rangle\langle\Phi_{AA}^+|], \quad (5.11)$$

where $|\Phi_{AA}^+\rangle = \frac{1}{\sqrt{d_A}} \sum_x |x, x\rangle$ is a maximally entangled state and $d_A = \dim \mathcal{H}_A$.

The Stinespring representation has a nice conceptual interpretation. It is by definition a composition of applying an isometry and then forgetting a subsystem (partial trace). In Exercise 5.5 you will show that you can reinterpret the isometry as a composition of first adding a pure state in a reference system and then applying a *unitary*:²



That means that every quantum channel can be constructed as a composition of adding states, applying unitaries and discarding subsystems, which shows that the formalism of quantum channels is equivalent to unitary quantum mechanics on pure states where we may add and forget subsystems, which is not at all clear from the original definition of a quantum channel. Thus we may feel sufficiently confident to state the following axiom:

Axiom 5.6 (Channels). Any quantum channel $\Phi_{A \rightarrow B}$ can be realized physically. That is, in principle, there exists a device that, given as input an arbitrary state ρ_A , outputs the state $\Phi_{A \rightarrow B}[\rho_A]$.

Apart from being conceptually insightful, the Stinespring representation also often simplifies proofs tremendously. Indeed, to show a certain property holds for quantum channels it suffices to show that it holds for isometries (which is often trivial) and for partial traces. In Exercise 5.8 you will use this proof strategy to show the following lemma:

²At first glance this “equation” seems to have a syntactical error: The left-hand side is a quantum channel (a superoperator), while the middle and right-hand side are an isometry and unitary, respectively. However, we can identify the latter with quantum channels, as explained in Eq. (4.28).

Lemma 5.7 (Monotonicity of distance measures). *For all states ρ_A, σ_A and channels $\Phi_{A \rightarrow B}$,*

$$T(\Phi_{A \rightarrow B}[\rho_A], \Phi_{A \rightarrow B}[\sigma_A]) \leq T(\rho_A, \sigma_A) \quad \text{and} \quad F(\Phi_{A \rightarrow B}[\rho_A], \Phi_{A \rightarrow B}[\sigma_A]) \geq F(\rho_A, \sigma_A).$$

Remark 5.8 (Uniqueness of the Stinespring and Kraus representations). It is interesting to ask how much freedom we have in choosing the Stinespring and Kraus representations. Any two Stinespring isometries $V_{A \rightarrow BE}, \tilde{V}_{A \rightarrow BE}$ of a channel $\Phi_{A \rightarrow B}$ are related by a unitary U_E on the system E that is discarded, in the sense that

$$\tilde{V}_{A \rightarrow BE} = (I_B \otimes U_E)V_{A \rightarrow BE}. \quad (5.12)$$

This follows from Lemma 2.18, because $|\Phi_{ABE}\rangle := (I_A \otimes V_{A \rightarrow BE})|\Phi_{AA}^+\rangle$ and $|\tilde{\Phi}_{ABE}\rangle := (I_A \otimes \tilde{V}_{A \rightarrow BE})|\Phi_{AA}^+\rangle$ are both purifications of the Choi state in Eq. (5.11) and hence they are related by a unitary U_E on E , so $|\tilde{\Phi}_{ABE}\rangle = (I_{AB} \otimes U_E)|\Phi_{ABE}\rangle$. It is an exercise for the reader to check that this indeed implies Eq. (5.12).

As a consequence, any two sets $\{X_i\}_{i=1}^r, \{Y_i\}_{i=1}^r$ of Kraus operators for a channel $\Phi_{A \rightarrow B}$ are related by a unitary matrix $U \in U(\mathbb{C}^r)$ in the sense that $X_i = \sum_j U_{ij} Y_j$ for $i = 1, \dots, r$. This can be seen by constructing the Stinespring isometries corresponding to these Kraus operators as in the proof of Theorem 5.3.

One can also compare Stinespring isometries with different auxilliary systems or sets of Kraus operators of different cardinalities, in which case the unitary on the reference system is replaced by an isometry.

Recall from Eq. (4.4) that the Hilbert-Schmidt inner product on $L(\mathcal{H})$ is given by $\langle M, N \rangle = \text{Tr}[M^\dagger N]$. This allows us to define the *adjoint* of a superoperator:

Definition 5.9 (Adjoint superoperator). The *adjoint* $\Phi^\dagger \in L(L(\mathcal{H}_B), L(\mathcal{H}_A))$ of a superoperator $\Phi \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$ is defined such that

$$\langle M_A, \Phi^\dagger[N_B] \rangle = \langle \Phi[M_A], N_B \rangle \quad \forall M_A \in L(\mathcal{H}_A), N_B \in L(\mathcal{H}_B).$$

Like for any adjoint, it holds that $(\Phi^\dagger)^\dagger = \Phi$. The following lemma summarizes some properties of the adjoint. You will prove it in Exercise 5.7.

Lemma 5.10 (Properties of the adjoint). *Let $\Phi \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$ be a superoperator.*

- (a) Φ is completely positive if and only if Φ^\dagger is completely positive.
- (b) Φ is trace-preserving if and only if Φ^\dagger is unital (meaning that $\Phi^\dagger[I_B] = I_A$).

5.3 Exercises

5.1 Depolarizing and dephasing channels: The *completely depolarizing channel* on $L(\mathcal{H})$ is given by

$$\mathcal{D}[M] = \text{Tr}[M] \frac{I}{d} \quad \forall M \in L(\mathcal{H}),$$

where $d = \dim \mathcal{H}$. For $\mathcal{H} = \mathbb{C}^\Sigma$, the *completely dephasing channel* is defined by

$$\Delta[M] = \sum_x \langle x|M|x \rangle |x\rangle\langle x| \quad \forall M \in L(\mathcal{H}).$$

- (a) Compute the Choi operator of either channel.
- (b) What is the result of acting by either channel on half of a maximally entangled state?
- (c) For qubits, $\mathcal{H} = \mathbb{C}^2$, how does either channel act on Bloch vectors?

5.2 Kraus and Stinespring: Find Kraus and Stinespring representations for the following quantum channels:

- (a) *Trace:* $\Phi[M] = \text{Tr}[M]$
- (b) *Add pure state:* $\Phi[M_A] = M_A \otimes |\phi\rangle\langle\phi|_B$ for a unit vector $|\phi_B\rangle \in \mathcal{H}_B$.
- (c) *Completely dephasing channel:* $\Delta[M] = \sum_x \langle x|M|x\rangle |x\rangle\langle x|$ (same as above).

5.3 Kraus and Stinespring composition: Given Kraus or Stinespring representations of two channels $\Phi_{A \rightarrow B}$ and $\Psi_{B \rightarrow C}$, explain how to obtain the same representation for $\Psi_{B \rightarrow C} \circ \Phi_{A \rightarrow B}$.

5.4 More channel representations:

- (a) Find a Kraus representation of the following channel, which sends a single qubit to two qubits: $\Phi[\omega] = \frac{1}{2}\omega \otimes |0\rangle\langle 0| + \frac{1}{2}\text{Tr}[\omega] |0\rangle\langle 0| \otimes |1\rangle\langle 1|$.
- (b) Compute the Choi operator of the single-qubit channel with the following two Kraus operators:

$$\sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \sqrt{1-p} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

5.5 Stinespring with unitaries: Use the Stinespring representation to prove that any quantum channel $\Phi_{A \rightarrow B}$ can be written in the following form:

$$\Phi_{A \rightarrow B}[M_A] = \text{Tr}_E[U_{AC \rightarrow BE}(M_A \otimes \sigma_C)U_{AC \rightarrow BE}^\dagger] \quad \forall M_A,$$

where $\mathcal{H}_C, \mathcal{H}_E$ are auxiliary Hilbert spaces, $\sigma_C \in \mathcal{D}(\mathcal{H}_C)$ is a *pure* state, and $U_{AC \rightarrow BE}$ a *unitary*.

5.6 Entry-wise channels: Let $\mathcal{H} = \mathbb{C}^d$ and fix an operator $A \in \mathcal{L}(\mathcal{H})$. Consider the superoperator $\Phi: \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{H})$ defined as $\Phi[X] = A \odot X$ for all $X \in \mathcal{L}(\mathcal{H})$, where $A \odot X$ denotes the entry-wise product of A and X : $(A \odot X)_{ij} = A_{ij}X_{ij}$ for all $i, j \in \{0, \dots, d-1\}$.

- (a) Compute $\Phi[X]$ when A is the identity operator. Under what name did you encounter this channel before?
- (b) Show that if A is positive semidefinite then Φ is completely positive.
Hint: One possibility is by computing the Choi matrix of Φ .
- (c) Show that if Φ is completely positive then A is positive semidefinite.
- (d) Under what condition on A is Φ trace-preserving? Show that Φ is trace-preserving if and only if your condition holds.

5.7 Adjoint superoperator:

- (a) Given a Kraus representation of a completely positive map Φ , explain how to find a Kraus representation for Φ^\dagger .
- (b) Prove part (a) of Lemma 5.10.
- (c) Prove part (b) of Lemma 5.10.

5.8 Monotonicity of distance measures: Prove Lemma 5.7.

- 5.9 **Fidelity and composition of channels:** Let $\Phi_{A \rightarrow B}, \Psi_{B \rightarrow C}$ be channels and let ρ_A, ρ_B, ρ_C be states. Show that if $F(\Phi_{A \rightarrow B}[\rho_A], \rho_B) \geq 1 - \delta$ and $F(\Psi_{B \rightarrow C}[\rho_B], \rho_C) \geq 1 - \delta$ for some $\delta > 0$ then $F((\Psi_{B \rightarrow C} \circ \Phi_{A \rightarrow B})[\rho_A], \rho_C) \geq 1 - 4\delta$.

Hint: Exercise 4.8.

- 5.10 **Depolarizing channel:** Consider the following trace-preserving superoperator on $L(\mathcal{H})$, where $\dim \mathcal{H} = d$ and $\lambda \in \mathbb{R}$ is a parameter:

$$\mathcal{D}_\lambda[M] = \lambda M + (1 - \lambda) \text{Tr}[M] \frac{I}{d}$$

- (a) Compute the Choi operator of \mathcal{D}_λ for any value of λ .
 - (b) For which values of λ is \mathcal{D}_λ a quantum channel?
- 5.11 **Kraus and Stinespring:** Find Kraus and Stinespring representations for the following quantum channels:
- (a) *Partial trace:* $\Phi[M_{AE}] = \text{Tr}_E[M_{AE}]$
 - (b) *Add state:* $\Phi[M_A] = M_A \otimes \sigma_B$ for a state σ_B .
 - (c) *Measure and prepare:* $\Phi[M] = \sum_{x \in \Sigma} \langle x | M | x \rangle \sigma_{B,x}$, where $|x\rangle$ denotes the standard basis of $\mathcal{H}_A = \mathbb{C}^\Sigma$ and $\sigma_{B,x}$ is an arbitrary state for each $x \in \Sigma$.
- 5.12 **Classical-quantum states and quantum-to-classical channels:** Let $\mathcal{H}_X = \mathbb{C}^\Sigma$. We say that a state ρ_{XB} is *classical on subsystem X*, or that it is a *classical-quantum state on XB*, if it can be written in the form

$$\rho_{XB} = \sum_{x \in \Sigma} p(x) |x\rangle\langle x| \otimes \rho_{B,x}$$

for a probability distribution p on Σ and states $\rho_{B,x}$ on \mathcal{H}_B . By convention, we will always denote subsystems by X, Y, \dots if we know them to be classical (and A, B, \dots otherwise).

- (a) Discuss how this generalizes the notion of classical states.
- (b) Show that the fidelity between two classical-quantum states $\rho_{XB} = \sum_{x \in \Sigma} p(x) |x\rangle\langle x| \otimes \rho_{B,x}$ and $\sigma_{XB} = \sum_{x \in \Sigma} q(x) |x\rangle\langle x| \otimes \sigma_{B,x}$ is

$$F(\rho_{XB}, \sigma_{XB}) = \sum_{x \in \Sigma} \sqrt{p(x)q(x)} F(\rho_{B,x}, \sigma_{B,x}).$$

- (c) Show that ρ_{XB} is classical on subsystem X if and only if $(\Delta_X \otimes \mathcal{I}_B)[\rho_{XB}] = \rho_{XB}$, where $\Delta_X[M] = \sum_{x \in \Sigma} |x\rangle\langle x| M |x\rangle\langle x|$ is the completely dephasing channel on the X -system.
- (d) Assume that $\Phi_{A \rightarrow X}$ is a *quantum-to-classical channel*, i.e. $\Phi_{A \rightarrow X}[\rho_A]$ is classical for every state ρ_A . Show that there exists a measurement $\mu_A: \Sigma \rightarrow \text{PSD}(\mathcal{H}_A)$ such that

$$\Phi_{A \rightarrow X}[\rho_A] = \sum_{x \in \Sigma} \text{Tr}[\mu_A(x) \rho_A] |x\rangle\langle x| \quad \forall \rho_A.$$

Hint: Use Exercise 1.18.

- (e) Let $\Phi_{A \rightarrow X}$ be the channel corresponding to a measurement μ_A , as in Exercise 5.12 (d). Show that, for any system B and any state ρ_{AB} , the state $\rho_{XB} = (\Phi_{A \rightarrow X} \otimes \mathcal{I}_B)[\rho_{AB}]$ is a classical-quantum state and compute the probabilities $p(x)$ and the states $\rho_{B,x}$. Compare your result with Axiom 2.15.


5.13 **Complementary channels:** Recall that any channel $\Phi_{A \rightarrow B} \in C(\mathcal{H}_A, \mathcal{H}_B)$ has a Stinespring representation

$$\Phi_{A \rightarrow B}[M] = \text{Tr}_E[VMV^\dagger] \quad \forall M \in L(\mathcal{H}_A),$$

where $V: \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$ is an isometry and E is an auxiliary system that is discarded after applying the isometry. If we instead discard the original output system B , we obtain the following channel $\Phi_{A \rightarrow E}^c \in C(\mathcal{H}_A, \mathcal{H}_E)$, which is called a *complementary channel* of $\Phi_{A \rightarrow B}$:

$$\Phi_{A \rightarrow E}^c[M] = \text{Tr}_B[VMV^\dagger] \quad \forall M \in L(\mathcal{H}_A).$$

- (a) Find a Stinespring representation and complementary channel of the completely dephasing channel $\Phi_{A \rightarrow B}[M] = \sum_{i=1}^d \langle i|M|i \rangle |i\rangle\langle i|$, where $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^d$.
- (b) Find a Stinespring representation and complementary channel of the channel $\Phi_{A \rightarrow B}[M] = \text{Tr}[M]\tau_B$, which prepares the maximally mixed state $\tau_B \in D(\mathcal{H}_B)$.
- (c) We say that a channel $\Phi_{A \rightarrow B}$ is *degradable* if there is another channel $\Psi_{E \rightarrow B}$ such that $\Phi_{A \rightarrow B} = \Psi_{E \rightarrow B} \circ \Phi_{A \rightarrow E}^c$. Which of the two channels discussed in (a) and (b) is degradable? Justify your answer.
- (d) So far we considered specific channels. Suppose now you are given a general channel $\Phi_{A \rightarrow B}$ with Kraus representation $\Phi_{A \rightarrow B}[M] = \sum_{i=1}^r X_i M X_i^\dagger$. Find a Stinespring representation of $\Phi_{A \rightarrow B}$ and a Kraus representation of the complementary channel $\Phi_{A \rightarrow E}^c$.

5.14  **Practice:** In the files 05-choi-matrix-1.txt, 05-choi-matrix-2.txt you will find two Choi matrices, which represent superoperators from 5 qubits to 1 qubit.

- (a) Apply these superoperators to the input state $|00010\rangle$.
- (b) Compute the fidelity between these two outputs of the two channels.
- (c) Do these Choi matrices describe valid quantum channels?

Lecture 6

Shannon entropy and data compression

Over the past month, we have learned the basic formalism and toolbox of quantum information theory (e.g., note that all objects in the cartoon on p. 7 are now well-defined). From this week on we will discuss information theory proper. Today we will discuss the classical theory of data compression due to Shannon. Next week, we will generalize Shannon's results and learn how to optimally compress quantum information. For more information on classical information theory see, e.g., the lecture notes at <https://staff.fnwi.uva.nl/m.walter/iit19/>.

6.1 Shannon entropy

Today we will work with classical probability distributions a lot. Recall that

$$\mathcal{P}(\Sigma) = \left\{ p: \Sigma \rightarrow \mathbb{R}_{\geq 0} : \sum_{x \in \Sigma} p(x) = 1 \right\}$$

denotes the set of all probability distributions on a finite set Σ . If X is a random variable then write $X \sim p$ to say that X is distributed according to p , i.e., $\Pr(X = x) = p(x)$ for all $x \in \Sigma$. As usual, we write $E[X] = \sum_{x \in \Sigma} p(x)x$ for the *expectation value* and $\text{Var}(X) = E[X^2] - E[X]^2$ for the *variance* of a numerical random variable X . We now define the Shannon entropy.

Definition 6.1 (Shannon entropy). The *Shannon entropy* of a probability distribution $p \in \mathcal{P}(\Sigma)$ is defined by

$$H(p) := \sum_{x \in \Sigma} p(x) \log \frac{1}{p(x)} = - \sum_{x \in \Sigma} p(x) \log p(x). \quad (6.1)$$

Throughout these lecture notes, \log always denotes the logarithm to *base 2* (i.e., $\log 2 = 1$).

As stated, Eq. (6.1) is only well-defined if all $p(x) > 0$. However, note that $q \log \frac{1}{q} = -q \log q$ is continuous in $q > 0$ and tends to 0 as $q \rightarrow 0$, as illustrated in Fig. 6.1 (a). We can thus extend the definition of $H(p)$ by continuity, i.e., defining

$$p(x) \log \frac{1}{p(x)} = -p(x) \log p(x) = 0 \quad \text{for} \quad p(x) = 0$$

in Eq. (6.1). Then $H(p)$ is a *continuous* function of $p \in \mathcal{P}(\Sigma)$. This definition is also compatible with the interpretation that the Shannon entropy can be written as

$$H(p) = E \left[\log \frac{1}{p(X)} \right] = -E[\log p(X)], \quad (6.2)$$

where $X \sim p$, since probability-zero outcomes do not impact the expectation value.

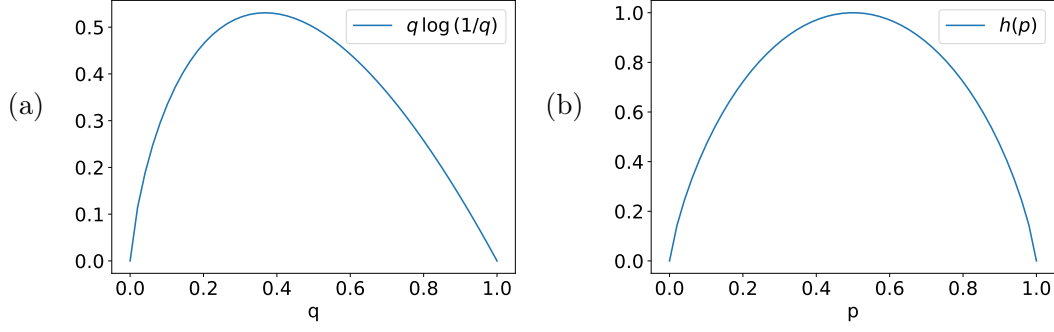


Figure 6.1: (a) The function $f(q) = q \log \frac{1}{q}$. As is apparent, $f(q) \rightarrow 0$ as $q \rightarrow 0$. (b) The binary entropy function $h(p) = H(\{p, 1-p\})$ defined as in Eq. (6.3).

Example 6.2 (Binary entropy function). The Shannon entropy of a probability distribution with two possible outcomes is given by the so-called *binary entropy function*,

$$h(p) := H(\{p, 1-p\}) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}, \quad (6.3)$$

where p is the probability of one of the outcomes. This function is visualized in Fig. 6.1 (b). Note that it is continuous, but not Lipschitz continuous (you can prove this in Exercise 6.7).

We will now list some further properties of the Shannon entropy. Before we state these properties, recall that a function $f: \mathbb{D} \rightarrow \mathbb{R}$ defined on a convex set $\mathbb{D} \subseteq \mathbb{R}^n$ (e.g., an interval if $n = 1$) is called *concave* if it holds that $qf(a) + (1-q)f(b) \leq f(qa + (1-q)b)$ for any $q \in [0, 1]$ and $a, b \in \mathbb{D}$. It is called *strictly concave* if equality only holds for $a = b$ or $q \in \{0, 1\}$. If \mathbb{D} is an interval and f is twice differentiable on its interior with $f'' \leq 0$ then f is concave. If $f'' < 0$ then f is strictly concave.

Jensen's inequality states that, for any concave function f as above,

$$\sum_{x \in \Sigma} p(x) f(a(x)) \leq f\left(\sum_{x \in \Sigma} p(x) a(x)\right) \quad (6.4)$$

for any probability distribution $p \in \mathcal{P}(\Sigma)$ and function $a: \Sigma \rightarrow \mathbb{D}$. (If $|\Sigma| = 2$ then this simply restates the definition of concavity.) Moreover, if f is strictly concave then equality in Eq. (6.4) holds if and only if a is constant on the set $\{x \in \Sigma : p(x) > 0\}$. We can also state Eq. (6.4) in probabilistic terms. If f is a concave function on \mathbb{D} and A a random variable on \mathbb{D} then

$$E[f(A)] \leq f(E[A]),$$

and for a strictly concave function we have equality iff A is constant.

Lemma 6.3 (Properties of the Shannon entropy).

- (a) Nonnegativity: $H(p) \geq 0$. Moreover, $H(p) = 0$ if and only if p is deterministic (i.e., $p(x) = 1$ for one x and all other probabilities are zero).
- (b) Upper bound: $H(p) \leq \log |\{x : p(x) > 0\}| \leq \log |\Sigma|$. Moreover, $H(p) = \log |\Sigma|$ if and only if p is uniform, i.e., $p(x) = 1/|\Sigma|$ for all $x \in \Sigma$.
- (c) Concavity: The Shannon entropy is a strictly concave function of $p \in \mathcal{P}(\Sigma)$.

Proof.

- (a) The lower bound holds since $f(q) = q \log \frac{1}{q} \geq 0$ for any $q \in [0, 1]$. Moreover, $f(q) = 0$ iff $q \in \{0, 1\}$, which implies the second claim. See also the figure above.
- (b) This follows from Jensen's inequality, applied to the concave log function and $a(x) = 1/p(x)$. Indeed,

$$H(p) = \sum_{x \in \Sigma, p(x) > 0} p(x) \log \frac{1}{p(x)} \leq \log \sum_{x \in \Sigma, p(x) > 0} p(x) \frac{1}{p(x)} = \log |\{x : p(x) > 0\}|,$$

with equality if and only if all nonzero $p(x)$ are equal. Now the rest is clear.

- (c) This follows if we can show that

$$f(q) = q \log \frac{1}{q} = -\frac{1}{\ln 2} q \ln q$$

is strictly concave on $q \in [0, 1]$. Indeed, for $q > 0$,

$$f'(q) = -\frac{1}{\ln 2} (\ln q + 1) \quad \text{and so} \quad f''(q) = -\frac{1}{\ln 2} \frac{1}{q} < 0. \quad \square$$

Definition 6.4 (Subscripts, entropy of subsystems). When dealing with joint distributions, it is often useful to use subscripts to denote the distribution of a random variable. Thus, if X and Y are random variables then we might write p_{XY} for their joint distribution and p_X, p_Y for their marginal distributions, etc. That is,

$$\begin{aligned} p_{XY}(x, y) &= \Pr(X = x, Y = y), \\ p_X(x) &= \Pr(X = x) = \sum_y p_{XY}(x, y), \\ p_Y(y) &= \Pr(Y = y) = \sum_x p_{XY}(x, y). \end{aligned}$$

We already discussed and used this convention in Eq. (4.24). It will also be useful to write Σ_X for the space of outcomes of a random variable X , i.e., if $p_X \in \mathcal{P}(\Sigma_X)$. This is completely analogous to our notation and conventions in the quantum case, see Definitions 2.5 and 2.12.

Similarly, we will denote the entropies of subsets of the random variables by

$$H(XY) := H(p_{XY}), \quad H(X) := H(p_X), \quad H(Y) := H(p_Y).$$

Sometimes we will also write $H(XY)_p, H(X)_p$, etc. if we want to be explicit about the underlying probability distribution.

Today we only use this notation to state the following lemma, which you can prove in Exercise 6.6.

Lemma 6.5 (Monotonicity and subadditivity of the Shannon entropy). *Given random variables X and Y , the following inequalities for the Shannon entropy hold:*

- (a) Monotonicity: $H(XY) \geq H(Y)$.
- (b) Subadditivity: $H(X) + H(Y) \geq H(XY)$.

We now turn towards today's main goal, which is to give an interpretation of the Shannon entropy in the context of compression.

6.2 Lossy and lossless compression

Consider a data source modeled by a random variable $X \sim p \in \mathcal{P}(\Sigma)$. We would like to compress one sample from X into a bitstring of length ℓ . By this we mean that we would like to come up with an encoder E and a decoder D such that $\tilde{X} := D(E(X))$ is equal to X (i.e., first compressing and then decompressing does recover the original input). This is illustrated in the following picture:



How small can we choose ℓ to be? The answer is given by the *raw bit content* of p , which is defined as follows:

$$H_0(p) := \log |\{x \in \Sigma : p(x) > 0\}|.$$

Indeed, the encoder E needs to assign a distinct bitstring in $\{0,1\}^\ell$ to each element x that occurs with nonzero probability – this can be done if and only if $\ell \geq H_0(p)$. Clearly, this is not a very interesting result – we are not doing any compression at all. How can we do better? There are two main options:

- (a) *Lossy fixed-length compression:* We could allow a small probability of error, i.e., only demand that $\Pr(\tilde{X} \neq X) \leq \delta$ for some $\delta > 0$.
- (b) *Lossless variable-length compression:* We could use bitstrings of different lengths $\ell = \ell(x)$ depending on the element x that is sampled and try to minimize the average length.

Here is a concrete example:

Example 6.6. Consider the following distribution on $\Sigma = \{A, B, C\}$:

$$p(A) = 0.98, \quad p(B) = 0.01, \quad p(C) = 0.01$$

Clearly, $H_0(p) = \log 3 \approx 1.58$, so we need at least $\ell = 2$ bits to achieve (6.5). Let us discuss the two options: (a) If we are willing to tolerate a probability of error $\delta = 0.01$ then we can compress into a single bit ($\ell = 1$). For example, we might define the encoder and decoder by

x	E(x)	s	D(s)
A	0	0	A
B	1	1	B
C	1		

(b) If we are willing to use bitstrings of varying length then the following encoder and decoder

x	E(x)	s	D(s)
A	0	0	A
B	10	10	B
C	11	11	C

achieves an average length of $0.98 \times 1 + 0.02 \times 2 = 1.02$ with *no* error.

Note that none of the ‘codewords’ $E(x)$ is a prefix of any other – this ensures that we can decode a given bitstring without having to use an additional ‘end of input’ symbol.

This goes already in the right direction but is still not very impressive. For example, suppose that we have a source that emits two symbols A and B with probabilities

$$p(A) = 0.75, \quad p(B) = 0.25.$$

There should clearly be some potential for savings, since this situation seems much less random than if the two probabilities were the same. But neither of the two options above seem very helpful – for a lossy protocol we would need to allow a probability of failure of $\delta = 25\%$, while for a lossless protocol there is no better way than sending $\ell = 1$ bit for both messages (since we cannot send partial bitstrings).

How can we do better? The key idea is to try to compress not a single symbol at a time but to focus on *blocks* of many symbols. We will discuss how this can be done in detail for lossy compression and defer a discussion of the lossless case to Exercise 6.4.

6.3 Block codes, Shannon’s source coding theorem, typical sets

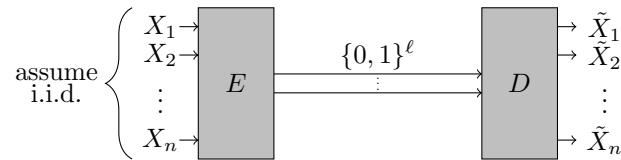
The basic assumption will be that our data source is *IID* (or *memoryless*), which means that it emits symbols

$$X_1, X_2, \dots, X_n \stackrel{\text{IID}}{\sim} p$$

for some $p \in \mathcal{P}(\Sigma)$. This notation means that the X_i are *independent and identically distributed* (IID) random variables such that each X_i has distribution p .

Remark 6.7. While the IID assumption may not necessarily be a realistic assumption when it comes to a concrete data source (e.g., typical data sources may exhibit correlations or may change over time), it is a very useful base case. For more sophisticated compression schemes, see <https://staff.fnwi.uva.nl/m.walter/iit19/>.

Schematically, what we would like to achieve is the following. We would like to find an encoder and decoder, now operating on a block or sequence of n symbols, as in the following figure,



such that

$$\Pr(\tilde{X}^n \neq X^n) \leq \delta.$$

Here and below we use the notation $X^n = (X_1, \dots, X_n)$ for sequences of length n if we want to emphasize their length. Note that for $n = 1$ the above reduces to Eq. (6.5). Our goal now is to minimize the *compression rate*

$$\frac{\ell}{n} = \frac{\text{number of bits}}{\text{block length}}.$$

We now formalize the above in a definition and state Shannon’s central theorem, which shows that the optimal compression rate is directly related to the Shannon entropy (if we allow $n \rightarrow \infty$).

Definition 6.8 (Code). An (n, R, δ) -code for $p \in \mathcal{P}(\Sigma)$ is a pair of functions

$$E: \Sigma^n \rightarrow \{0, 1\}^{\lfloor nR \rfloor} \quad \text{and} \quad D: \{0, 1\}^{\lfloor nR \rfloor} \rightarrow \Sigma^n$$

such that

$$\Pr(D(E(X^n)) \neq X^n) \leq \delta \tag{6.6}$$

for $X^n \stackrel{\text{iid}}{\sim} p$.

Note that the left-hand side of Eq. (6.6) can also be written as

$$\Pr(D(E(X^n)) \neq X^n) = \sum_{x^n \in \Sigma^n: D(E(x^n)) \neq x^n} p(x^n) = \sum_{x^n \in \Sigma^n: D(E(x^n)) \neq x^n} p(x_1) \cdots p(x_n),$$

where we write $p(x^n) := p(x_1) \cdots p(x_n)$ for the joint distribution of a sequence $x^n \in \Sigma^n$.

Theorem 6.9 (Shannon's source coding). Let $p \in \mathcal{P}(\Sigma)$ and $\delta \in (0, 1)$. Then:

- (a) If $R > H(p)$ then there exists n_0 such that there exists an (n, R, δ) -code for all $n \geq n_0$.
- (b) If $R < H(p)$ then there exists n_0 such that no (n, R, δ) -codes exist for $n \geq n_0$.

That is, the *optimal asymptotic compression rate* for an *iid source* described by a probability distribution is given by its Shannon entropy.

To prove Theorem 6.9, we need to make use of the fact that not all sequences x^n are equally likely. For example, for large n , we might expect that with high probability the number of times that any given symbol x appears in X^n is $\approx n(p(x) \pm \varepsilon)$. The following definition captures a closely related property of 'typical' sequences:

Definition 6.10 (Typical set). For $p \in \mathcal{P}(\Sigma)$, $n \in \mathbb{N}$, and $\varepsilon > 0$, define the *typical set*

$$\begin{aligned} T_{n,\varepsilon}(p) &:= \left\{ x^n \in \Sigma^n : \left| \frac{1}{n} \log \frac{1}{p(x^n)} - H(p) \right| \leq \varepsilon \right\} \\ &= \left\{ x^n \in \Sigma^n : \left| \frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(x_i)} - H(p) \right| \leq \varepsilon \right\} \end{aligned}$$

The following lemma summarizes the most important properties of the typical sets.

Lemma 6.11 (Asymptotic Equipartition Property, AEP). The following properties hold:

- (a) $2^{-n(H(p)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(p)-\varepsilon)}$ for all $x^n \in T_{n,\varepsilon}(p)$.
- (b) $|T_{n,\varepsilon}(p)| \leq 2^{n(H(p)+\varepsilon)}$.
- (c) For $X^n \stackrel{\text{iid}}{\sim} p$, it holds that $\Pr(X^n \notin T_{n,\varepsilon}(p)) \leq \frac{\sigma^2}{n\varepsilon^2}$. Here, $\sigma^2 = \text{Var}(\log \frac{1}{p(X_i)})$ is a constant that only depends on p .

Proof. (a) This is just restating the definition.

(b) This follows from

$$1 \geq \Pr(X^n \in T_{n,\varepsilon}(p)) \geq |T_{n,\varepsilon}(p)| 2^{-n(H(p)+\varepsilon)},$$

where the last step is the lower bound in part (a).

- (c) Define the random variables $R_i := \log \frac{1}{p(X_i)}$. Then the R_1, \dots, R_n are IID, with expectation value $\mu = E[R_i] = H(p)$ (Eq. (6.2)) and variance $\text{Var}(R_i) = \sigma^2$. Now,

$$\Pr(X^n \notin T_{n,\varepsilon}(p)) = \Pr\left(\left|\frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(X_i)} - H(p)\right| > \varepsilon\right) = \Pr\left(\left|\frac{1}{n} \sum_{i=1}^n R_i - \mu\right| > \varepsilon\right).$$

The weak law of large number states that the right-hand side converges to zero for large n . Let us recall its proof to get a concrete bound. For this, define $Y := \frac{1}{n} \sum_{i=1}^n R_i$. Then,

$$E[Y] = \mu \quad \text{and} \quad \text{Var}(Y) = \frac{1}{n^2} \text{Var}(R_1 + \dots + R_n) = \frac{1}{n} \text{Var}(R_i) = \frac{\sigma^2}{n},$$

using that the variance of a sum of independent random variables is simply the sum of the individual variances. Now we can use the Chebyshev inequality, which states that

$$\Pr(|Y - E[Y]| > \varepsilon) \leq \frac{\text{Var}(Y)}{\varepsilon^2}$$

to conclude the proof. □

We are now in a good position to prove Shannon's source coding theorem.

Proof of Theorem 6.9. To prove part (a), let us choose $\varepsilon = \frac{R-H(p)}{2}$, noting that $\varepsilon > 0$. Then, using part (b) of Lemma 6.11,

$$|T_{n,\varepsilon}(p)| \leq 2^{n(H(p)+\varepsilon)} = 2^{n(R-\varepsilon)} \leq 2^{\lfloor nR \rfloor},$$

the final inequality holds provided we assume that $n \geq \frac{1}{\varepsilon}$. The above implies that there exists an injective map $E: T_{n,\varepsilon} \rightarrow \{0,1\}^{\lfloor nR \rfloor}$. Let us denote by $D: \{0,1\}^{\lfloor nR \rfloor} \rightarrow \Sigma^n$ its left inverse (i.e., $D(E(x^n)) = x^n$ for $x^n \in T_{n,\varepsilon}$). Finally, extend E arbitrarily to all of Σ^n . Then,

$$\Pr(D(E(X^n)) \neq X^n) \leq \Pr(X^n \notin T_{n,\varepsilon}(p)) \leq \frac{\sigma^2}{n\varepsilon^2} \leq \delta,$$

where we first used that only sequences outside the typical set can lead to errors (since $D(E(x^n)) = x^n$ for $x^n \in T_{n,\varepsilon}$) and then part (c) of Lemma 6.11; the final inequality holds if we assume that $n \geq \frac{\sigma^2}{\varepsilon^2\delta}$. Thus we have proved that there exists an (n, R, δ) -code for any $n \geq n_0 := \max\{\frac{1}{\varepsilon}, \frac{\sigma^2}{\varepsilon^2\delta}\}$. We emphasize that n_0 only depends on p , δ , and R , as it should.

How about the proof of part (b)? This is your Exercise 6.8! □

In Exercise 6.3 you can reflect on the practicalities of using typical sets for compression. In Exercise 6.4 you can discuss how to translate an (n, R, δ) -code into a corresponding lossless variable-length compression protocol.

Remark 6.12. The typical sets constructed in the proof are in general not the smallest sets S_n with the property that $\Pr(X^n \in S_n) \geq 1 - \delta$. However, they are easy to handle mathematically as $n \rightarrow \infty$ and still small enough (this is the content of part (b) of Theorem 6.9).

To obtain the smallest possible S_n , we could sort the strings x^n by decreasing probability and add one string after the other until we reach probability $1 - \delta$.

Next week we will discuss how to translate the above ideas into the quantum realm. Here there are many challenges, e.g., the states emitted by a quantum data source need not be orthogonal, so cannot be perfectly distinguished by the encoder, and at any rate the encoder is not allowed to measure the information as we typically destroy quantum information when we measure it – but we will see that all these challenges can be overcome!

6.4 Exercises

6.1 Joint distributions and entropies: Consider the following joint distribution of two random variables X and Y over $\{0, 1, 2\}$ and $\{0, 1\}$, respectively:

x	y	$p(x, y)$
0	0	$1/2$
0	1	0
1	0	0
1	1	$1/4$
2	0	0
2	1	$1/4$

- Compute the joint entropy $H(X, Y)$.
 - Compute the marginal probability distributions $p(x)$ and $p(y)$.
 - Compute the entropies $H(X)$ and $H(Y)$.
 - Are X and Y independent?
- 6.2 Entropy and typical sets:** Let p be the probability distribution with three possible outcomes 0, 1, 2 and probabilities $p(0) = 1/2$, $p(1) = 1/4$, $p(2) = 1/4$. Let X_1, X_2 be independent and identically distributed (IID) according to p .
- Compute $H(X_1) = H(X_2) = H(p)$. What is $H(X_1, X_2)$?
 - Make a table that lists the joint probability $p(x_1, x_2)$ and the quantity $\frac{1}{2} \log \frac{1}{p(x_1, x_2)}$ for all possible outcomes x_1 and x_2 .
 - Write down all elements of the typical set $T_{2, \varepsilon}(p)$ for $\varepsilon = 0.12345$.
- 6.3 How to compress it?** Suppose you would like to compress an IID source. In class we showed how such a source can in principle be compressed by using typical sets. Discuss how this can be applied in practice. What parameters have to be fixed? How do the encoder and decoder work? What if you don't know the distribution of symbols emitted by the source? Is this a *practical* way of compressing?
- 6.4 Lossy vs. lossless compression:** Given an (n, R, δ) -code defined as in Definition 6.8, can you construct a lossless compression protocol with average rate $\approx R$? You may assume that n is large and δ is small.
- 6.5 Typical sets:** Let p be a probability distribution on an set Σ . For $t \geq 0$, define

$$S_{n,t} = \{x^n \in \Sigma^n : p(x^n) \geq 2^{-nt}\}.$$

- Show that $S_{n,t}$ contains no more than 2^{nt} strings.
 - Show that, if $t > H(p)$, then $S_{n,t}$ contains a typical set $T_{n, \varepsilon}(p)$ for some $\varepsilon > 0$.
- 6.6 Properties of the Shannon entropy:** Prove Lemma 6.5. Can you interpret the two inequalities in the context of compression? *Hint: For both (a) and (b), write the left-hand side minus the right-hand side of the inequality as a single expectation value. For (b), use Jensen's inequality.*
- 6.7 Binary entropy function:** Is the binary entropy (Example 6.2) Lipschitz continuous? That is, is there a constant $L > 0$ such that $|h(p) - h(q)| \leq L|p - q|$ for all $0 \leq p, q \leq 1$?

6.8 Optimality of the Shannon entropy: In this problem, you will prove the converse part of Shannon's source coding theorem which states that it is impossible to compress at rates below the entropy of the source. Given a probability distribution p on a finite set Σ , recall that an (n, R, δ) -code consists of functions $E: \Sigma^n \rightarrow \{0, 1\}^{\lfloor nR \rfloor}$ and $D: \{0, 1\}^{\lfloor nR \rfloor} \rightarrow \Sigma^n$ such that $\sum_{x^n \in \Sigma^n: D(E(x^n)) = x^n} p(x_1) \cdots p(x_n) \geq 1 - \delta$. Show that:

- (a) For any (n, R, δ) -code, there are at most 2^{nR} many strings x^n such that $D(E(x^n)) = x^n$.
- (b) For fixed $\delta \in (0, 1)$ and $R < H(p)$, (n, R, δ) -codes can only exist for finitely many n .


Hint: Distinguish between typical and atypical sequences.

6.9 Lexicographic order (for the bonus problem): The lexicographic order \leq_{lex} on $\{0, 1\}^n$ is defined as follows: Given bitstrings x^n and y^n , we let $x^n \leq_{\text{lex}} y^n$ if either $x^n = y^n$ or $x_i < y_i$ for the smallest i such that $x_i \neq y_i$. For example, $001 \leq_{\text{lex}} 010$. The lexicographic order defines a total order on $\{0, 1\}^n$, hence also on the bitstrings of length n with k ones, which we denote by $B(n, k)$.

- (a) Write down $B(5, 2)$ in lexicographic order (smallest element first).
- (b) How can you recursively compute the m -th element of $B(n, k)$?
- (c) How can you recursively compute the index of a given element in $B(n, k)$?

Hint: $|B(n, k)| = \binom{n}{k}$. Moreover, $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$ for all $1 \leq k \leq n-1$.

6.10 Practice: A binary image of size $r \times s$ can be represented by a bitstring of length rs , where we list the pixel values (0=black pixel, 1=white pixel) row by row, starting with the top row. We can thus compress the image in the following *lossless* fashion: First, compute the number k of ones in the bitstring. Next, compute the index $m \in \{0, 1, \dots, \binom{rs}{k} - 1\}$ of the bitstring in the lexicographically sorted list of all bitstrings of length rs that contain k ones. The quadruple (r, s, k, m) defines the compression of the image.

For example, the 2×3 -image  corresponds to the bitstring 000100. There are six strings with $k = 1$ ones. In lexicographic order: 000001, 000010, 000100, 001000, 010000, and 100000. The index of our bitstring in this list is $m = 2$. Thus, we would compress this picture by $(2, 3, 1, 2)$.

- (a) What is the bitstring corresponding to the following image? What is its compression?



- (b) Can you decompress the image given by $(r, s, k, m) = (7, 8, 8, 243185306)$?

Lecture 7

From classical to quantum compression

Last week we discussed how to compress a classical data source which emits a symbol IID according to a known probability distribution. We discussed two paradigms for compression – lossy fixed-length compression and lossless variable-length compression – and their relation. We then zoomed into the lossy paradigm and proved Shannon’s source coding theorem, which states that, in the limit of large block lengths, the optimal compression rate of a source is computed by its Shannon entropy (see Theorem 6.9 for a precise statement). Today, we will see the quantum analogs of these results. We will define the von Neumann entropy of quantum states, the notion of a quantum code, and prove Schumacher’s theorem that computes the optimal compression rate in the quantum scenario.

7.1 Von Neumann Entropy

As last week, we will first define the entropy and then discuss how it naturally arises in the context of compression.

Definition 7.1 (von Neumann entropy). The *von Neumann entropy* of a quantum state $\rho \in D(\mathcal{H})$ is defined as the Shannon entropy of its eigenvalues (cf. Definition 6.1). That is,

$$H(\rho) := H(p) \tag{7.1}$$

where $p = (p(1), \dots, p(d))$ is a probability distribution whose entries are the eigenvalues of ρ , repeated according to their multiplicity, and $d = \dim \mathcal{H}$.

We can also write the von Neumann entropy more intrinsically in the following way:

$$H(\rho) = -\operatorname{Tr}[\rho \log \rho]. \tag{7.2}$$

Let us discuss how “ $\rho \log \rho$ ” is defined. In general, if Q is positive definite then its *logarithm*, denoted $\log Q$ or $\log(Q)$, is the Hermitian operator with the same eigenvectors but eigenvalues the logarithm of those of Q . That is, if $Q = \sum_i \lambda_i |e_i\rangle\langle e_i|$ is an eigendecomposition then $\log Q = \sum_i \log(\lambda_i) |e_i\rangle\langle e_i|$. This is a special case of Definition 1.6 and completely analogous to the definition of the square root \sqrt{Q} . Note that $\log Q$ is always Hermitian but typically not PSD. You can practice this definition in Exercises 7.1 and 7.2. If ρ is positive definite then we can use this definition to define $\log \rho$ and hence $\rho \log \rho$.

If ρ has some zero eigenvalues then $\log \rho$ is ill-defined. However, recall from the discussion below Definition 6.1 that the function $f(q) = q \log q$ can be extended to $q \geq 0$ by continuity.

Thus we can still define $\rho \log \rho$ for all $\rho \in \mathcal{D}(\mathcal{H})$. If ρ is positive definite then this definition coincides with the one given above. Thus, Eq. (7.2) is well-defined and holds for all $\rho \in \mathcal{D}(\mathcal{H})$.

We now state some properties of the von Neumann entropy that are analogous to Lemma 6.3 for Shannon entropy.

Lemma 7.2 (Properties of von Neumann entropy).

(a) Nonnegativity: $H(\rho) \geq 0$. Moreover, $H(\rho) = 0$ if and only if ρ is pure (i.e., $\rho = |\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle \in \mathcal{H}$).

(b) Upper bound:

$$H(\rho) \leq \log \text{rank}(\rho) \leq \log \dim \mathcal{H}.$$

Moreover, $H(\rho) = \log(\dim \mathcal{H})$ if and only if ρ is maximally mixed (i.e., $\rho = \frac{I}{\dim \mathcal{H}}$).

(c) Invariance under isometries: $H(\rho) = H(V\rho V^\dagger)$ for any isometry V .

(d) Continuity: The von Neumann entropy is continuous.

(e) Concavity: The von Neumann entropy is a strictly concave function of $\rho \in \mathcal{D}(\mathcal{H})$.

Proof. The first two follow immediately from the corresponding properties of the Shannon entropy in Lemma 6.3. The invariance under isometries holds since the entropy only depends on the nonzero eigenvalues – but the latter are the same for ρ and $V\rho V^\dagger$. The continuity follows because the Shannon entropy is continuous and the sorted eigenvalues of a Hermitian operator depend continuously on the operator (but we will not prove this). A quantitative bound is stated below in Theorem 7.3. You will prove concavity in Exercise 7.10 (c) and strict concavity in Exercise 8.4. See p. 76 in Lecture 6 for the definition of concavity and strict concavity. \square

The following theorem, which we do not prove, gives a quantitative bound for the continuity of the von Neumann entropy:

Theorem 7.3 (Fannes–Audenaert). For all $\rho, \sigma \in \mathcal{D}(\mathcal{H})$,

$$|H(\rho) - H(\sigma)| \leq t \log(\dim \mathcal{H} - 1) + h(t),$$

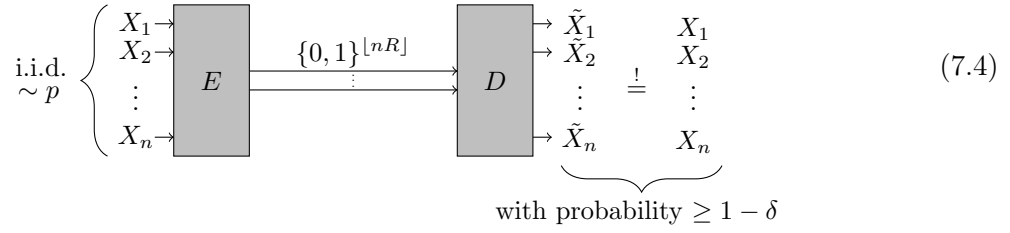
where $t = T(\rho, \sigma)$ is the trace distance between the two states and $h(t)$ denotes the binary Shannon entropy discussed in Example 6.2 and Exercise 6.7.

7.2 Motivation: Classical compression and correlations

Before we turn to compressing quantum data, let us briefly revisit the classical case. Recall from Definition 6.8 that an (n, R, δ) -code for a probability distribution $p \in \mathcal{P}(\Sigma)$ consists of functions $E: \Sigma^n \rightarrow \{0, 1\}^{\lfloor nR \rfloor}$ and $D: \{0, 1\}^{\lfloor nR \rfloor} \rightarrow \Sigma^n$ such that

$$\Pr(\tilde{X}^n \neq X^n) \leq \delta \tag{7.3}$$

for $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p$, where $\tilde{X}^n := D(E(X^n))$. Pictorially:



Shannon's source coding theorem asserts that $H(p)$ is the optimal rate R for compression in this context (see Theorem 6.9 for the precise statement).

How about if X^n is *correlated* to another random variable Y ? For example, suppose that $Y = X_1$, or $Y = X_1 \oplus \dots \oplus X_n$ or even $Y = X^n$. Are these correlations *preserved* if we replace X^n by \tilde{X}^n ? To state this question precisely, let $p_{X^n Y}$ denote the joint distribution of (X^n, Y) and let $p_{\tilde{X}^n Y}$ denote the joint distribution of (\tilde{X}^n, Y) . Then we would like to ask if it is true that $p_{X^n Y} \approx p_{\tilde{X}^n Y}$. This can be quantified by using the trace distance for probability distributions which is defined as follows:

Definition 7.4 (Trace distance). Given probability distributions $p, q \in \mathcal{P}(\Sigma)$, their (*normalized*) *trace distance* or *total variation distance* is defined as

$$T(p, q) := \frac{1}{2} \sum_{z \in \Sigma} |p(z) - q(z)| = \frac{1}{2} \|p - q\|_1,$$

where $\|x\|_1 = \sum_{z \in \Sigma} |x_z|$ denotes the ℓ_1 -norm of vector x .

Note that this is nothing but the trace distance (see Definition 4.6) of the corresponding classical states. In Exercise 7.3, you will prove the following two properties:

- (a) If Z, \tilde{Z} are random variables over Σ with distributions p, q , respectively, then

$$T(p, q) = \max_{S \subseteq \Sigma} (\Pr(Z \in S) - \Pr(\tilde{Z} \in S)). \quad (7.5)$$

- (b) If Z and \tilde{Z} are as above and have a joint distribution then it holds that

$$T(p, q) \leq \Pr(Z \neq \tilde{Z}). \quad (7.6)$$

Eq. (7.6) is known as the *coupling inequality*. This is because, in probability theory, a joint distribution of a given pair of marginal distributions is often called a *coupling*.

Then we have the following lemma, which shows that not only are correlations preserved in a precise quantitative sense but that this in fact *characterizes* a reliable code!

Lemma 7.5. Let $p \in \mathcal{P}(\Sigma)$ and $E: \Sigma^n \rightarrow \{0, 1\}^{[Rn]}$, $D: \{0, 1\}^{[Rn]} \rightarrow \Sigma^n$ be an arbitrary pair of functions. Then, (E, D) is an (n, R, δ) -code for p if and only if

$$T(p_{X^n Y}, p_{\tilde{X}^n Y}) \leq \delta$$

for any joint distribution $p_{X^n Y}$ of random variables $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} p$ and Y , where $p_{\tilde{X}^n Y}$ denotes the joint distribution of $\tilde{X}^n = D(E(X^n))$ and Y .

Proof. (\Rightarrow): Using the coupling inequality Eq. (7.6) for $Z = (X^n, Y)$ and $\tilde{Z} = (\tilde{X}^n, Y)$,

$$T(p_{X^n Y}, p_{\tilde{X}^n Y}) \leq \Pr(Z \neq \tilde{Z}) = \Pr(X^n \neq \tilde{X}^n) \leq \delta,$$

where the last inequality is Eq. (7.3), using that (E, D) is by assumption an (n, R, δ) -code.

(\Leftarrow): Choose $Y = X^n$. Then,

$$\Pr(\tilde{X}^n \neq X^n) = \Pr(\tilde{X}^n \neq Y) = \Pr(\tilde{X}^n \neq Y) - \underbrace{\Pr(X^n \neq Y)}_{=0} \leq T(p_{\tilde{X}^n Y}, p_{X^n Y}) \leq \delta,$$

where the first inequality is Eq. (7.5) for the event $S = \{(x^n, y) : x^n \neq y\}$. \square

7.3 Quantum codes and compression

We just saw that good codes are characterized by the property that they approximately preserve all correlations. We will take this as the definition in the quantum case. Recall from Definition 4.15 that $C(\mathcal{H}_A, \mathcal{H}_B)$ denotes the set of all quantum channels from \mathcal{H}_A to \mathcal{H}_B .

Definition 7.6 (Quantum code). An (n, R, δ) -quantum code for $\rho \in D(\mathcal{H}_A)$ is a pair of channels

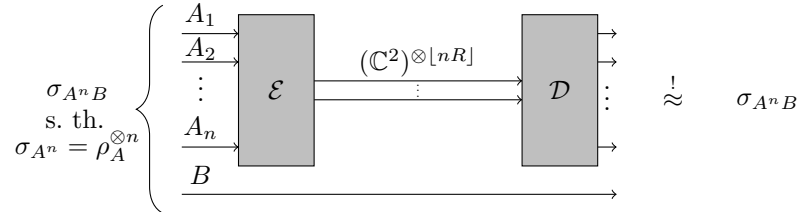
$$\mathcal{E} \in C(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}) \quad \text{and} \quad \mathcal{D} \in C((\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}, \mathcal{H}_A^{\otimes n})$$

such that

$$F(\sigma_{A^n B}, (\mathcal{D} \circ \mathcal{E} \otimes \mathcal{I}_B)[\sigma_{A^n B}]) \geq 1 - \delta \quad (7.7)$$

for all finite-dimensional \mathcal{H}_B and states $\sigma_{A^n B} \in D(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B)$ such that $\sigma_{A^n} = \rho_A^{\otimes n}$.

Here we use the fidelity rather than the trace distance – otherwise Definition 7.6 is completely analogous to the condition in Lemma 7.5. The following pictures illustrates the definition:



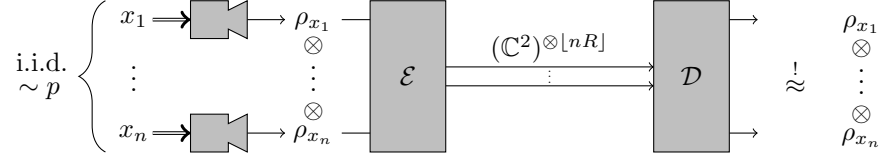
Definition 7.6 is perhaps surprising and raises three immediate questions:

- (1) What does the definition have to do with compression in the ‘ordinary’ sense of compressing the output of a source?
- (2) Is there any way to simplify the condition in Eq. (7.7) so that it no longer refers to infinitely many options for $\sigma_{A^n B}$?
- (3) What is the optimal rate of compression – is there an analog to Shannon’s theorem?

We will address these questions one after the other.

First, let us relate Definition 7.6 to compression of a source. In analogy to the discussion in Lecture 6, we imagine that a *quantum source* emits states $\rho_x \in D(\mathcal{H}_A)$ for $x \in \Sigma$ according to a known probability distribution $p \in P(\Sigma)$. We will further imagine the source to be *IID (or memoryless)*, which means that it emits states $\rho_{x_1} \otimes \dots \otimes \rho_{x_n}$ according to the IID distribution $p(x^n) = p(x_1) \dots p(x_n)$. What would it mean to compress such a quantum source?

Clearly, we would like to have



on average or even with high probability. For example, we might like to show that

$$\sum_{x^n \in \Sigma^n} p(x_1) \cdots p(x_n) F(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}, \mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]]) \geq 1 - \delta. \quad (7.8)$$

This looks similar to Eqs. (7.3) and (7.4), except that we are now happy to recover $\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}$ approximately (since we are dealing with quantum states it turns out that we cannot in general hope for equality).

We will now show that Eq. (7.8) can indeed be achieved by using quantum codes. For this, suppose that $(\mathcal{E}, \mathcal{D})$ is an (n, R, δ) -quantum code for the *average output state* of the source, i.e.,

$$\rho = \sum_{x \in \Sigma} p(x) \rho_x.$$

Why does this help? To make use of Eq. (7.7), we need to construct a state that extends $\rho^{\otimes n}$. We will consider the following state

$$\sigma_{A^n X^n} := \sum_{x^n \in \Sigma^n} p(x^n) \rho_{x_1} \otimes \cdots \otimes \rho_{x_n} \otimes |x^n\rangle\langle x^n|,$$

on $D(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_X^{\otimes n})$, where $\mathcal{H}_X = \mathbb{C}^\Sigma$. Both the state $\sigma_{A^n X^n}$ and

$$(\mathcal{D} \circ \mathcal{E} \otimes \mathcal{I}_{X^n})[\sigma_{A^n X^n}] = \sum_{x^n \in \Sigma^n} p(x^n) \mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]] \otimes |x^n\rangle\langle x^n|$$

are classical on the X^n -system, with the same probability distribution. Thus,

$$\sum_{x^n \in \Sigma^n} p(x^n) F(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}, \mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]]) = F(\sigma_{A^n X^n}, (\mathcal{D} \circ \mathcal{E} \otimes \mathcal{I}_{X^n})[\sigma_{A^n X^n}]) \geq 1 - \delta,$$

where the equality holds thanks to Exercise 5.12 (b) and the inequality is simply by Eq. (7.7) in the definition of a quantum code, applied to the state $\sigma_{A^n X^n}$.

Thus we have proved that Eq. (7.7) implies Eq. (7.8), meaning that a quantum code for ρ can be used for compressing any quantum source with average output state ρ . In Exercise 7.7 you will show that in general the converse is *not* true. This makes sense, since Eq. (7.8) refers to a single source, while we just proved that Eq. (7.7) ensures that *any* source with average output state ρ can be compressed reliably.

We close this section with some warnings to avoid some common traps that one can fall into when thinking about compressing quantum sources:

- In general, there is no relation between the number of states ρ_x and the Hilbert space dimension (i.e., in general $|\Sigma| \neq \dim \mathcal{H}_A$).
- The states ρ_x for $x \in \Sigma$ need *not* be pure nor pairwise orthogonal.
- The $p(x)$ need *not* be the eigenvalues of the average state $\rho = \sum_x p(x) \rho_x$.

7.4 Channel fidelity

We now turn to the second question raised above – how can we check the condition in Eq. (7.7) without having to consider all possible states $\sigma_{A^n B}$? We start with a definition that abstracts the situation.

Definition 7.7 (Channel fidelity). Given a channel $\mathcal{T}_A \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_A)$ and a state ρ_A , define the *channel fidelity* as

$$F(\mathcal{T}_A, \rho_A) := \inf \left\{ F(\sigma_{AB}, (\mathcal{T}_A \otimes \mathcal{I}_B)[\sigma_{AB}]) : \mathcal{H}_B, \sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) \text{ such that } \sigma_A = \rho_A \right\}.$$

Given this definition, we can rephrase Eq. (7.7) in the definition of a quantum code as

$$F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geq 1 - \delta. \quad (7.9)$$

Why is this progress? It turns out that we can always compute the channel fidelity by considering an arbitrary purification.

Lemma 7.8. Let $\sigma_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ be an arbitrary purification of ρ_A . Then,

$$F(\mathcal{T}_A, \rho_A) = F(\sigma_{AB}, (\mathcal{T}_A \otimes \mathcal{I}_B)[\sigma_{AB}]).$$

Proof. This follows readily from the fidelity's monotonicity and invariance under isometries. \square

As a consequence we find a simple expression in terms of a Kraus representation.

Corollary 7.9. Let $\mathcal{T}_A[M_A] = \sum_i X_i M_A X_i^\dagger$ be a Kraus representation. Then,

$$F(\mathcal{T}_A, \rho_A) = \sqrt{\sum_i |\text{Tr}[X_i \rho_A]|^2}.$$

Proof. Let $\sigma_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ be an arbitrary purification of ρ_A . Then,

$$\begin{aligned} F(\mathcal{T}_A, \rho_A)^2 &= F(\sigma_{AB}, (\mathcal{T}_A \otimes \mathcal{I}_B)[\sigma_{AB}])^2 \\ &= \langle \Psi_{AB} | (\mathcal{T}_A \otimes \mathcal{I}_B) [|\Psi_{AB}\rangle\langle\Psi_{AB}|] | \Psi_{AB} \rangle \\ &= \sum_i \langle \Psi_{AB} | (X_i \otimes I_B) | \Psi_{AB} \rangle \langle \Psi_{AB} | (X_i^\dagger \otimes I_B) | \Psi_{AB} \rangle \\ &= \sum_i |\langle \Psi_{AB} | X_i \otimes I_B | \Psi_{AB} \rangle|^2 = \sum_i |\text{Tr}[X_i \rho_A]|^2, \end{aligned}$$

where we first used Lemma 7.8, then Eq. (4.11) to evaluate the fidelity, and finally the Kraus representation of \mathcal{T}_A . \square

7.5 Schumacher's theorem and typical subspaces

With the preceding theory in hand we shall now address the third and main question of today's lecture – what is the optimal rate of quantum compression? The following theorem due to Schumacher gives a precise solution.

Theorem 7.10 (Schumacher compression). *Let $\rho \in D(\mathcal{H}_A)$ and $\delta \in (0, 1)$. Then:*

- (a) *If $R > H(\rho)$ then there exists n_0 such that there exists an (n, R, δ) -quantum code for all $n \geq n_0$.*
- (b) *If $R < H(\rho)$ then there exists n_0 such that no (n, R, δ) -quantum codes exist for $n \geq n_0$.*

Just like Shannon's theorem was proved using typical sets, we will prove Schumacher's theorem by using the closely related notion of a typical subspace.

Definition 7.11 (Typical subspace and projector). For $\rho \in D(\mathcal{H}_A)$, $n \in \mathbb{N}$, and $\varepsilon > 0$, define the *typical subspace*

$$S_{n,\varepsilon}(\rho) = \text{span} \{ |e_{y_1}\rangle \otimes \cdots \otimes |e_{y_n}\rangle : y^n \in T_{n,\varepsilon}(q) \},$$

where $\rho = \sum_{y=1}^d q(y) |e_y\rangle\langle e_y|$ is an eigendecomposition of ρ and $d = \dim \mathcal{H}_A$.

Moreover, we define the *typical projector* $\Pi_{n,\varepsilon}(\rho)$ as the orthogonal projection onto the typical subspace $S_{n,\varepsilon}(\rho) \subseteq \mathcal{H}_A^{\otimes n}$. We will often abbreviate it by $\Pi_{n,\varepsilon}$.

To motivate this definition, note that

$$\begin{aligned} \rho^{\otimes n} &= \sum_{y^n} q(y_1) \cdots q(y_n) (|e_{y_1}\rangle \otimes \cdots \otimes |e_{y_n}\rangle) (\langle e_{y_1}| \otimes \cdots \otimes \langle e_{y_n}|) \\ &= \sum_{y^n} q(y_1) \cdots q(y_n) |e_{y_1}\rangle\langle e_{y_1}| \otimes \cdots \otimes |e_{y_n}\rangle\langle e_{y_n}|, \end{aligned} \quad (7.10)$$

so we recognize that the eigenvalues of $\rho^{\otimes n}$ are precisely given by the IID probabilities $q(y^n) := q(y_1) \cdots q(y_n)$. It is useful to note that the typical projector is diagonal in the same basis, since

$$\Pi_{n,\varepsilon} = \sum_{y^n \in T_{n,\varepsilon}(q)} |e_{y_1}\rangle\langle e_{y_1}| \otimes \cdots \otimes |e_{y_n}\rangle\langle e_{y_n}|. \quad (7.11)$$

In particular, $\Pi_{n,\varepsilon}$ and $\rho^{\otimes n}$ commute with each other. The following lemma summarizes the most important properties of the typical subspaces.

Lemma 7.12 (Quantum Asymptotic Equipartition Property, QAEP). *With notation as above, the following properties hold:*

- (a) *The nonzero eigenvalues of $\Pi_{n,\varepsilon} \rho^{\otimes n} \Pi_{n,\varepsilon} = \Pi_{n,\varepsilon} \rho^{\otimes n} = \rho^{\otimes n} \Pi_{n,\varepsilon}$ are within $2^{-n(H(\rho) \pm \varepsilon)}$,*
- (b) *$\text{rank } \Pi_{n,\varepsilon} = \dim S_{n,\varepsilon}(\rho) = |T_{n,\varepsilon}(q)| \leq 2^{n(H(\rho) + \varepsilon)}$,*
- (c) *$\text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}] \geq 1 - \frac{\sigma^2}{n\varepsilon}$, where σ^2 is a constant that only depends on the eigenvalues of ρ .*

Proof. These properties follow from the corresponding properties in Lemma 6.11. For property (b), this is immediate. To prove the other properties, note that Eqs. (7.10) and (7.11) imply that

$$\Pi_{n,\varepsilon} \rho^{\otimes n} \Pi_{n,\varepsilon} = \Pi_{n,\varepsilon} \rho^{\otimes n} = \rho^{\otimes n} \Pi_{n,\varepsilon} = \sum_{y^n \in T_{n,\varepsilon}(q)} q(y^n) |e_{y_1}\rangle\langle e_{y_1}| \otimes \cdots \otimes |e_{y_n}\rangle\langle e_{y_n}|.$$

This is an eigendecomposition, so we obtain property (a) from the corresponding property in Lemma 6.11. And since the preceding implies that

$$\text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}] = \sum_{y^n \in T_{n,\varepsilon}(q)} q(y^n) = \Pr(Y^n \in T_{n,\varepsilon}(q)),$$

where $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} p$, property (c) likewise follows from Lemma 6.11. □

We now prove Schumacher's theorem.

Proof of Theorem 7.10. To prove part (a), we start as in the proof of Shannon's source coding theorem and choose $\varepsilon = \frac{R-H(q)}{2} = \frac{R-H(\rho)}{2}$, which is $\varepsilon > 0$ by assumption. Then, using part (b) of Lemma 7.12,

$$\text{rank } \Pi_{n,\varepsilon} \leq 2^{n(H(\rho)+\varepsilon)} = 2^{n(R-\varepsilon)} \leq 2^{\lfloor nR \rfloor} = \dim((\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}); \quad (7.12)$$

the final inequality holds for large enough n (e.g., if $n \geq \frac{1}{\varepsilon}$). Eq. (7.12) implies that there exists a linear map $V: \mathcal{H}_A^{\otimes n} \rightarrow (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}$ such that

$$V^\dagger V = \Pi_{n,\varepsilon}.$$

Indeed, we can simply set $V = \sum_{i=1}^D |\psi_i\rangle \langle \varphi_i|$, where $D = \text{rank } \Pi_{n,\varepsilon}$, $\{|\varphi_i\rangle\}_{i=1}^D$ is a basis of the typical subspace and $\{|\psi_i\rangle\}_{i=1}^D$ some arbitrary set of orthonormal vectors in $(\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}$.¹ Finally, define the compressor and decompressor by

$$\begin{aligned} \mathcal{E}[M] &:= V M V^\dagger + \text{Tr}[\sqrt{I - V^\dagger V} M \sqrt{I - V^\dagger V}] \alpha, \\ \mathcal{D}[M] &:= V^\dagger M V + \text{Tr}[\sqrt{I - V V^\dagger} M \sqrt{I - V V^\dagger}] \beta, \end{aligned}$$

where α and β are arbitrary states. Note that $I - V^\dagger V$ and $I - V V^\dagger$ are PSD (since $V^\dagger V = \Pi_{n,\varepsilon}$ and $V V^\dagger$ have the same nonzero eigenvalues and the former is a projection), so that the square roots are well-defined PSD operators. It follows that \mathcal{E} and \mathcal{D} are completely positive and it is also easy to see that they are trace-preserving. Thus, we have defined channels $\mathcal{E} \in \mathcal{C}(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor})$ and $\mathcal{D} \in \mathcal{C}((\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}, \mathcal{H}_A^{\otimes n})$.

It remains to verify Eq. (7.7) or, equivalently, Eq. (7.9). For this, note that \mathcal{E} has a Kraus representation that includes the operator V , and \mathcal{D} has a Kraus representation that includes the operator V^\dagger . By Exercise 5.3, this means that $\mathcal{D} \circ \mathcal{E}$ has a Kraus representation starting with $V^\dagger V = \Pi_{n,\varepsilon}$. Hence, Corollary 7.9 implies that

$$F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geq |\text{Tr}[V^\dagger V \rho^{\otimes n}]| = \text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}].$$

But now property (c) in Lemma 7.12 shows that the right-hand side is $\geq 1 - \delta$ if we choose n sufficiently large. This concludes the proof of part (a).

We now prove part (b). Fix $\delta \in (0, 1)$ and $R < H(\rho)$. First, note that if P is an arbitrary orthogonal projection of rank $\leq 2^{nR}$ then

$$\begin{aligned} \text{Tr}[P \rho^{\otimes n}] &= \text{Tr}[P \Pi_{n,\varepsilon} \rho^{\otimes n}] + \text{Tr}[P(I - \Pi_{n,\varepsilon}) \rho^{\otimes n}] \\ &\leq \underbrace{\|P\|_1}_{\leq 2^{nR}} \underbrace{\|\Pi_{n,\varepsilon} \rho^{\otimes n}\|_\infty}_{\leq 2^{-n(H(\rho)-\varepsilon)}} + \underbrace{\text{Tr}[(I - \Pi_{n,\varepsilon}) \rho^{\otimes n}]}_{1 - \text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}]} \\ &\leq 2^{-n\varepsilon} + (1 - \text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}]) \end{aligned} \quad (7.13)$$

if we choose $\varepsilon = \frac{H(\rho)-R}{2}$. Here we estimated the left-hand side term using the Hölder inequality for operators from Eq. (4.9) and the operator norm using property (a) in Lemma 7.12. For the right-hand side term, we simply used that $P \leq I$ and rewrote the result. In view of property (c) in Lemma 7.12, the expression Eq. (7.13) converges to 0 as $n \rightarrow \infty$.

¹For example, we can use $V = \sum_{y^n \in T_{n,\varepsilon}(q)} |E(y^n)\rangle (\langle e_{y_1}| \otimes \cdots \otimes \langle e_{y_n}|)$, where $E: T_{n,\varepsilon} \rightarrow \{0, 1\}^{\lfloor nR \rfloor}$ is an arbitrary injective map and $|E(y^n)\rangle$ denotes the standard basis vector in $(\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}$ corresponding to $E(y^n) \in \{0, 1\}^{\otimes \lfloor nR \rfloor}$.

Now suppose that $(\mathcal{E}, \mathcal{D})$ is an (n, R, ε) -code. If $\{X_i\}$ are Kraus operators for \mathcal{E} and $\{Y_j\}$ are Kraus operators for \mathcal{D} , then $\{Z_k\} = \{Y_j X_i\}$ are Kraus operators for $\mathcal{D} \circ \mathcal{E}$. Since $X_i \in \mathcal{L}(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor})$, it has necessarily rank $\leq 2^{nR}$. Thus the same is true for the Kraus operators Z_k of $\mathcal{D} \circ \mathcal{E}$. Finally, let P_k denote the orthogonal projections onto the range of Z_k , so that the rank of P_k is likewise $\leq 2^{nR}$. We now evaluate the channel fidelity using Corollary 7.9 and obtain

$$\begin{aligned} F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n})^2 &= \sum_k |\text{Tr}[Z_k \rho^{\otimes n}]|^2 = \sum_k |\text{Tr}[P_k Z_k \rho^{\otimes n}]|^2 \\ &= \sum_k |\text{Tr}[Z_k \sqrt{\rho^{\otimes n}} \sqrt{\rho^{\otimes n}} P_k]|^2 \leq \sum_k \text{Tr}[Z_k^\dagger Z_k \rho^{\otimes n}] \text{Tr}[P_k \rho^{\otimes n}], \end{aligned}$$

where the inequality is by the Cauchy-Schwarz inequality for operators [Eq. (4.8)]. Since $\mathcal{D} \circ \mathcal{E}$ is a quantum channel, it is trace-preserving, so $\sum_k Z_k^\dagger Z_k = I$ by Lemma 5.5. This implies that $r(k) := \text{Tr}[Z_k^\dagger Z_k \rho^{\otimes n}]$ is a probability distribution. But then,

$$F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n})^2 \leq \sum_k r(k) \text{Tr}[P_k \rho^{\otimes n}] \leq 2^{-n\varepsilon} + (1 - \text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}])$$

by Eq. (7.13). By property (c) in Lemma 7.12, the right-hand side converges to 0 as $n \rightarrow \infty$. As a consequence, $F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geq 1 - \delta$ can only hold for finitely many n . In other words, (n, R, δ) -codes can only exist for finitely many values of n . \square

7.6 Exercises

7.1 Operator logarithm: Compute the logarithm of the following matrix: $\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$.

Hint: Hadamard basis.

7.2 Operator logarithm: Verify the following properties:

- (a) $\log(cI) = \log(c)I$ for every $c \geq 0$.
- (b) $\log(Q \otimes R) = \log(Q) \otimes I_B + I_A \otimes \log(R)$ for all positive definite $Q \in \mathcal{L}(\mathcal{H}_A)$, $R \in \mathcal{L}(\mathcal{H}_B)$.
- (c) $\log(\sum_{x \in \Sigma} p_x |x\rangle\langle x| \otimes \rho_x) = \sum_{x \in \Sigma} \log(p_x) |x\rangle\langle x| \otimes I_B + \sum_{x \in \Sigma} |x\rangle\langle x| \otimes \log(\rho_x)$ for every ensemble $\{p_x, \rho_x\}_{x \in \Sigma}$ of positive definite operators $\rho_x \in \mathcal{D}(\mathcal{H}_B)$.

Warning: It is in general not true that $\log(QR) = \log(Q) + \log(R)$. Indeed, QR is in general not even positive definite.

7.3 Trace distance of probability distributions: We defined the (normalized) trace distance between two probability distributions $p, q \in \mathcal{P}(\Sigma)$ by $T(p, q) := \frac{1}{2} \sum_{x \in \Sigma} |p(x) - q(x)|$.

- (a) Show that $T(p, q) = T(\rho, \sigma)$, where $\rho = \sum_x p(x) |x\rangle\langle x|$ and $\sigma = \sum_x q(x) |x\rangle\langle x|$.
- (b) Let X, Y be random variables with distributions p, q , respectively. Show that

$$T(p, q) = \max_{S \subseteq \Sigma} (\Pr(X \in S) - \Pr(Y \in S)).$$

Do you recognize this as the probability theory analog of a formula that you proved for quantum states?

- (c) Suppose X, Y are random variables as above and have a joint distribution. Use part (b) to show that $T(p, q) \leq \Pr(X \neq Y)$. This beautiful inequality is known as the *coupling inequality*.

7.4 Measurements and trace distance: In this problem, you will revisit how to distinguish quantum states by using measurements. Given states $\rho, \sigma \in D(\mathcal{H})$ and a measurement $\mu: \Omega \rightarrow \text{PSD}(\mathcal{H})$, let $p, q \in P(\Omega)$ denote the corresponding probability distributions of measurement outcomes.

- (a) Prove that $T(p, q) \leq T(\rho, \sigma)$.
- (b) Show that, for any ρ and σ , there exists Ω and a measurement μ such that equality holds.

Hint: Recall Helstrom's theorem.

7.5 Typical subspaces: Consider the state $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1| + \frac{1}{4}|2\rangle\langle 2|$ on $\mathcal{H} = \mathbb{C}^3$. Determine the dimension of the typical subspace $S_{n,\varepsilon}(\rho)$ for $n = 2$ and arbitrary $0 < \varepsilon < \frac{1}{2}$.

7.6 On the definition of quantum codes: The definition of an (n, R, δ) -quantum code in Definition 7.6 was perhaps surprising. Why did we not simply demand that $F(\mathcal{D}[\mathcal{E}[\rho^{\otimes n}], \rho^{\otimes n}) \geq 1 - \delta$? Argue that such a definition would not correspond to a reliable compression protocol. What is the probability theory analog of this condition?

7.7 Compression and correlations: Let $\rho = \sum_{x \in \Sigma} p(x) \rho_x$, where $p \in P(\Sigma)$ is a probability distribution and ρ_x a state for each $x \in \Sigma$. In class, we showed that if \mathcal{E} and \mathcal{D} are channels such that $F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geq 1 - \delta$ then

$$\sum_{x^n} p(x_1) \cdots p(x_n) F(\mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}], \rho_{x_1} \otimes \cdots \otimes \rho_{x_n}) \geq 1 - \delta.$$

Show that the converse is not necessarily true.

Hint: There are even counterexamples for $n = 1$ and $\delta = 0$.

7.8 Non-monotonicity of the von Neumann entropy: Given a quantum state ρ_{AB} , we write $H(AB)$ for its entropy and $H(A), H(B)$ for the entropies of its reduced states.

- (a) Find a state ρ_{AB} such that $H(AB) > H(B)$.
- (b) Find a state ρ_{AB} such that $H(AB) < H(B)$.

Thus, the von Neumann entropy does *not* satisfy the same monotonicity as the Shannon entropy.

7.9 Subadditivity of the von Neumann entropy: Use Schumacher's theorem to show that, for all states ρ_{AB} ,

$$H(A) + H(B) \geq H(AB),$$

using the same notation as in Exercise 7.8. Thus, the von Neumann entropy is *subadditive*.

Hint: Exercise 4.8.

7.10 Classical-quantum states and concavity: Given a probability distribution $p \in P(\Sigma)$ and states $\rho_x \in D(\mathcal{H})$ for $x \in \Sigma$, we can consider the cq state $\rho_{XB} = \sum_{x \in \Sigma} p(x) |x\rangle\langle x| \otimes \rho_x$ in $D(\mathcal{H}_X \otimes \mathcal{H}_B)$, where $\mathcal{H}_X = \mathbb{C}^\Sigma$ and $\mathcal{H}_B = \mathcal{H}$. See Exercise 5.12.

- (a) Show that $H(XB) = H(p) + \sum_{x \in \Sigma} p(x) H(\rho_x)$.
- (b) Conclude that $H(XB) \geq H(X)$. When does equality hold?
- (c) Show that the von Neumann entropy is a concave function on $D(\mathcal{H})$.

Hint: Evaluate the subadditivity inequality from Exercise 7.9 for a classical-quantum state.

7.11 **Compression without error:** In this problem you will study how well one can compress a quantum state without making any error. Let us say that $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ can be compressed into n qubits if there exists a Hilbert space \mathcal{H}_R of dimension $\leq 2^n$ and quantum channels $\mathcal{E} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_R)$ and $\mathcal{D} \in \mathcal{C}(\mathcal{H}_R, \mathcal{H}_A)$ such that $(\mathcal{D}\mathcal{E} \otimes \mathcal{I}_B)(\sigma_{AB}) = \sigma_{AB}$ for every register B and $\sigma_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $\text{Tr}_B(\sigma_{AB}) = \rho_A$.

(a) Motivate briefly why we do not simply demand that $\mathcal{D}(\mathcal{E}(\rho_A)) = \rho_A$.


For any given state ρ , define the following quantity:

$$N(\rho) := \lceil \log(\text{rank}(\rho)) \rceil = \min\{n \in \mathbb{Z} : n \geq \log(\text{rank}(\rho))\}.$$

- (b) Compute $N(\omega)$, $N(\omega^{\otimes 2})$, and $N(\omega^{\otimes 3})$ for the completely mixed state $\omega = I/3 \in \mathcal{D}(\mathbb{C}^3)$.
(c) Compute $\lim_{n \rightarrow \infty} \frac{1}{n} N(\rho^{\otimes n})$ for a general state $\rho \in \mathcal{D}(\mathcal{H}_A)$ as some simple function of the state.

It turns out that $N(\rho_A)$ corresponds to the *minimal* number of qubits that ρ_A can be compressed into according to the above definition. To prove this, show the following facts:

- (d) Show that if $n \geq \log(\text{rank}(\rho_A))$ then ρ_A can be compressed into n qubits.
(e) Show that if ρ_A can be compressed into n qubits then $n \geq \log(\text{rank}(\rho_A))$.

7.12  **Practice:** In this problem, you can explore the properties of typical subspaces. Consider the qubit state $\rho = \frac{1}{2} |0\rangle\langle 0| + \frac{1}{2} |+\rangle\langle +|$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

- (a) Compute the largest eigenvalue λ as well as the von Neumann entropy $H(\rho)$ of ρ .
(b) Plot the following functions of $k \in \{0, 1, \dots, n\}$ for $n = 100$ as well as for $n = 1000$:

$$d(k) = \binom{n}{k}, \quad r(k) = \frac{1}{n} \log \binom{n}{k}, \quad q(k) = \binom{n}{k} \lambda^k (1 - \lambda)^{n-k}$$

- (c) Plot the following functions of $n \in \{1, \dots, 1000\}$ for $\varepsilon = 0.1$ as well as for $\varepsilon = 0.01$:

$$r(n) = \frac{1}{n} \log \dim S_{n,\varepsilon}, \quad p(n) = \text{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}],$$

where $\Pi_{n,\varepsilon}$ denotes the orthogonal projection onto the typical subspace $S_{n,\varepsilon}$ of ρ .

Lecture 8

Entropy and subsystems

Last week we discussed the definition of the von Neumann entropy, which generalizes the Shannon entropy to quantum states, as well as the problem of compressing quantum information. The main result was Schumacher's theorem (Theorem 7.10), which states that the von Neumann entropy is the 'optimal' compression rate.

Today we will discuss how the entropies of subsystems are related to the entropy of the overall system. As you already saw in last week's Exercise 7.9, these entropies are not independent but constrained by *entropy inequalities*, and we will discuss several of those. Then we will introduce the *mutual information*, which is a very useful correlation measure, and discuss its mathematical properties.

8.1 Entropies of subsystems

To study the entropies of subsystems, it is useful to first introduce some notation.

Definition 8.1 (Entropy of subsystems). Given a quantum state ρ_{AB} , we define

$$H(AB)_\rho := H(\rho_{AB}), \quad H(A)_\rho := H(\rho_A), \quad H(B)_\rho := H(\rho_B).$$

We use analogous notation for more than two subsystems. We will very often leave out the subscript and write $H(AB)$, $H(A)$, $H(B)$ when the state is clear. In fact, we already introduced and used this convention in Exercise 7.8, as well as for the Shannon entropy (Definition 6.4).

How are these entropies related? Let us first consider two very extreme cases:

- If ρ_{AB} is pure then $H(AB) = 0$ and

$$H(A) = H(B). \tag{8.1}$$

The latter is often called the *entanglement entropy* of ρ_{AB} .

Proof. The former holds because the eigenvalues of a pure state are $1, 0, \dots, 0$. The latter follows from the Schmidt decomposition, which implies that ρ_A and ρ_B have the same nonzero eigenvalues (Corollary 2.21). \square

Definition 8.2 (Entanglement entropy). For a pure state, $H(A) = H(B)$ is called the *entanglement entropy*.

Note that Eq. (8.1) generalizes also to *pure* multi-partite states. For example, $H(A) = H(BC)$, $H(B) = H(AC)$, and $H(C) = H(AB)$ for any pure state on ABC . We will come back to entanglement entropy in Lecture 12 and derive its operational interpretation in Theorem 12.5.

- If ρ_{AB} is a product state (equivalently, $\rho_{AB} = \rho_A \otimes \rho_B$) then $H(AB) = H(A) + H(B)$. We say that the entropy is *additive* with respect to tensor products.

Proof. If ρ_A has eigenvalues $(p_i)_{i=1}^{d_A}$ and ρ_B has eigenvalues $(q_j)_{j=1}^{d_B}$ then $\rho_{AB} = \rho_A \otimes \rho_B$ has eigenvalues $(p_i q_j)_{i,j}$. Thus,

$$\begin{aligned} H(AB) &= \sum_{i,j} p_i q_j \log \frac{1}{p_i q_j} = \sum_{i,j} p_i q_j \log \frac{1}{p_i} + \sum_{i,j} p_i q_j \log \frac{1}{q_j} \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} = H(A) + H(B). \quad \square \end{aligned}$$

Next, we list some general properties. We first discuss the extent to which the subadditivity and monotonicity properties of the Shannon entropy (see Lemma 6.5) generalize to the quantum case.

- *Subadditivity:*

$$H(A) + H(B) \geq H(AB). \quad (8.2)$$

Moreover, equality holds if *and only if* $\rho_{AB} = \rho_A \otimes \rho_B$. The term “subadditivity” means that when the two systems are combined or “added” together, the entropy of the joint system is generally smaller (“sub”) than the sum of the entropies of the two parts.

You proved this inequality in Exercise 7.9 and we discussed above that equality holds for product states. Why does equality hold *only* for product states? We will prove this next week.

- The von Neumann entropy is *not* monotonic. That is, in general, $H(AB) \not\geq H(A)$ and $H(AB) \not\geq H(B)$. You discussed this in Exercise 7.8.
- However, for classical-quantum states ρ_{XB} we do have the monotonicity inequalities

$$H(XB) \geq H(X) \quad \text{and} \quad H(XB) \geq H(B). \quad (8.3)$$

You proved the first inequality in Exercise 7.10; the second will follow from Lemma 9.3.

- *Araki-Lieb (or triangle) inequality:*

$$H(AB) \geq |H(A) - H(B)|. \quad (8.4)$$

We can think of Eq. (8.4) as a weaker form of monotonicity (not to be confused with Eq. (8.6) below). Indeed, if $H(AB) \geq H(A)$ and $H(AB) \geq H(B)$ were true then these would imply Eq. (8.4).

Proof. Choose any purification ρ_{ABC} of ρ_{AB} . Then:

$$H(AB) = H(C) \geq H(BC) - H(B) = H(A) - H(B),$$

where the first and last step hold since ρ_{ABC} is pure [Eq. (8.1)] and the inequality is subadditivity [Eq. (8.2)]. Likewise,

$$H(AB) = H(C) \geq H(AC) - H(A) = H(B) - H(A),$$

which proves the other half of Eq. (8.4). \square

It turns out that there is a stronger variant of the subadditivity inequality which is very powerful:

- *Strong subadditivity*: For all ρ_{ABC} , it holds that

$$H(AC) + H(BC) \geq H(ABC) + H(C). \quad (8.5)$$

Clearly, this inequality reduces to Eq. (8.2) if there is no C system, which justifies the terminology. Eq. (8.5) is much harder to prove than Eq. (8.2) and we will not have time to do this in the lecture (cf. the closely related monotonicity property of the quantum relative entropy [Eq. (9.6)] that we will discuss in Lecture 9).

- *Weak monotonicity*: For all ρ_{ABC} , it holds that

$$H(AC) + H(BC) \geq H(A) + H(B). \quad (8.6)$$

This inequality follows from Eq. (8.5) by using a purification – in the same way that Eq. (8.4) follows from Eq. (8.2) – as you get to prove in Exercise 8.3. The name is justified since if $H(AC) \geq H(A)$ and $H(BC) \geq H(B)$ were true then these would imply Eq. (8.6).

8.2 Mutual information

In this section we will discuss the mutual information, which is a useful way to quantify correlations in quantum states.

Definition 8.3 (Mutual information). The *mutual information* of a quantum state ρ_{AB} is defined as

$$I(A : B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho. \quad (8.7)$$

As for individual entropies, we will mostly leave the subscript out and write $I(A : B)$ if the state is clear.

We can use the same formula to define the mutual information $I(X : Y)_p$ of a joint probability distribution. These definitions are of course compatible: If $\rho_{XY} = \sum_{x,y} p(x,y) |x,y\rangle\langle x,y|$ is the classical state corresponding to a joint distribution $p(x,y)$ then $I(X : Y)_\rho = I(X : Y)_p$.

Example 8.4. Let us revisit the three states from Table 2.1 and compute their mutual information.

- *Maximally entangled state*:

$$\rho_{AB} = |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+| = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

where $|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Note that $H(A) = H(B) = 1$ since $\rho_A = \rho_B = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ are maximally mixed, and $H(AB) = 0$ since ρ_{AB} is pure. Hence, $I(A : B) = 1 + 1 - 0 = 2$.

- *Maximally correlated classical state:*

$$\rho_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|)_{AB} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which corresponds to a pair of perfectly correlated uniformly random bits. In this case $\rho_A = \rho_B = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ are also maximally mixed, so again $H(A) = H(B) = 1$. However, this time ρ_{AB} is not pure – its eigenvalues are $(\frac{1}{2}, 0, 0, \frac{1}{2})$, which we easily see because ρ_{AB} is diagonal, so $H(AB) = 1$. Hence, $I(A : B) = 1 + 1 - 1 = 1$.

- *Two independent uniformly random bits:*

$$\rho_{AB} = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)_A \otimes \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)_B = \frac{1}{4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Since $\rho_{AB} = \rho_A \otimes \rho_B$ is a product state, its entropy is additive: $H(AB) = H(A) + H(B)$, which immediately implies that $I(A : B) = H(A) + H(B) - H(AB) = 0$. Alternatively, we again observe that $\rho_A = \rho_B = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ so $H(A) = H(B) = 1$. However, this time the eigenvalues of ρ_{AB} are $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, which is easily seen since ρ_{AB} is diagonal. This constitutes a uniform distribution on two bits so $H(AB) = 2$. Hence, $I(A : B) = 1 + 1 - 2 = 0$.

We now list some useful properties, several of which follow directly from the results of Section 8.1:

- *Nonnegativity:* $I(A : B) \geq 0$. Moreover, $I(A : B) = 0$ if and only if ρ_{AB} is a product state (i.e., $\rho_{AB} = \rho_A \otimes \rho_B$). This is a first indication that the mutual information is a useful correlation measure.

Proof. This is simply a restatement of subadditivity [Eq. (8.2)], including the condition for equality. \square

- *Invariance under isometries:* For any state ρ_{AB} and isometries $V_{A \rightarrow A'}$, $W_{B \rightarrow B'}$, we have

$$I(A : B)_\rho = I(A' : B')_{\sigma},$$

where $\sigma_{A'B'} := (V_{A \rightarrow A'} \otimes W_{B \rightarrow B'})\rho_{AB}(V_{A \rightarrow A'}^\dagger \otimes W_{B \rightarrow B'}^\dagger)$.

Proof. This follows from the invariance of the von Neumann entropy under isometries (see Lemma 7.2) once we recognize that $\sigma_{A'} = V\rho_A V^\dagger$ and $\sigma_{B'} = W\rho_B W^\dagger$. \square

- *Pure states:* If ρ_{AB} is pure then $I(A : B) = 2H(A) = 2H(B)$.

Proof. Recall that $H(AB) = 0$ and $H(A) = H(B)$ if ρ_{AB} is pure. \square

- *Upper bound:* Let $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$. Then,

$$I(A : B) \leq 2 \min \{H(A), H(B)\} \leq 2 \log \min \{d_A, d_B\}. \quad (8.8)$$

For classical-quantum states ρ_{XB} , we have the stronger upper bound

$$I(X : B) \leq \min \{H(X), H(B)\} \leq \log \min \{d_X, d_B\}. \quad (8.9)$$

In particular, Eq. (8.9) holds for classical states and joint probability distributions. In Exercises 8.6 and 8.7 you will investigate under which conditions the upper bounds in Eqs. (8.8) and (8.9) hold with equality.

Proof. The first inequality in Eq. (8.8) follows from the Araki-Lieb inequality [Eq. (8.4)]. Indeed, $H(A) + H(B) - H(AB) = I(A : B) \leq 2H(A)$ is equivalent to $H(AB) \geq H(B) - H(A)$, and similarly for the other bound.

Similarly, the first bound in Eq. (8.9) is equivalent to the monotonicity inequalities in Eq. (8.3). \square

- *Monotonicity:* For all ρ_{ACE} ,

$$I(A : CE) \geq I(A : C). \quad (8.10)$$

(We label the subsystems ACE rather than ABC to avoid confusion in the below.)

Proof. This is simply a rewriting of strong subadditivity [Eq. (8.5)]. \square

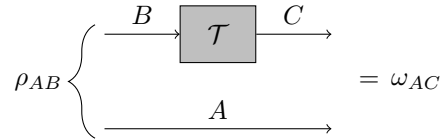
The latter property is equivalent to the following general result:

Lemma 8.5 (Data processing inequality). *Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a state, $\mathcal{T}_{B \rightarrow C} \in \mathcal{C}(\mathcal{H}_B, \mathcal{H}_C)$ a channel, and $\omega_{AC} = (\mathcal{I}_A \otimes \mathcal{T}_{B \rightarrow C})[\rho_{AB}]$. Then,*

$$I(A : B)_\rho \geq I(A : C)_\omega.$$

By symmetry, a similar inequality holds if we apply a channel on A rather than on B.

The data processing inequality is very intuitive, as it states that we can never increase correlations by acting locally. The following figure illustrates the situation:



Clearly, Lemma 8.5 reduces to the monotonicity property of the mutual information (simply choose $B = CE$ and $\mathcal{T} = \text{Tr}_E$).

Proof of Lemma 8.5. Any channel has a Stinespring representation $\mathcal{T}_{B \rightarrow C}[M_B] = \text{Tr}_E[V M_B V^\dagger]$, where $V = V_{B \rightarrow CE} \in \mathcal{L}(\mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_E)$ is an isometry [Eq. (5.7) and Lemma 5.5]. Note that

$$\omega_{ACE} = (I_A \otimes V_{B \rightarrow CE}) \rho_{AB} (I_A \otimes V_{B \rightarrow CE})^\dagger$$

is an extension of ω_{AC} (i.e., $\text{Tr}_E[\omega_{ACE}] = \omega_{AC}$). As a consequence,

$$I(A : B)_\rho = I(A : CE)_\omega \geq I(A : C)_\omega,$$

using that the mutual information is invariant under isometries and monotonic. \square

Next week we will discuss a nice application of the data processing inequality known as Holevo's Theorem (Theorem 9.4). It introduces a quantity that characterizes how much classical information can be extracted from a quantum state. The same quantity also turns out to capture the rate at which classical information can be transmitted through a quantum channel.

8.3 Exercises

8.1 **Computing entropy and mutual information:** Let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$ and consider the following two-qubit state given with respect to the basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$:

$$\rho_{AB} = \frac{1}{10} \begin{pmatrix} 4 & 0 & 0 & 4 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 4 & 0 & 0 & 4 \end{pmatrix}.$$

- (a) Is ρ_{AB} pure or mixed?
- (b) Compute the entropy $H(AB)$.

Consider the three-qubit state $|\Gamma_{ABC}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$.

- (c) Compute the entropies $H(A)$, $H(B)$, and $H(C)$.
- (d) Compute the mutual information $I(A : B)$.
- (e) Give an example of a state ρ_{AB} which is separable and for which $I(A : B)_\rho > 0$. Justify your answer.

8.2 **Mutual information upper bound:** Eq. (8.8) implies that $I(A : B) \leq \log d_A + \log d_B$, where $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$. Give a simple proof of this fact without using the Araki-Lieb inequality.

8.3 **Weak monotonicity:** Use a purification to deduce the weak monotonicity inequality $H(AC) + H(BC) \geq H(A) + H(B)$ from the strong subadditivity inequality, and vice versa.

8.4 **Strict concavity of the von Neumann entropy:** In Exercise 7.10 you proved that $H(\rho)$ is a concave function of $\rho \in D(\mathcal{H})$. Revisit your proof and show that it is strictly concave using the equality condition for the subadditivity inequality discussed today.

8.5 **Equality condition for monotonicity:** In Exercise 6.6, you proved that the Shannon entropy (unlike the von Neumann entropy) satisfies the following monotonicity inequality: $H(XY) \geq H(X)$ for any probability distribution p_{XY} . Show that equality holds if and only if $p_{XY}(x, y) = p_X(x)\delta_{f(x), y}$ for a function $f : \Sigma_X \rightarrow \Sigma_Y$.

This means that $Y = f(X)$, i.e., the second random variable is a function of the first!

8.6 **Classical mutual information:** From Eq. (8.9), we know that $I(X : Y) \leq \log d$ for every distribution $p_{XY} \in P(\Sigma_X \times \Sigma_Y)$ with $|\Sigma_X| = |\Sigma_Y| = d$. Show that $I(X : Y) = \log d$ if and only if $p_{XY}(x, y) = \frac{1}{d}\delta_{f(x), y}$ for a bijection $f : \Sigma_X \rightarrow \Sigma_Y$. Such probability distributions p_{XY} are called *maximally correlated*.

Hint: Exercise 8.5.

8.7 **Quantum mutual information:** From class, we know that $I(A : B) \leq 2 \log d$ for every state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^d$. Show that $I(A : B) = 2 \log d$ if and only if is maximally entangled (Definition 3.5).

Hint: Inspect your solution to Exercise 8.2.

8.8 **Entropic uncertainty relation:** Here you can prove another uncertainty relation. Let $\rho \in D(\mathbb{C}^2)$ and denote by p_{Std} and p_{Had} the probability distributions of outcomes when measuring ρ in the standard basis and Hadamard basis, respectively. You will show:

$$H(p_{\text{Std}}) + H(p_{\text{Had}}) \geq H(\rho) + 1 \quad (8.11)$$

- (a) Why is it appropriate to call (8.11) an *uncertainty relation*?
(b) Find a state ρ for which the uncertainty relation is saturated (i.e., an equality).

To start, recall the Pauli matrices $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

- (c) Verify that $\frac{1}{2}(\rho + Z\rho Z) = \begin{pmatrix} \langle 0|\rho|0\rangle & 0 \\ 0 & \langle 1|\rho|1\rangle \end{pmatrix}$ and deduce that $H(p_{\text{Std}}) = H(\frac{1}{2}(\rho + Z\rho Z))$.
(d) Show that, similarly, $H(p_{\text{Had}}) = H(\frac{1}{2}(\rho + X\rho X))$. *Hint: $|\pm\rangle$ is the eigenbasis of X .*

Now consider the following three-qubit state,


$$\omega_{ABC} = \frac{1}{4} \sum_{a=0}^1 \sum_{b=0}^1 |a\rangle\langle a|_A \otimes |b\rangle\langle b|_B \otimes (X^a Z^b \rho Z^b X^a)_C,$$

where we denote $X^0 = I$, $X^1 = X$, $Z^0 = I$, $Z^1 = Z$. Note that subsystems A & B are classical.

- (e) Show that $H(ABC) = 2 + H(\rho)$. Use parts (c) and (d) to verify that $H(AC) = 1 + H(p_{\text{Std}})$, $H(BC) = 1 + H(p_{\text{Had}})$, and $H(C) = 1$ in state ω_{ABC} .

Hint: Use the formula for the entropy of classical-quantum states from Exercise 7.10.

- (f) Use part (e) and the strong subadditivity inequality to deduce (8.11).

8.9  **Practice:** Let $|\Psi_{ABC}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C = \mathbb{C}^5 \otimes \mathbb{C}^2 \otimes \mathbb{C}^3$ be a unit vector given by

$$|\Psi_{ABC}\rangle = \sum_{k=0}^4 c_k |k\rangle_A \otimes |\beta_k\rangle_B \otimes |\gamma_k\rangle_C$$

where $(c_0, c_1, c_2, c_3, c_4) = (0.3, 0.4, 0.5, 0.5, 0.5)$ and

$$|\beta_k\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + \exp\left(\frac{2\pi i}{5}k\right) |1\rangle \right), \quad |\gamma_k\rangle = |k \bmod 3\rangle.$$

- (a) Compute the mutual information $I(A : B)$ of $\rho_{AB} = \text{Tr}_C[|\Psi\rangle\langle\Psi|_{ABC}]$.
(b) Let $\Phi \in \mathcal{C}(\mathcal{H}_B)$ be a quantum channel that acts on any $M \in \mathcal{L}(\mathcal{H}_B)$ as

$$\Phi(M) = \frac{1}{2}M + \frac{1}{2}XMX$$

where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Compute the mutual information $I(A : B)$ of $\sigma_{AB} = (\mathcal{I}_A \otimes \Phi_B)[\rho_{AB}]$.

8.10 **Entanglement measure:** In this problem we consider the following quantity which is called the *squashed entanglement* of a state ρ_{AB} :

$$E(A : B)_\rho := \inf_{\substack{\mathcal{H}_R, \omega_{ABR} \text{ s.t.} \\ \text{Tr}_R[\omega_{ABR}] = \rho_{AB}}} \frac{1}{2} \left(I(A : BR)_\omega - I(A : R)_\omega \right).$$

The infimum is over arbitrary finite-dimensional Hilbert spaces \mathcal{H}_R and quantum states ω_{ABR} such that $\text{Tr}_R[\omega_{ABR}] = \rho_{AB}$. We write $E(A : B)$ if the state is clear from the context. Your task is to prove a number of properties which indicate that $E(A : B)$ is a useful measure of the entanglement between systems A and B in state ρ_{AB} .

- (a) Show that $E(A : B) \geq 0$ for all states ρ_{AB} .
(b) Show that $E(A : B) \leq H(A)_\rho$ for all states ρ_{AB} .

- (c) Show that if $\sigma_{AB'}$ is obtained by applying a channel $B \rightarrow B'$ to ρ_{AB} then $E(A : B')_\sigma \leq E(A : B)_\rho$.
- (d) Show that $E(A : B) + E(A : C) \leq E(A : BC)$ for all states ρ_{ABC} .
- (e) Show that if ρ_{AB} is pure then $E(A : B) = H(A)$.
- (f) Show that if ρ_{AB} is separable then $E(A : B) = 0$.
Hint: Any separable state can be written in the form $\rho_{AB} = \sum_i p_i \rho_{A,i} \otimes \rho_{B,i}$ where $\rho_{A,i}$ and $\rho_{B,i}$ are pure. Can you find an extension ω_{ABR} of ρ_{AB} that is classical on R ?

Lecture 9

Holevo bound and relative entropy

Last week we discussed various entropic quantities in the quantum case and inequalities between them. In particular, for a multipartite state ρ_{ABC} one can consider the entropies of the reduced states (e.g., $H(AB) = H(\rho_{AB})$ where $\rho_{AB} = \text{Tr}_C[\rho_{ABC}]$) and the inequalities among them, such as the strong subadditivity:

$$H(AB) + H(BC) \geq H(ABC) + H(B).$$

Strong subadditivity is equivalent to the monotonicity of the mutual information $I(A : B) = H(A) + H(B) - H(AB)$, namely

$$I(A : B) \leq I(A : BC).$$

This in turn was equivalent to the data processing inequality for the mutual information:

$$I(A : B)_\rho \geq I(A : C)_\omega \quad (9.1)$$

where $\omega_{AC} = (\mathcal{I}_A \otimes \mathcal{T}_{B \rightarrow C})[\rho_{AB}]$ is obtained by applying a channel $\mathcal{T}_{B \rightarrow C}$ on the B system of ρ_{AB} . Intuitively, processing quantum information locally can only decrease the mutual information.

9.1 Holevo bound

Assume we draw an element $x \in \Sigma$ with probability p_x , record its value in a classical system X with Hilbert space $\mathcal{H}_X = \mathbb{C}^\Sigma$, and create an arbitrary state $\rho_x \in \mathcal{D}(\mathcal{H}_B)$ associated to x in a separate system B . Then the resulting classical-quantum (cq) state

$$\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \in \mathcal{D}(\mathcal{H}_X \otimes \mathcal{H}_B) \quad (9.2)$$

represents the ensemble $\{p_x, \rho_x\}$ (see Exercise 5.12). To such an ensemble, or the corresponding cq-state, we associate the so-called Holevo χ -quantity.

Definition 9.1 (Holevo χ -quantity). The *Holevo χ -quantity* of an ensemble $\{p_x, \rho_x\}$ is

$$\chi(\{p_x, \rho_x\}) := I(X : B) = H\left(\sum_x p_x \rho_x\right) - \sum_x p_x H(\rho_x),$$

where the mutual information is computed in the cq-state $\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$.

To verify the second equality, use that $H(XB) = H(p) + \sum_x p_x H(\rho_x)$, as you proved in Exercise 7.10 (a). The following lemma states upper and lower bounds on the Holevo quantity.

Lemma 9.2. For any ensemble $\{p_x, \rho_x\}$ of states $\rho_x \in \mathcal{D}(\mathcal{H}_B)$, we have

$$0 \leq \chi(\{p_x, \rho_x\}) \leq H\left(\sum_x p_x \rho_x\right) \leq \log \dim \mathcal{H}_B.$$

Proof. The lower bound on the Holevo quantity holds since the mutual information is nonnegative. The first upper bound follows by leaving out a nonnegative term (or by observing that it is equivalent to $H(XB) \geq H(X)$, which you proved in Exercise 7.10). The second one is clear. \square

We also have the following upper bound, which you will prove in Exercise 9.3.

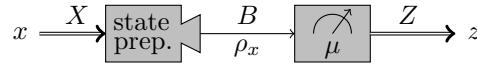
Lemma 9.3. For any ensemble $\{p_x, \rho_x\}$ of states $\rho_x \in \mathcal{D}(\mathcal{H}_B)$, we have

$$\chi(\{p_x, \rho_x\}) \leq H(p).$$

Moreover, equality holds if and only if the states ρ_x with $p_x > 0$ have pairwise orthogonal image.

In terms of the cq-state corresponding to the ensemble, the upper bound in Lemma 9.3 can also be written as $H(XB) \geq H(B)$. This confirms our claim in Eq. (8.3).

Why is the Holevo χ -quantity useful? For this, let us revisit the following fundamental question: How much information can Alice communicate to Bob by sending a quantum state? We will consider the following setup:



Here, Alice has a classical message $x \in \Sigma$ with distribution $p \in \mathcal{P}(\Sigma)$ which she would like to communicate to Bob. For this, she sends Bob a quantum state $\rho_x \in \mathcal{D}(\mathcal{H}_B)$, and Bob applies a measurement $\mu: \Gamma \rightarrow \text{PSD}(\mathcal{H}_B)$. Using Born's rule, we see that the joint distribution of Alice's random message X and Bob's random measurement outcome Z on the set $\Sigma \times \Gamma$ is given by

$$p(x, z) = p(x) \text{Tr}[\rho_x \mu(z)].$$

In Exercise 2.3, you proved the so-called Nayak bound: if $x \in \{0, 1\}^m$ is chosen uniformly at random and $\mathcal{H}_B = (\mathbb{C}^2)^{\otimes n}$ then $\Pr(X = Z) \leq 2^{n-m}$, i.e., we need to send $n \geq m$ qubits to communicate m bits reliably. But how about if the distribution of x is not uniform?

It turns out that there is a general useful bound on the mutual information $I(X : Z)$ between Alice's message and Bob's measurement result. This is the content of the following theorem:

Theorem 9.4 (Holevo). $I(X : Z) \leq \chi(\{p_x, \rho_x\})$ for any ensemble $\{p_x, \rho_x\}$ and measurement μ .

Holevo's theorem is a simple consequence of the data processing inequality (Lemma 8.5), which in turn relies on the very nontrivial strong subadditivity inequality.

Proof. Let ρ_{XB} be the cq-state from Eq. (9.2) that represents the ensemble $\{p_x, \rho_x\}$, let $\mu: \Gamma \rightarrow \text{PSD}(\mathcal{H}_B)$ be an arbitrary measurement on system B , and let $\Phi_{B \rightarrow Z}[\sigma] := \sum_{z \in \Gamma} \text{Tr}[\sigma \mu(z)] |z\rangle\langle z|$, with output space \mathcal{H}_Z , be the quantum channel corresponding to μ . Then by the data processing inequality for the mutual information, Eq. (9.1),

$$\chi(\{p_x, \rho_x\}) = I(X : B)_\rho \geq I(X : Z)_\omega$$

where

$$\omega_{XZ} = (\mathcal{I}_X \otimes \Phi_{B \rightarrow Z})[\rho_{XB}] = \sum_{x \in \Sigma} p_x |x\rangle\langle x| \otimes \Phi[\rho_x] = \sum_{x \in \Sigma, z \in \Gamma} p(x, z) |x\rangle\langle x| \otimes |z\rangle\langle z|$$

is the resulting output state after the measurement. That was easy! \square

Here is a concrete consequence of the Holevo bound. In the situation above, when can Bob exactly recover Alice's message? That is, when is $X = f(Z)$ for some function f ? We know from Exercise 8.5 that this is the case precisely when $H(XZ) = H(Z)$, that is, when $I(X : Z) = H(X)$. But $I(X : Z) \leq \chi(\{p_x, \rho_x\}) \leq H(p) = H(X)$ by the Holevo bound and Lemma 9.3. Thus, Bob can exactly recover Alice's message *only* if $\chi(\{p_x, \rho_x\}) = H(p)$. (In Exercise 9.4 you can prove that this is also sufficient.) But note that Lemma 9.3 also asserts that in this case the states ρ_x with $p_x > 0$ have pairwise orthogonal image. Accordingly, $\dim H_B \geq |\{x \in \Sigma : p_x > 0\}|$. This means that if we want to communicate m bits perfectly, we need to send at least m qubits, even if the distribution of the messages is not uniform!

Remark 9.5. The above considerations are closely related to one of the most fundamental problems in quantum information theory: Given access to a quantum channel $\mathcal{N}_{A \rightarrow B}$ (which could, e.g., describe an optical fiber), what is the optimal rate at which we can use it to communicate classical information? This rate is called the *classical capacity* of the channel. The Holevo-Schumacher-Westmoreland theorem computes this capacity in terms of the Holevo quantity. To state this result, note that for any ensemble of input states $\{p_x, \rho_{A,x}\}$ we get an ensemble of output states $\{p_x, \sigma_{B,x}\}$, where $\sigma_{B,x} := \mathcal{N}_{A \rightarrow B}[\rho_{A,x}]$. Let $\chi(\mathcal{N}_{A \rightarrow B})$ denote the supremum of $\chi(\{p_x, \sigma_{B,x}\})$ over all ensembles obtained in this way. Then the classical capacity of $\mathcal{N}_{A \rightarrow B}$ is given by $\lim_{n \rightarrow \infty} \frac{1}{n} \chi(\mathcal{N}_{A \rightarrow B}^{\otimes n})$. Proving this result is out of scope for this introductory lecture, but you can consult the books by Watrous or Wilde for details.

9.2 Relative entropy

Quantum relative entropy is a useful mathematical tool for analyzing the von Neumann entropy. Let us first consider its classical version (also known as *Kullback-Leibler divergence*).

Definition 9.6 (Relative entropy). Let $p, q \in \mathcal{P}(\Sigma)$ be probability distributions. The *relative entropy* of p with respect to q is

$$D(p||q) = \begin{cases} \sum_{x \in \Sigma} p(x) \log \frac{p(x)}{q(x)} & \text{if } \{x : q(x) = 0\} \subseteq \{x : p(x) = 0\}, \\ \infty & \text{otherwise.} \end{cases} \quad (9.3)$$

To make sense of the expression $p(x) \log \frac{p(x)}{q(x)}$ for all possible values of $p(x), q(x) \in [0, 1]$, recall from p. 75 that $\lim_{a \rightarrow 0} a \log a = 0$. So the expression is equal to 0 whenever $p(x) = 0$, and has a finite non-zero value when both $p(x) > 0$ and $q(x) > 0$. The only problematic case is when $p(x) > 0$ but $q(x) = 0$, in which case the value becomes infinite.

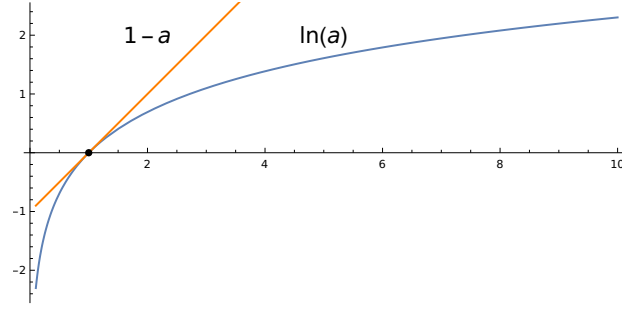
$p(x)$	$q(x)$	$p(x) \log \frac{p(x)}{q(x)}$
= 0	= 0	0
= 0	> 0	0
> 0	= 0	∞
> 0	> 0	finite

The condition for when $D(p||q)$ in Eq. (9.3) is finite can also be stated as $\forall x : q(x) = 0 \Rightarrow p(x) = 0$ or equivalently as $\forall x : p(x) > 0 \Rightarrow q(x) > 0$.

Here are some basic properties of the relative entropy:

- *Nonnegativity:* $D(p||q) \geq 0$, with equality iff $p = q$.

Proof. Without loss of generality, we assume that $p(x) \geq 0$ for all x . Note that $\ln a \leq a - 1$, with equality iff $a = 1$.



Since $\log a = \frac{\ln a}{\ln 2}$,

$$\begin{aligned} D(p\|q) &= \sum_x p(x) \left(-\log \frac{q(x)}{p(x)} \right) \\ &\geq \frac{1}{\ln 2} \sum_x p(x) \left(1 - \frac{q(x)}{p(x)} \right) \\ &= \frac{1}{\ln 2} \left(\sum_x p(x) - \sum_x q(x) \right) = 0, \end{aligned}$$

completing the proof. \square

In statistics, functions with this property are known as *divergences*. A divergence is a weaker notion than a distance since it does not need to be symmetric or obey the triangle inequality. For example, the cost of a plane ticket from one destination to another is generally a divergence but not a distance.

- Note that $D(p\|q)$ is *not* symmetric, i.e., generally $D(p\|q) \neq D(q\|p)$.

Let us consider two simple applications of the classical relative entropy. Let $p, q \in \mathcal{P}(\Sigma)$ be probability distributions and assume that $q(x) = 1/|\Sigma|$ is uniform. Then

$$\begin{aligned} D(p\|q) &= \sum_{x \in \Sigma} p(x) \log p(x) - \sum_{x \in \Sigma} p(x) \log q(x) \\ &= -H(p) - \log \frac{1}{|\Sigma|}, \end{aligned}$$

implying that $H(p) \leq \log |\Sigma|$, with equality if and only if p is uniform. We proved this already in Item (b) of Lemma 6.3.

As another application, let $p_{XY} \in \mathcal{P}(\Sigma \times \Gamma)$ be an arbitrary distribution and let $q_{XY}(x, y) = p_X(x)p_Y(y)$ be the product of its marginals (recall that the marginals are obtained by summing over the remaining indices: $p_X(x) = \sum_{y \in \Sigma} p_{XY}(x, y)$ and $p_Y(y) = \sum_{x \in \Sigma} p_{XY}(x, y)$). Then

$$\begin{aligned} D(p_{XY}\|q_{XY}) &= -H(p_{XY}) - \sum_{x \in \Sigma} \sum_{y \in \Gamma} p_{XY}(x, y) \log(p_X(x)p_Y(y)) \\ &= -H(p_{XY}) - \sum_{x \in \Sigma} p_X(x) \log p_X(x) - \sum_{y \in \Sigma} p_Y(y) \log p_Y(y) \\ &= H(p_X) + H(p_Y) - H(p_{XY}) \\ &= I(X : Y)_{p_{XY}}, \end{aligned}$$

implying that $I(X : Y)_{p_{XY}} \geq 0$, with equality if and only if p_{XY} is a product distribution, i.e., $p_{XY}(x, y) = p_X(x)p_Y(y)$ (that is, X and Y are independent).

9.3 Quantum relative entropy

Now that we are familiar with the classical relative entropy, we can define the quantum version by noting that $p(x) \log \frac{p(x)}{q(x)} = p(x) \log p(x) - p(x) \log q(x)$ and replacing probability distributions by density matrices.

Definition 9.7 (Quantum relative entropy). Let $\rho, \sigma \in D(\mathcal{H})$ be quantum states. The *quantum relative entropy* of ρ with respect to σ is

$$D(\rho \parallel \sigma) = \begin{cases} \text{Tr}[\rho \log \rho] - \text{Tr}[\rho \log \sigma] & \text{if } \ker \sigma \subseteq \ker \rho, \\ \infty & \text{otherwise.} \end{cases}$$

The interpretation here is similar to the classical case. Note that the first term is equal to $-H(\rho)$ so we only need to make sense of $\rho \log \sigma$ in the second term. It is unambiguous how $\rho \log \sigma$ acts on $(\ker \sigma)^\perp$ since $\log \sigma$ there is well-defined. Assuming $\ker \sigma \subseteq \ker \rho$, we can define $\rho \log \sigma$ as zero on $\ker \sigma$. If this condition is not met, the expression becomes infinite just like in the classical case. Note that the condition $\ker \sigma \subseteq \ker \rho$ is equivalent to $\text{im } \rho \subseteq \text{im } \sigma$. For example, $D(|0\rangle\langle 0| \parallel |+\rangle\langle +|) = \infty$ since $\text{span}\{|0\rangle\} \not\subseteq \text{span}\{|+\rangle\}$.

Here is a list of various properties of quantum relative entropy:

- *Classical states:* If $\rho = \sum_x p(x) |x\rangle\langle x|$ and $\sigma = \sum_x q(x) |x\rangle\langle x|$ then $D(\rho \parallel \sigma) = D(p \parallel q)$.
- *Monotonicity:* For any $\rho, \sigma \in D(\mathcal{H})$ and any $\Phi \in C(\mathcal{H}, \mathcal{H}')$:

$$D(\rho \parallel \sigma) \geq D(\Phi[\rho] \parallel \Phi[\sigma]). \quad (9.4)$$

This property is very important and could well be called the “fundamental theorem of quantum information theory” (it even implies the strong subadditivity inequality as we will discuss below). Unfortunately, we will not have time to prove since it would require a separate lecture (see p. 280 of Watrous’ book).

- *Nonnegativity (Klein’s inequality):*

$$D(\rho \parallel \sigma) \geq 0, \quad (9.5)$$

with equality iff $\rho = \sigma$.

Proof. Let $\mu: \Omega \rightarrow \text{PSD}(\mathcal{H})$ be a quantum measurement and denote by $\Phi \in C(\mathcal{H}, \mathcal{X})$ where $\mathcal{X} = \mathbb{C}^\Omega$ the quantum channel

$$\Phi[\omega] := \sum_{x \in \Omega} \text{Tr}[\mu(x)\omega] |x\rangle\langle x|$$

that implements the measurement μ . Denote by p and q the probability distributions resulting from measuring ρ and σ , respectively:

$$p(x) := \text{Tr}[\mu(x)\rho], \quad q(x) := \text{Tr}[\mu(x)\sigma].$$

Note that

$$\Phi[\rho] = \sum_{x \in \Omega} p(x) |x\rangle\langle x|, \quad \Phi[\sigma] = \sum_{x \in \Omega} q(x) |x\rangle\langle x|$$

are diagonal. By monotonicity,

$$D(\rho\|\sigma) \geq D(\Phi[\rho]\|\Phi[\sigma]) = D(p\|q) \geq 0,$$

where we used the fact that the output states $\Phi[\rho]$ and $\Phi[\sigma]$ are diagonal to reduce to the classical nonnegativity inequality. For the equality condition, note that if $\rho = \sigma$ then $D(\rho\|\sigma) = 0$. To prove the converse, we can use Exercise 7.4 which shows that, for any $\rho, \sigma \in \mathcal{D}(\mathcal{H})$, there exists a measurement whose output distributions p and q on the two states satisfy $\|p - q\|_1 = \|\rho - \sigma\|_1$. In particular, if $\rho \neq \sigma$ then $\|p - q\|_1 = \|\rho - \sigma\|_1 > 0$, meaning that $p \neq q$. Since the classical relative entropy is a divergence, $D(\rho\|\sigma) \geq D(p\|q) > 0$ by a similar argument as above. Hence the quantum relative entropy is also a divergence. \square

- *Joint convexity:* Let Σ be a finite set, $p \in \mathcal{P}(\Sigma)$ a probability distribution, and $(\rho_x)_{x \in \Sigma}$ and $(\sigma_x)_{x \in \Sigma}$ families of states in $\mathcal{D}(\mathcal{H})$.

$$D\left(\sum_{x \in \Sigma} p_x \rho_x \left\| \sum_{x \in \Sigma} p_x \sigma_x\right.\right) \leq \sum_{x \in \Sigma} p_x D(\rho_x \|\sigma_x). \quad (9.6)$$

You will show this in Exercise 9.6.

- Just like in the classical case, $D(\rho\|\sigma)$ is *not* symmetric, i.e., generally $D(\rho\|\sigma) \neq D(\sigma\|\rho)$.

Along the same lines as in the classical case, we can use the quantum relative entropy to quickly derive some entropy inequalities we have seen earlier. Let $\rho, \sigma \in \mathcal{D}(\mathbb{C}^d)$ where $\sigma = I/d$ is the maximally mixed state. You will show in Exercise 9.2 that

$$D(\rho\|\sigma) = \log d - H(\rho), \quad (9.7)$$

implying that $H(\rho) \leq \log d$, with equality iff $\rho = I/d$ is maximally mixed. We know this already from Lemma 7.2. Next, let ρ_{AB} be a bipartite state with marginals $\rho_A = \text{Tr}_B \rho_{AB}$ and $\rho_B = \text{Tr}_A \rho_{AB}$. You will show in Exercise 9.2 that

$$D(\rho_{AB}\|\rho_A \otimes \rho_B) = I(A : B)_{\rho_{AB}}, \quad (9.8)$$

implying $I(A : B)_{\rho_{AB}} \geq 0$, with equality iff $\rho_{AB} = \rho_A \otimes \rho_B$ is a product state. Thus we recover not only the subadditivity inequality but also characterize when equality holds. This proves a claim made below Eq. (8.2) in Lecture 8.

We can also derive the monotonicity of the mutual information [Eq. (8.10)] from the monotonicity of the relative entropy [Eq. (9.4)]. Namely, by choosing $\Phi = \text{Tr}_E$, we obtain

$$I(A : BE)_{\rho_{ABE}} = D(\rho_{ABE}\|\rho_A \otimes \rho_{BE}) \geq D(\rho_{AB}\|\rho_A \otimes \rho_B) = I(A : B)_{\rho_{AB}}.$$

As discussed last week, this inequality is in turn equivalent to strong subadditivity [Eq. (8.5)].

Finally, the Klein inequality [Eq. (9.5)] implies that entropy never decreases under basis measurements:

Lemma 9.8. Let $\rho \in \mathcal{D}(\mathcal{H})$ and let $\mathcal{M}[X] = \sum_{x \in \Sigma} \langle x|X|x\rangle |x\rangle\langle x|$ denote the measurement channel for an arbitrary orthonormal basis $|x\rangle$ of \mathcal{H} . Then,

$$H(\mathcal{M}[\rho]) \geq H(\rho),$$

with equality if and only if $\mathcal{M}[\rho] = \rho$.

Proof. By the Klein inequality, we have

$$0 \leq D(\rho\|\mathcal{M}[\rho]) = -H(\rho) - \text{Tr} \rho \log \mathcal{M}[\rho] = -H(\rho) - \text{Tr} \mathcal{M}[\rho] \log \mathcal{M}[\rho] = -H(\rho) + H(\mathcal{M}[\rho]).$$

The second inequality holds as $\mathcal{M}[\rho]$ is diagonal in the basis $|x\rangle$. \square

9.4 Exercises

9.1 Relative entropy warmup:

- (a) Compute the $D(\rho\|\sigma)$ for $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ and $\sigma = \frac{1}{4}|+\rangle\langle +| + \frac{3}{4}|-\rangle\langle -|$.
- (b) Show that if ρ and σ are both pure states then $D(\rho\|\sigma) \in \{0, \infty\}$.

9.2 From relative entropy to entropy and mutual information:

Prove Eqs. (9.7) and (9.8). For Eq. (9.8) you may assume that $\rho_{AB}, \rho_A, \rho_B$ are positive definite. *Hint: Exercise 7.2.*

9.3 Entropy and ensembles:

The goal of this exercise is to prove Lemma 9.3. It is useful to write $\chi(\{p_x, \rho_x\}) \leq H(p)$ as $H\left(\sum_x p_x \rho_x\right) \leq H(p) + \sum_x p_x H(\rho_x)$.

- (a) First prove the lemma assuming each ρ_x is a pure state, i.e., $\rho_x = |\psi_x\rangle\langle\psi_x|$.
Hint: Consider the pure state $|\Phi_{XB}\rangle = \sum_x \sqrt{p_x}|x\rangle \otimes |\psi_x\rangle$, and compare the entropy of the X system before and after a standard basis measurement.
- (b) Then use part (a) to prove the lemma for general ρ_x .
Hint: Consider an ensemble obtained from the eigendecompositions of all the ρ_x .

9.4 Holevo χ -quantity:

Alice wants to communicate a classical message to Bob by sending a quantum state. She chooses one state $\rho_x \in D(\mathcal{H})$ for each possible message $x \in \Sigma$ that she may want to send, and Bob chooses a measurement $\mu: \Sigma \rightarrow \text{PSD}(\mathcal{H})$ that he uses to decode.

- (a) Write down a formula for the probability that Bob successfully decodes the message if the message is drawn according to an arbitrary probability distribution $p \in P(\Sigma)$.

In class, we used the Holevo bound to prove that if this probability is 100% then, necessarily, the Holevo χ -quantity of the ensemble $\{p_x, \rho_x\}$ must be equal to $H(p)$.

- (b) Show that this condition is also sufficient: If $\chi(\{p_x, \rho_x\}) = H(p)$ then there exists a measurement μ such that Bob decodes the message with 100% probability of success.

9.5 Entropy need not increase:

Find a state ρ and a channel Φ such that $H(\Phi[\rho]) < H(\rho)$.

9.6 Applications of monotonicity:

Show the following two statements by using the monotonicity of the quantum relative entropy:

- (a) *Entropy increase:* If $\Phi \in C(\mathcal{H}, \mathcal{H}')$ is a *unital* channel then we have $H(\Phi[\rho]) \geq H(\rho)$ for any $\rho \in D(\mathcal{H})$. Recall that a channel is *unital* if $\Phi[I_{\mathcal{H}}] = I_{\mathcal{H}'}$.
- (b) *Joint convexity:* For any probability distribution $(p_x)_{x \in \Sigma}$ and families of states $(\rho_x)_{x \in \Sigma}$ and $(\sigma_x)_{x \in \Sigma}$ in $D(\mathcal{H})$, it holds that $D(\sum_{x \in \Sigma} p_x \rho_x \| \sum_{x \in \Sigma} p_x \sigma_x) \leq \sum_{x \in \Sigma} p_x D(\rho_x \| \sigma_x)$. You may assume that the operators ρ_x and σ_x are positive definite.

Hint: In Exercise 7.2 you computed the logarithm of a cq-state.

9.7 Compressing ensembles:

Consider an ensemble $\{p_x, \rho_x\}_{x \in \Sigma}$ of quantum states $\rho_x \in D(\mathcal{H}_A)$. We define an (n, R, δ) -ensemble code to be a pair of channels

$$\mathcal{E} \in C(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes [nR]}) \quad \text{and} \quad \mathcal{D} \in C((\mathbb{C}^2)^{\otimes [nR]}, \mathcal{H}_A^{\otimes n})$$

such that the following holds:

$$\sum_{x^n \in \Sigma^n} p_{x_1} \cdots p_{x_n} F(\mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]], \rho_{x_1} \otimes \cdots \otimes \rho_{x_n}) \geq 1 - \delta. \quad (9.9)$$

We say that an ensemble can be compressed at *rate* R if there exists a sequence of (n, R, δ_n) -ensemble codes, one for each $n \in \mathbb{N}$, such that $\lim_{n \rightarrow \infty} \delta_n = 0$. By Schumacher's theorem and the discussion in Section 7.3, we know that we can compress an ensemble at any rate $R > H(\rho)$, where $\rho = \sum_{x \in \Sigma} p_x \rho_x$. However, in some cases one can do better. In part (d), you will prove that the ensemble's Holevo quantity is an ultimate lower bound on the rate.

- (a) Under what condition is $\chi(\{p_x, \rho_x\})$ equal to $H(\rho)$?
- (b) Give an example of an ensemble $\{p_x, \rho_x\}$ and of a corresponding $(1, R, 0)$ -ensemble code with $R < H(\rho)$.

Now consider an arbitrary ensemble $\{p_x, \rho_x\}$, with cq state $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$.

- (c) Show that if $\sigma_{X^n A^n} = (\mathcal{I}_{X^n} \otimes (\mathcal{D} \circ \mathcal{E}))[\rho_{XA}^{\otimes n}]$ for an (n, R, δ) -ensemble code $(\mathcal{E}, \mathcal{D})$ then

$$R \geq \frac{1}{n} I(X^n : A^n)_\sigma.$$

Hint: Use the data processing inequality and an upper bound on the mutual information.

- (d) Show that it is impossible to compress an ensemble $\{p_x, \rho_x\}$ at a rate smaller than $\chi(\{p_x, \rho_x\})$. You may use without proof that, if ρ_{AB} and σ_{AB} are quantum states with fidelity $F(\rho_{AB}, \sigma_{AB}) \geq 1 - \delta$, then

$$|I(A : B)_\rho - I(A : B)_\sigma| \leq 2\sqrt{\delta} \log(\dim(\mathcal{H}_A) \dim(\mathcal{H}_B)) + 3h(\sqrt{\delta}),$$

where h is the binary Shannon entropy.

Lecture 10

LOCC and separable channels

In Lecture 3 we discussed entanglement – a notion that applies to quantum systems with at least two subsystems, A and B . Recall from Definition 3.1 that a state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is entangled if it is *not* separable, and we call ρ_{AB} separable if

$$\rho_{AB} = \sum_{i \in I} p_i \rho_{A,i} \otimes \rho_{B,i}$$

for some probability distribution $(p_i)_{i \in I}$ and states $\rho_{A,i} \in \mathcal{D}(\mathcal{H}_A)$ and $\rho_{B,i} \in \mathcal{D}(\mathcal{H}_B)$. Entanglement is a synonym for “quantum correlations” – the correlations between subsystems A and B that do not have classical origin. In particular, entanglement cannot be created or increased by the following operations:

- *local operations*, such as unitary operations, isometries, measurement or, more generally, a quantum channel applied to one of the subsystems (e.g., $\Phi_{A \rightarrow A'}$ or $\Psi_{B \rightarrow B'}$);
- *classical communication* (exchanging classical messages between the two subsystems), which can increase classical correlations but not quantum.

In contrast, *global operations* and *quantum communication* can create or increase entanglement.

We refer to the set of operations that include both Local Operations and Classical Communication as LOCC. LOCC plays a central role in quantum information theory. For example, we can then think of entanglement as the resource that cannot be created (or increased) by LOCC. This is a very useful perspective, in particular when it comes to comparing or measuring the amount of entanglement in different states – if a state $|\Psi_{AB}\rangle$ can be converted to some other state $|\Psi'_{AB}\rangle$ by LOCC then we know that $|\Psi_{AB}\rangle$ has at least as much entanglement as $|\Psi'_{AB}\rangle$, since LOCC could not increase the entanglement. For another example, note that the teleportation protocol discussed in Section 3.3 is nothing but an LOCC operation.

Right now our discussion has been somewhat informal. We will now make it more precise.

10.1 Instruments and LOCC channels

Before we can formally define LOCC, let us first introduce the most general type of operation that produces a classical outcome as well as a leftover quantum state (you can think of this as smashing together the notions of a quantum channel and a measurement).

Definition 10.1 (Instrument). An *instrument* is a collection of completely positive maps $\{\Phi_{A \rightarrow B,x}\}_{x \in \Omega} \subseteq \mathcal{CP}(\mathcal{H}_A, \mathcal{H}_B)$ such that $\sum_{x \in \Omega} \Phi_{A \rightarrow B,x} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B)$.

What is the interpretation of an instrument? For this, let $\mathcal{H}_X = \mathbb{C}^\Omega$, and consider the channel $\Phi_{A \rightarrow BX} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_X)$ defined by

$$\Phi_{A \rightarrow BX}[M_A] := \sum_{x \in \Omega} \Phi_{A \rightarrow B, x}[M_A] \otimes |x\rangle\langle x|_X \quad (\forall M_A \in \mathcal{L}(\mathcal{H}_A)) \quad (10.1)$$

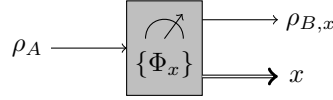
Note that for any state ρ_A , the channel output $\rho_{BX} := \Phi_{A \rightarrow BX}[\rho_A]$ is classical on X . Indeed,

$$\rho_{BX} = \sum_{x \in \Omega} p_x \rho_{B, x} \otimes |x\rangle\langle x|,$$

where

$$p_x = \text{Tr}[\Phi_{A \rightarrow B, x}[\rho_A]] \quad \text{and} \quad \rho_{B, x} = \frac{\Phi_{A \rightarrow B, x}[\rho_A]}{\text{Tr}[\Phi_{A \rightarrow B, x}[\rho_A]]}.$$

In other words, ρ_{BX} is the classical-quantum (cq) state corresponding to the ensemble $\{p_x, \rho_{B, x}\}$. It describes precisely the ensemble of post-measurement states obtained when performing a basis measurement on X (Axioms 2.7 and 2.15). We can visualize the situation as follows:



Note that if we trace over the X system, we obtain

$$\text{Tr}_X \circ \Phi_{A \rightarrow BX} = \sum_{x \in \Omega} \Phi_{A \rightarrow B, x},$$

which is an arbitrary quantum channel (by definition of an instrument!). On the other hand, tracing over the B system gives

$$(\text{Tr}_B \circ \Phi_{A \rightarrow BX})[M_A] = \sum_{x \in \Omega} \text{Tr}[\Phi_{A \rightarrow B, x}[M_A] |x\rangle\langle x|] = \sum_{x \in \Omega} \text{Tr}[\mu_A(x) M_A] |x\rangle\langle x| \quad (\forall M_A \in \mathcal{L}(\mathcal{H}_A)),$$

where we have introduced the measurement

$$\mu_A: \Omega \rightarrow \text{PSD}(\mathcal{H}_A), \quad \mu_A(x) := \Phi_{A \rightarrow B, x}^\dagger[I_B]. \quad (10.2)$$

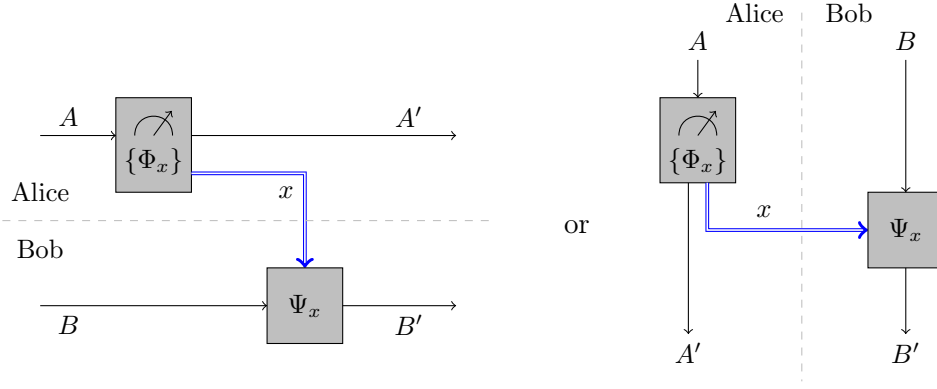
Thus, $\text{Tr}_B \circ \Phi_{A \rightarrow B}$ is precisely the measurement channel (4.29) corresponding to this measurement. This makes it precise that an instrument combines the notions of a quantum channel and a measurement.

It is clear that any combination of applying quantum channels and performing measurements on subsystems can be described by an instrument. Moreover:

Lemma 10.2. *Let $\mathcal{H}_X = \mathbb{C}^\Omega$ for some finite set Ω . If $\Phi_{A \rightarrow BX} \in \mathcal{C}(\mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_X)$ is a channel such that $\Phi_{A \rightarrow BX}[\rho_A]$ is classical on X for every state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$, then it has the form (10.1) for an instrument $\{\Phi_{A \rightarrow B, x}\}_{x \in \Omega}$.*

Thus, instruments are indeed the most general type of operation that produce a classical outcome along with a quantum state. You can verify all claims above in Exercise 10.1.

We can now formally define the set of LOCC quantum channels that can be implemented only by local operations and classical communication. The basic idea is the following: Suppose Alice applies an instrument $\{\Phi_{A \rightarrow A', x}\}_{x \in \Omega}$ and sends her *classical* measurement outcome $x \in \Omega$ over to Bob. Bob then applies an arbitrary quantum channel $\Psi_{B \rightarrow B', x}$ to his system depending on the value of x . We can visualize this as follows:



In the first picture time goes from left to right, as is usual in quantum circuits. In the second picture it goes from top to bottom, which is often used in quantum communication protocols. This is called a *one-way LOCC channel* from Alice to Bob, or a *one-way right LOCC channel* (if we imagine Alice sitting to the left of Bob, as in the second figure). We can similarly define one-way LOCC channels that go the other way around. A general LOCC channel is then a composition of one-way LOCC channels of both types. We now define this formally:

Definition 10.3 (LOCC channels). A channel $\Xi_{AB \rightarrow A'B'} \in C(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ is called:

- *one-way LOCC channel from Alice to Bob, or one-way right LOCC channel*, if

$$\Xi_{AB \rightarrow A'B'} = \sum_{x \in \Omega} \Phi_{A \rightarrow A', x} \otimes \Psi_{B \rightarrow B', x},$$

for an instrument $\{\Phi_{A \rightarrow A', x}\}_{x \in \Omega} \subseteq \text{CP}(\mathcal{H}_A, \mathcal{H}_{A'})$ and channels $\Psi_{B \rightarrow B', x} \in C(\mathcal{H}_B, \mathcal{H}_{B'})$.

- *one-way LOCC channel from Bob to Alice, or one-way left LOCC channel*, if it is of the same form except that now $\{\Psi_{B \rightarrow B', x}\}_{x \in \Omega}$ is an instrument and each $\Phi_{A \rightarrow A', x}$ is a channel.
- *LOCC channel* if it is a composition of any number of one-way LOCC channels of either type.

We write $\text{LOCC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'}) \subseteq C(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ for the set of LOCC channels defined as above. We also abbreviate $\text{LOCC}(\mathcal{H}_A : \mathcal{H}_B) := \text{LOCC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_A : \mathcal{H}_B)$.

Just like for separable states, the colon “:” in the notation signifies how the systems are split between the two parties. By definition, the set of LOCC channels is closed under composition.

Again it is instructive to observe that the teleportation protocol discussed in Section 3.3 is nothing but one-way LOCC channel from Alice to Bob. In Exercise 10.4 you can explore another example of an LOCC channel. In Exercise 10.5 you can show that the set of states that can be created by LOCC (starting from any separable state, or from no state at all) are precisely the separable states.

10.2 Separable channels and operators

Unfortunately, LOCC is very hard to deal with mathematically. Therefore we often relax it to a somewhat larger class of operations known as separable operations or SepC:

Definition 10.4 (Separable maps and channels). A completely positive map $\Xi_{AB \rightarrow A'B'} \in$

$\text{CP}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ is called *separable* if it is of the form

$$\Xi_{AB \rightarrow A'B'} = \sum_{i \in I} \Phi_{A \rightarrow A', i} \otimes \Psi_{B \rightarrow B', i},$$

where $\Phi_{A \rightarrow A', i} \in \text{CP}(\mathcal{H}_A, \mathcal{H}_{A'})$ and $\Psi_{B \rightarrow B', i} \in \text{CP}(\mathcal{H}_B, \mathcal{H}_{B'})$ for every $i \in I$.

We denote the set of separable completely positive maps by $\text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$, and the set of separable channels by $\text{SepC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$. Finally, we abbreviate $\text{SepCP}(\mathcal{H}_A : \mathcal{H}_B) = \text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_A : \mathcal{H}_B)$ and $\text{SepC}(\mathcal{H}_A : \mathcal{H}_B) = \text{SepC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_A : \mathcal{H}_B)$.

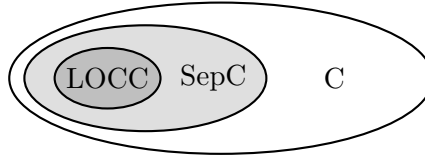
By definition, the sets of separable completely positive maps and separable channels are related as follows:

$$\text{SepC}(\mathcal{H}_A, \mathcal{H}_B : \mathcal{H}_{A'}, \mathcal{H}_{B'}) = \text{SepCP}(\mathcal{H}_A, \mathcal{H}_B : \mathcal{H}_{A'}, \mathcal{H}_{B'}) \cap \text{C}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}).$$

Clearly, any one-way LOCC channel is separable. Since the separable channels are also closed under composition, it follows that the LOCC channels are a subset of the separable channels. This is stated in the following lemma (which you will prove in Exercise 10.2) and corollary:

Lemma 10.5 (Composition of separable maps). *If $\Xi \in \text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$ and $\Gamma \in \text{SepCP}(\mathcal{H}_{A'} : \mathcal{H}_{B'}, \mathcal{H}_{A''} : \mathcal{H}_{B''})$, then $\Gamma \circ \Xi \in \text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A''} : \mathcal{H}_{B''})$. The same holds if we replace SepCP by SepC .*

Corollary 10.6 (LOCC in separable). $\text{LOCC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'}) \subseteq \text{SepC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$.



We will now give two mathematical characterizations of separable maps. The first is in terms of the Kraus representation, and you will prove it in Exercise 10.3.

Lemma 10.7 (Kraus vs separable). *Let $\Xi \in \text{CP}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$. Then, Ξ is separable if and only if it has a Kraus representation with Kraus operators of the form $Y \otimes Z$, where $Y \in \text{L}(\mathcal{H}_A, \mathcal{H}_{A'})$ and $Z \in \text{L}(\mathcal{H}_B, \mathcal{H}_{B'})$.*

The second characterization is in terms of the Choi operator. Namely, a completely positive map is separable if and only if the corresponding Choi operator is separable. Since the Choi operator need not be a state, we first generalize the definition of separability from states to arbitrary PSD operators:

Definition 10.8 (Separable operator). Let \mathcal{H}_A and \mathcal{H}_B be two Hilbert spaces. Then an operator $M_{AB} \in \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is *separable (between A and B)* if

$$M_{AB} = \sum_{i \in I} P_{A,i} \otimes Q_{B,i} \tag{10.3}$$

for $P_{A,i} \in \text{PSD}(\mathcal{H}_A)$, $Q_{B,i} \in \text{PSD}(\mathcal{H}_B)$. We denote the set of separable operators by $\text{Sep}(\mathcal{H}_A : \mathcal{H}_B)$.

It is easy to see that the separable states (Definition 3.1) are just the separable operators that are states:

$$\text{SepD}(\mathcal{H}_A : \mathcal{H}_B) = \text{Sep}(\mathcal{H}_A : \mathcal{H}_B) \cap \text{D}(\mathcal{H}_A \otimes \mathcal{H}_B).$$

Then we have the following result:

Lemma 10.9 (Choi vs separable). *Let $\Xi \in \text{CP}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$. Let*

$$J_{ABA'B'}^\Xi := \sum_{a,b,\tilde{a},\tilde{b}} |ab\rangle\langle\tilde{a}\tilde{b}| \otimes \Xi[|ab\rangle\langle\tilde{a}\tilde{b}|] \in \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$$

denote its Choi operator, defined as in (5.1) with respect to a product basis $\{|a\rangle \otimes |b\rangle\}$ of $\mathcal{H}_A \otimes \mathcal{H}_B$. Then, Ξ is separable if and only if $J_{ABA'B'}^\Xi$ is separable between AA' and BB' .¹

Proof. If Ξ is separable then by definition this means that we can write $\Xi = \sum_i \Phi_i \otimes \Psi_i$ for completely positive maps Φ_i and Ψ_i . Then,

$$\begin{aligned} J_{ABA'B'}^\Xi &= \sum_i \sum_{a,b,\tilde{a},\tilde{b}} |ab\rangle\langle\tilde{a}\tilde{b}| \otimes (\Phi_i \otimes \Psi_i)[|ab\rangle\langle\tilde{a}\tilde{b}|] \\ &= \sum_i \sum_{a,b,\tilde{a},\tilde{b}} |a\rangle\langle\tilde{a}| \otimes |b\rangle\langle\tilde{b}| \otimes (\Phi_i \otimes \Psi_i)[|a\rangle\langle\tilde{a}| \otimes |b\rangle\langle\tilde{b}|] \\ &= U^\dagger \left(\sum_i J_{AA'}^{\Phi_i} \otimes J_{BB'}^{\Psi_i} \right) U, \end{aligned}$$

where U is the unitary from the footnote. Since Φ_i and Ψ_i are completely positive, their Choi states are PSD. Thus we recognize that $J_{ABA'B'}^\Xi$ is separable between AA' and BB' .

Conversely, suppose that $J_{ABA'B'}^\Xi$ is separable between AA' and BB' . This means that we can write

$$J_{ABA'B'}^\Xi = \sum_i |v_i\rangle\langle v_i|, \quad \text{where} \quad |v_i\rangle = U^\dagger (|\alpha_i\rangle_{AA'} \otimes |\beta_i\rangle_{BB'})$$

for suitable vectors $|\alpha_i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A'}$ and $|\beta_i\rangle \in \mathcal{H}_B \otimes \mathcal{H}_{B'}$ that need not be normalized. We claim that Ξ has a Kraus representation with Kraus operators of tensor product form, so that Lemma 10.7 implies the claim. Indeed, recall from Eq. (5.9) in the proof of Theorem 5.3 that we obtain a Kraus representation by defining the Kraus operators

$$\begin{aligned} X_i &:= \sum_{a,b,a',b'} \langle a,b,a',b' | v_i \rangle |a',b'\rangle \langle a,b| \\ &= \sum_{a,b,a',b'} \langle a,a',b,b' | \alpha_i \otimes \beta_i \rangle |a',b'\rangle \langle a,b| \\ &= \left(\sum_{a,a'} \langle a,a' | \alpha_i \rangle |a'\rangle \langle a| \right) \otimes \left(\sum_{b,b'} \langle b,b' | \beta_i \rangle |b'\rangle \langle b| \right). \end{aligned}$$

This concludes the proof. □

¹This means that $U J_{ABA'B'}^\Xi U^\dagger \in \text{Sep}(\mathcal{H}_A \otimes \mathcal{H}_{A'} : \mathcal{H}_B \otimes \mathcal{H}_{B'})$, where $U \in \text{U}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}, \mathcal{H}_A \otimes \mathcal{H}_{A'} \otimes \mathcal{H}_B \otimes \mathcal{H}_{B'})$ interchanges systems B and A' .

10.3 Entanglement rank

Presumably some entangled states are more entangled than others, however we do not yet have any way of measuring this. The following definition provides a first (albeit somewhat rough) way to quantify the amount of entanglement of a general state. The idea essentially is to extend the notion of Schmidt rank to mixed states.

Recall from Theorem 2.20 that the Schmidt rank of $|\Psi_{AB}\rangle$ is the number of non-zero coefficients in a Schmidt decomposition of $|\Psi_{AB}\rangle$. For pure states, this is a meaningful measure of entanglement since the separable pure states are precisely those with Schmidt rank 1. We can extend this notion to mixed states and indeed to general PSD operator by decomposing a given operator as a sum of (unnormalized) pure states with as small Schmidt rank as possible:

Definition 10.10 (Entanglement rank). For any integer $r \geq 1$, we define $\text{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$ as the set of operators $M_{AB} \in \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that we can write

$$M_{AB} = \sum_{i \in I} |\Psi_{AB,i}\rangle \langle \Psi_{AB,i}|,$$

where each $|\Psi_{AB,i}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is a vector of Schmidt rank at most r . The *entanglement rank* of M_{AB} is by definition the smallest integer $r \geq 1$ such that $M_{AB} \in \text{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$.

Equivalently, $\text{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$ is the convex cone generated by the pure states in $D(\mathcal{H}_A \otimes \mathcal{H}_B)$.

For pure states, the entanglement rank coincides with the Schmidt rank since the decomposition consists only of a single term (if we combine all terms that are proportional to each other). Note that larger entanglement rank corresponds to more entanglement since

$$\text{Sep}(\mathcal{H}_A : \mathcal{H}_B) = \text{Ent}_1(\mathcal{H}_A : \mathcal{H}_B) \subset \text{Ent}_2(\mathcal{H}_A : \mathcal{H}_B) \subset \dots \subset \text{Ent}_d(\mathcal{H}_A : \mathcal{H}_B) = \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B),$$

where $d = \min(\dim \mathcal{H}_A, \dim \mathcal{H}_B)$. All inclusions are strict. Moreover, all these sets are convex cones. Indeed, Ent_r is the convex cone spanned by the pure states of Schmidt rank at most r .

Entanglement rank is only a rough measure of entanglement, since it only takes on integer values, $r \in \{1, \dots, n\}$. However, it is still meaningful. Indeed, the next theorem shows that separable channels cannot increase the entanglement rank (in particular, they cannot create entangled states out of separable ones).

Theorem 10.11 (Separable maps cannot increase entanglement rank). If $\Xi \in \text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$ and $M \in \text{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$, then $\Xi[M] \in \text{Ent}_r(\mathcal{H}_{A'} : \mathcal{H}_{B'})$.

Proof. Clearly we may assume that $M = |\Psi_{AB}\rangle \langle \Psi_{AB}|$, where $|\Psi_{AB}\rangle = \sum_{j=1}^r s_j |e_j\rangle \otimes |f_j\rangle$. Recall from Lemma 10.7 that there exists Kraus representation with Kraus operators $\{Y_i \otimes Z_i\}$. Then,

$$\Xi[M] = \sum_i (Y_i \otimes Z_i) |\Psi_{AB}\rangle \langle \Psi_{AB}| (Y_i \otimes Z_i)^\dagger = \sum_i |\Psi_{A'B',i}\rangle \langle \Psi_{A'B',i}|,$$

where

$$|\Psi_{A'B',i}\rangle := \sum_{j=1}^r s_j Y_i |e_j\rangle \otimes Z_i |f_j\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}.$$

While this need *not* be a Schmidt decomposition, the fact that there are at most r summands implies at once that each $|\Psi_{A'B',i}\rangle$ has Schmidt rank at most r . Thus, $\Xi[M]$ has entanglement rank at most r . \square

As a special case of this theorem, we see that the set of separable operators is preserved by separable maps:

Corollary 10.12 (Separable maps preserve separability). *If $\Xi \in \text{SepCP}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_{A'} : \mathcal{H}_{B'})$ and $M \in \text{Sep}(\mathcal{H}_A : \mathcal{H}_B)$, then $\Xi[M] \in \text{Sep}(\mathcal{H}_{A'} : \mathcal{H}_{B'})$.*

In particular, LOCC channels cannot increase the entanglement rank and preserve separability.

10.4 Separable and LOCC measurements

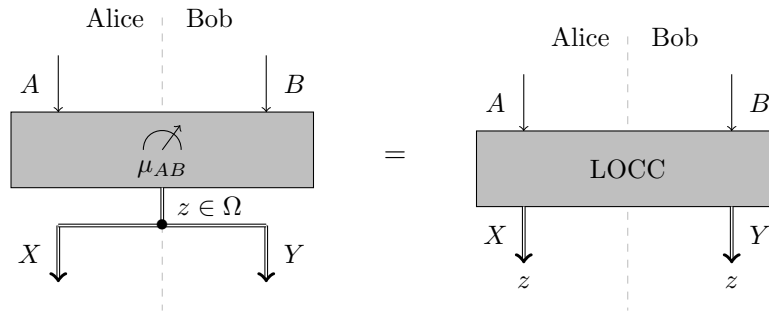
It is often useful to consider measurements that are either separable or LOCC. Imagine that Alice and Bob share a state $\rho_{AB,x}$ selected from some set of states, and they want to determine which state it is. However they cannot exchange any quantum information but can communicate only classically. This corresponds precisely to LOCC channels where at the end both Alice and Bob are left with a classical result (namely, their guess x for what state they started with). In this context, it is interesting to compare how well LOCC performs compared to the slightly more general separable operations. Since quantum-to-classical channels are precisely given by measurements, we can also use the language of measurements to study this problem:

Definition 10.13 (LOCC and separable measurements). Let $\mu_{AB} : \Omega \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a measurement. We say that μ_{AB} is an *LOCC measurement* if the corresponding quantum-to-classical channel $\Phi \in \mathcal{C}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_X \otimes \mathcal{H}_Y)$ defined by

$$\Phi[M_{AB}] = \sum_{z \in \Omega} \text{Tr}[\mu_{AB}(z) M_{AB}] |z\rangle\langle z|_X \otimes |z\rangle\langle z|_Y \quad (\forall M_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B))$$

is in $\text{LOCC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_X : \mathcal{H}_Y)$. Here, $\mathcal{H}_X = \mathcal{H}_Y = \mathbb{C}^\Omega$. Similarly, we say μ_{AB} is a *separable measurement* if $\Phi \in \text{SepC}(\mathcal{H}_A : \mathcal{H}_B, \mathcal{H}_X : \mathcal{H}_Y)$.

In other words, LOCC measurements are in one-to-one correspondence with the quantum-to-classical LOCC channels where Alice's output and Bob's output are perfectly correlated (are the same). The following figure illustrates the condition that μ_{AB} is LOCC:



In Exercises 10.6 and 10.9 you can design LOCC measurements that distinguish various (pure) quantum states.

While it is again difficult to characterize LOCC measurements, there is a clean criterion for verifying when a measurement is separable. You can prove this in Exercise 10.7.

Lemma 10.14. *Let $\mu_{AB} : \Omega \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a measurement. Then, μ_{AB} is separable if and only if $\mu_{AB}(z)$ is separable for all $z \in \Omega$.*

Above it did not matter whether Alice, Bob, or both of them (like in the definition) end up with the measurement outcome, since we can always by LOCC communicate the measurement outcome from one party to the other. We can also define one-way LOCC measurements. Here one party first performs a measurement, sends the outcome to the second party, who then adaptively performs another measurement that depends on the received value. The result of such a one-way LOCC measurement is by definition the outcome of this second measurement. We now define this formally:

Definition 10.15 (One-way LOCC measurements). A measurement $\mu_{AB}: \Omega \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called a *one-way LOCC measurement from Alice to Bob*, or *one-way right LOCC measurement*, if it is of the form

$$\mu_{AB}(y) = \sum_{x \in \Gamma} \nu_A(x) \otimes \pi_{B,x}(y)$$

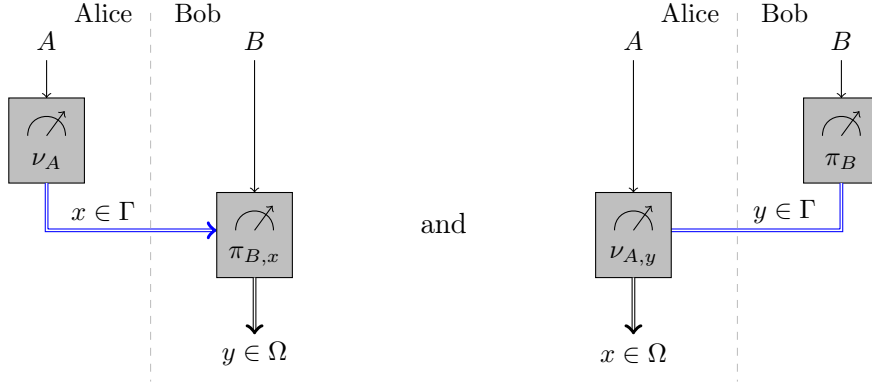
where Γ is some arbitrary finite set, $\nu_A: \Gamma \rightarrow \text{PSD}(\mathcal{H}_A)$ a measurement, and $\pi_{B,x}: \Omega \rightarrow \text{PSD}(\mathcal{H}_B)$ is a measurement for every $x \in \Gamma$.

Similarly, μ_{AB} is called a *one-way LOCC measurement from Bob to Alice*, or *one-way left LOCC measurement*, if it is of the form

$$\mu_{AB}(x) = \sum_{y \in \Gamma} \nu_{A,y}(x) \otimes \pi_B(y)$$

where Γ is again some arbitrary finite set, but now $\pi_B: \Gamma \rightarrow \text{PSD}(\mathcal{H}_B)$ is a single measurement, and $\nu_{A,y}: \Omega \rightarrow \text{PSD}(\mathcal{H}_A)$ is a measurement for every $y \in \Gamma$.

We can visualize the two types of one-way LOCC measurements as follows:



Remark 10.16. We can also define one-way LOCC measurements as in Definition 10.13. For example, a measurement $\mu_{AB}: \Omega \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is one-way right LOCC if and only if the channel

$$\Phi_{AB \rightarrow XY}[M_{AB}] = \sum_{z \in \Omega} \text{Tr}[\mu_{AB}(z) M_{AB}] |0\rangle\langle 0|_X \otimes |z\rangle\langle z|_Y \quad (\forall M_{AB} \in \text{L}(\mathcal{H}_A \otimes \mathcal{H}_B))$$

is one-way right LOCC, where $\mathcal{H}_X = \mathbb{C}$ and $\mathcal{H}_Y = \mathbb{C}^\Omega$. Note that unlike in Definition 10.13, only one side (the “receiver”) ends up with the measurement result. This is very natural in view of the asymmetric nature of one-way LOCC. Equivalently, one-way LOCC measurements correspond to quantum-to-classical one-way LOCC channels where the ‘sender’ only has a one-dimensional output system (equivalently, no output at all). Can you see why the above claims are true?

While one-way LOCC measurements may seem rather limited, they can perfectly discriminate any two orthogonal pure bipartite states, even if the states are entangled:

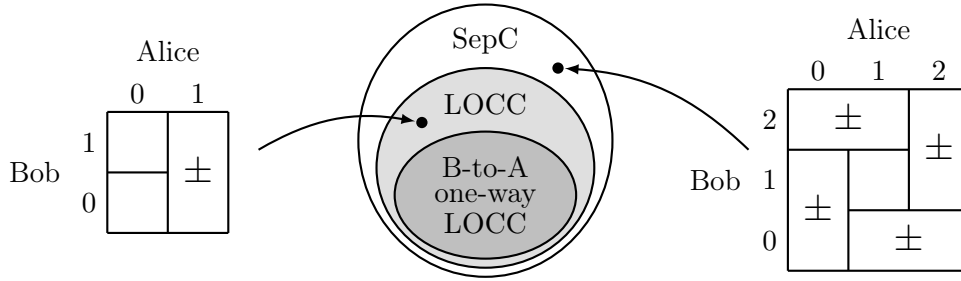
Theorem 10.17 (Perfect one-way LOCC measurement for discriminating orthogonal pure states). *For any pair of orthogonal pure states $\rho_{AB,x} = |\Psi_{AB,x}\rangle\langle\Psi_{AB,x}|$ (i.e., $\langle\Psi_{AB,0}|\Psi_{AB,1}\rangle = 0$), there is a one-way LOCC measurement $\mu_{AB}: \{0, 1\} \rightarrow \text{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\text{Tr}[\mu_{AB}(x)\rho_{AB,y}] = \delta_{x,y}$.*

This theorem has a nice proof that is not too complicated, but we will not prove it the class. It is quite surprising, since it shows that LOCC measurements are as good as global measurements for the task of discriminating *pairs* of orthogonal *pure* states.

Consider the following four (!) orthonormal product states:

$$|\Psi_{AB,1}\rangle = |0\rangle \otimes |0\rangle, \quad |\Psi_{AB,2}\rangle = |0\rangle \otimes |1\rangle, \quad |\Psi_{AB,3}\rangle = |1\rangle \otimes |+\rangle, \quad |\Psi_{AB,4}\rangle = |1\rangle \otimes |-\rangle$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. They are visualized by the left-hand tiles in the following picture:



In Exercise 10.9 you will show that these four states can be perfectly discriminated by a separable measurement or a one-way LOCC measurement from Alice to Bob, but not by a one-way LOCC measurement from Bob to Alice. Hence, as illustrated in the above diagram, the corresponding measurement is in LOCC but not in one-way LOCC from Bob to Alice. Using the same idea, one can come up with a slightly more complicated set of orthogonal product states in $\mathbb{C}^3 \otimes \mathbb{C}^3$ that cannot be perfectly discriminated by LOCC, even with two-way communication. However, the states can still be distinguished by a separable measurement. This shows that separable measurements are strictly more powerful than LOCC.

10.5 Exercises

10.1 Instruments:

- Verify that Eq. (10.1) defines a quantum channel.
- Verify that Eq. (10.2) defines a measurement.
- Prove Lemma 10.2.

10.2 Composition of separable maps: Prove Lemma 10.5.

10.3 Kraus representation of separable maps: Prove Lemma 10.7.

10.4 LOCC: Show that the following process can be implemented by an LOCC channel. Alice starts with a quantum system A and Bob with a quantum system B . Let (x, y) be drawn from an arbitrary probability distribution $p \in \mathcal{P}(\Sigma_X \times \Sigma_Y)$. Then Alice applies a channel $\Phi_{A \rightarrow A',x}$ on her system and Bob applies a channel $\Psi_{B \rightarrow B',y}$ on his system.

10.5 **LOCC vs separable:** Suppose Alice and Bob start out with an arbitrary separable state.² Show that the states they can create by LOCC are precisely the separable states.

10.6 **Discriminating Bell states by LOCC:** Recall the four Bell states from Eq. (3.4). Assume that Alice holds the first qubit of a Bell state and Bob holds the second qubit.

- (a) Find an LOCC protocol that can perfectly discriminate between $|\Phi^{(00)}\rangle$ and $|\Phi^{(10)}\rangle$.
- (b) Find an LOCC protocol that can perfectly discriminate between $|\Phi^{(00)}\rangle$ and $|\Phi^{(01)}\rangle$.

10.7 **Separable measurements:** Prove Lemma 10.14.

10.8 **Separable measurements:** Let $\{|\Psi_{AB,i}\rangle\}$ be an arbitrary orthonormal basis of $\mathcal{H}_A \otimes \mathcal{H}_B$. Show that the basis vectors can be perfectly distinguished by a separable measurement if and only if the basis consists of product states.

10.9 **One-way LOCC struggle:** Suppose Alice the first qubit and Bob holds the second qubit of a two-qubit system initialized in one of the following four states:

$$\begin{aligned} |\Psi_{AB,1}\rangle &= |0\rangle \otimes |0\rangle, \\ |\Psi_{AB,2}\rangle &= |0\rangle \otimes |1\rangle, \\ |\Psi_{AB,3}\rangle &= |1\rangle \otimes |+\rangle, \\ |\Psi_{AB,4}\rangle &= |1\rangle \otimes |-\rangle. \end{aligned}$$

- (a) Find a separable measurement that perfectly distinguishes the above four states.
- (b) Find a one-way LOCC measurement from Alice to Bob that perfectly distinguishes the above four states.
- (c) Show that there is *no* one-way LOCC measurement from Bob to Alice that can perfectly determine which of the four states they share.

10.10 **Operations on PPT states:** Suppose ρ_{AB} is an arbitrary state with positive semidefinite partial transpose (Exercise 3.7).

- (a) Show that if $\Xi_{AB \rightarrow A'B'}$ is a separable channel then the partial transpose of the state $\omega_{A'B'} := \Xi_{AB \rightarrow A'B'}[\rho_{AB}]$ is again positive semidefinite.
- (b) Show that it is not possible to transform ρ_{AB} into a maximally entangled state by LOCC.

²For example, they could start out with no quantum systems at all. This would be modeled by the trivial Hilbert spaces $\mathcal{H}_A = \mathbb{C}$ and $\mathcal{H}_B = \mathbb{C}$, and the (unique) state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B) = \mathcal{D}(\mathbb{C} \otimes \mathbb{C}) = \mathcal{D}(\mathbb{C}) = \{I\}$, where I denotes the identity operator on \mathbb{C} .

Lecture 11

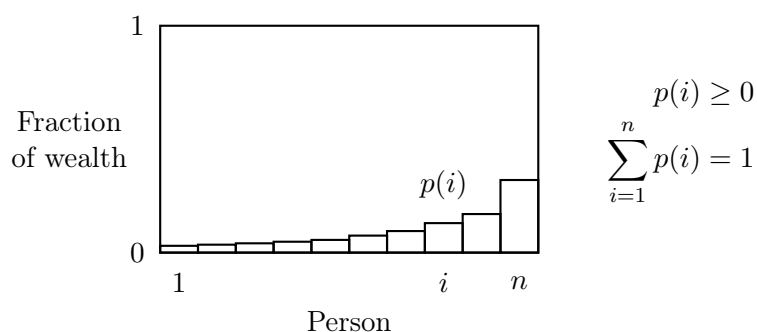
Majorization and Nielsen's theorem

Last week we looked at LOCC and separable channels as well as measurements. While separable maps are easier to define and work with mathematically, LOCC is more important from physical and operational perspective. Recall that LOCC is a subset of separable maps.

In this class, we will look at a different problem. Instead of trying to discriminate states by an LOCC measurement, we will try to perfectly convert one state into another. You can think of the discrimination problem as a special case of this, since a measurement effectively converts given states to different standard basis states. The general problem of converting one arbitrary set of states to another by LOCC is complicated, so we will only consider the case of converting a single pure state to another pure state. The answer to this problem is known, nontrivial, and closely tied with the notion of majorization. Thus we first need to learn the basics of majorization.

11.1 Wealth inequality and majorization

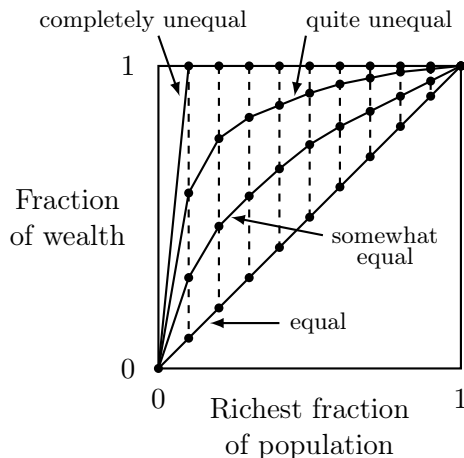
The concept of majorization is most intuitive in the context it was first introduced, namely as a way to measure wealth inequality. It is convenient to describe the distribution of wealth by a probability distribution p where $p(i)$ is the fraction of wealth owned by person $i \in \{1, \dots, n\}$. Alternatively, you can think of the total wealth as being normalized to 1 and $p(i)$ simply denoting the wealth of person i . You can depict the distribution p as follows:



Given two probability distributions p and q , how can we tell which one corresponds to a “more equal” distribution of wealth? Clearly, $p = (1, 0, \dots, 0)$ is the least equal distribution of wealth and $q = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ is the most equal. How about the rest and how can we compare two distributions?

One way to compare wealth distributions is by considering the fraction of wealth owned by the richest. More specifically, let us plot the cumulative wealth of the richest fraction of the population.

For this, we need to sort the probability distribution p so that $p(1) \geq p(2) \geq \dots \geq p(n)$, and let $f_p(k) = \sum_{i=1}^k p(i)$ be the total wealth of the k richest people. We can try to compare different wealth distributions p by plotting the corresponding cumulative wealth function f_p :



If f_q lies completely below f_p then the distribution q is more equal than p . Let us now turn this into a mathematical definition. It will be convenient to define it for arbitrary real vectors, not just for probability distributions.

Definition 11.1 (Majorization). Let $x, y \in \mathbb{R}^n$. Then we write $x \prec y$ if and only if

$$\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow \quad (\forall k \in \{1, \dots, n-1\}) \quad \text{and} \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i.$$

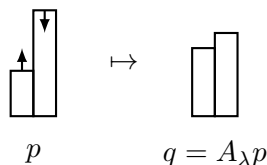
Here $x^\downarrow \in \mathbb{R}^n$ denotes the vector with the same entries as $x \in \mathbb{R}^n$, but sorted nonincreasingly, i.e., $x_1^\downarrow \geq \dots \geq x_n^\downarrow$. We say that y *majorizes* x or that x *is majorized by* y .

Note that majorization does not depend on the order of the vectors' entries (so it is really a property of two multisets). When restricted to probability distributions, we do not need to check the last condition since the entries automatically sum to one. Note that in general *neither* $x \prec y$ *nor* $y \prec x$ (Exercise 11.1). In terms of the picture above, this means that two cumulative wealth functions can intersect. You will derive other equivalent conditions in Exercise 11.2.

One obvious way to increase equality is to take from the rich and give to the poor. Let us call this a Robin-Hood move. Mathematically, it is described by the following 2×2 matrix, which should be applied to the corresponding two entries of the distribution:

$$A_\lambda = \lambda I + (1 - \lambda)X = \begin{pmatrix} \lambda & 1 - \lambda \\ 1 - \lambda & \lambda \end{pmatrix},$$

for some $\lambda \in [0, 1]$. If $\lambda \in (0, 1)$, applying this to the wealth of two individuals always has the effect of decreasing the gap between their wealth because the new values are convex combinations of the old ones¹:



¹When $\lambda > 1/2$, the roles of the two individuals in terms of their richness are swapped.

Any sequence of such Robin-Hood moves on a wealth distribution makes the distribution more equal. Note that the overall transformation amounts to a convex combination of permutations. Each Robin-Hood move has the effect of pushing the corresponding cumulative curve downwards. In fact, the same is true for a more general class of transformations called “doubly stochastic” matrices. We define these next.

Definition 11.2 (Doubly stochastic matrices). Let $A \in \mathbb{R}^{n \times n}$. We say that A is *stochastic* if its columns are probability distributions. It is called *doubly stochastic* if both A and A^T are stochastic, that is, if its rows as well as columns are probability distributions. In other words, A is doubly stochastic if its entries are nonnegative and all row and column sums are equal to one:

- $A_{ij} \geq 0$ for all $i, j \in \{1, \dots, n\}$,
- $\sum_{j=1}^n A_{ij} = 1$ for all $i \in \{1, \dots, n\}$,
- $\sum_{i=1}^n A_{ij} = 1$ for all $j \in \{1, \dots, n\}$.

We already know stochastic matrices under the name *classical channel*. Indeed, we saw in Section 4.3 that they are the most general linear transformations that map probability distributions to probability distributions.

How about the doubly stochastic matrices? To get some intuition, note that permutation matrices are doubly stochastic matrices:

Definition 11.3 (Permutation matrices). For any permutation $\pi \in S_n$, the *permutation matrix* P_π is defined as the matrix with entries $(P_\pi)_{i,j} = \delta_{i,\pi(j)}$ for all i, j . That is, $P_\pi = \sum_{j=1}^n |\pi(j)\rangle\langle j|$.

It is easy to see that the permutation matrices are the doubly stochastic matrices with all entries in $\{0, 1\}$. One can also show that they are the stochastic matrices that have a stochastic inverse, i.e., the classical channels that have an inverse channel.

Another example are the Robin-Hood matrices, which we now define formally.

Definition 11.4 (Robin-Hood matrices). A *Robin-Hood matrix* is a matrix of the form

$$\lambda I + (1 - \lambda)P_\tau,$$

where $\lambda \in [0, 1]$ and τ is a *transposition*, i.e., a permutation that fixes all but two elements.

The product of two permutation matrices is another permutation matrix. Moreover, they are closed under taking inverses. Doubly stochastic matrices are in general not invertible, e.g., note that $\frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is doubly stochastic (in fact, a Robin-Hood matrix) but not invertible. However, they are still closed under composition:

Lemma 11.5. *If A, B are doubly stochastic $n \times n$ matrices, then so is AB .*

Proof. The entries of AB are clearly nonnegative. Thus it suffices to verify its row and column sums: Indeed, for all $i \in \{1, \dots, n\}$ we can compute the corresponding row sum as

$$\sum_{j=1}^n (AB)_{ij} = \sum_{j=1}^n \sum_{k=1}^n A_{ik} B_{kj} = \sum_{k=1}^n A_{ik} \sum_{j=1}^n B_{kj} = \sum_{k=1}^n A_{ik} = 1.$$

A similar calculation shows that the column sums are all equal to one. □

It is true that any doubly stochastic matrix can be written as a finite product of Robin-Hood matrices, but we will not need this here.

The set of doubly stochastic matrices is compact and convex. In fact, the set of $n \times n$ doubly stochastic matrices is *convex polytope*, since it is defined by a finite number of linear equations and inequalities. Equivalently, it is the convex hull of finitely many extreme points, called the *vertices* of the polytope. These vertices are precisely the permutation matrices, as we show in the proof of the following theorem.

Theorem 11.6 (Birkhoff–von Neumann). *A matrix $A \in \mathbb{R}^{n \times n}$ is doubly stochastic if and only if there exists a probability distribution $(q_\pi)_{\pi \in S_n}$ on the set of permutations S_n such that*

$$A = \sum_{\pi \in S_n} q_\pi P_\pi. \quad (11.1)$$

Proof. Let DS_n denote the set of doubly stochastic $n \times n$ matrices. Since DS_n is convex and contains the permutation matrices, any matrix of the form (11.1) is contained in it.

To show that, conversely, any element in DS_n can be written in this form, it suffices to argue that the vertices of DS_n are the permutation matrices. For this we use some basic convex geometry. We start by noting that the space of matrices $A \in \mathbb{R}^{n \times n}$ has dimension n^2 . If we add the constraints that all row and column sums are equal to one, then A is restricted to an affine subspace of dimensions at least $n^2 - (2n - 1) = (n - 1)^2$, since there is at least one redundancy in these constraints (since $\sum_i \sum_j A_{ij} = \sum_j \sum_i A_{ij}$). Thus if $A \in \mathbb{R}^{n \times n}$ is a vertex then at least $(n - 1)^2$ of the inequalities $A_{ij} \geq 0$ have to hold with equality. That is, the matrix A has at least $(n - 1)^2$ zero entries. Since $(n - 1)^2 > n(n - 2)$, it follows that there must exist some i^* such that the i^* -th row contains $n - 1$ zero entries. Since the row sums are equal to one, it follows that there exists j^* such that $A_{i^*j^*} = 1$. Now let A' denote the matrix obtained by deleting the i^* -th row and the j^* -th column. It is not hard to verify that $A' \in DS_{n-1}$ and indeed an extreme point. By induction we find that A' and hence A is a permutation matrix. \square

We now show that majorization is intimately connected to doubly stochasticity. Not only is the output of a doubly stochastic matrix majorized by the input, but the converse is also true:

Theorem 11.7 (Hardy-Littlewood-Pólya). *Let $x, y \in \mathbb{R}^n$. Then, $x \prec y$ if and only if there exists a doubly stochastic matrix A such that $x = Ay$. Moreover, A can be taken to be a product of Robin-Hood matrices.*

Proof. For either direction we may without loss of generality assume that $x_1 \geq x_2 \geq \dots \geq x_n$ and $y_1 \geq y_2 \geq \dots \geq y_n$. (Can you see why this is true?)

First suppose that $x = Ay$ for a doubly stochastic matrix A . Then, we have, for any $k \in \{1, \dots, n - 1\}$, that

$$\sum_{i=1}^k x_i = \sum_{i=1}^k \sum_{j=1}^n A_{ij} y_j = \sum_{j=1}^n p_j y_j,$$

where we defined $p_j := \sum_{i=1}^k A_{ij}$. Note that $p_j \in [0, 1]$ and $\sum_{j=1}^n p_j = \sum_{i=1}^k \sum_{j=1}^n A_{ij} = k$, since A is doubly stochastic. Thus,

$$\sum_{j=1}^k y_j - \sum_{i=1}^k x_i = \sum_{j=1}^k y_j - \sum_{j=1}^n p_j y_j - y_k \left(k - \sum_{j=1}^n p_j \right)$$

$$\begin{aligned}
&= \sum_{j=1}^k (y_j - y_k) - \sum_{j=1}^n p_j (y_j - y_k) \\
&= \sum_{j=1}^k \underbrace{(y_j - y_k)}_{\geq 0} \underbrace{(1 - p_j)}_{\geq 0} + \sum_{j=k+1}^n \underbrace{p_j}_{\geq 0} \underbrace{(y_k - y_j)}_{\geq 0} \geq 0.
\end{aligned}$$

Since also $\sum_{i=1}^n x_i = \sum_{i,j=1}^n A_{ij} y_j = \sum_{j=1}^n y_j$ we conclude that $x \prec y$.

Now suppose that $x \prec y$ and $x \neq y$ (since otherwise we are done). We would like to construct a doubly stochastic matrix A such that $x = Ay$. Suppose that x and y differ in $\delta > 0$ entries. We claim that there exists a Robin-Hood matrix A such that x and $z := Ay$ differ in less than δ entries and, moreover, $x \prec z$. Then the result will follow by induction and Lemma 11.5. To prove the claim, let a be the largest index such that $x_a < y_a$, and let b be the smallest index larger than a such that $x_b > y_b$ (such an index always exists, since $x \prec y$ implies that $x_b > y_b$ for the largest index b such that $x_b \neq y_b$). Thus we have:

$$y_a > x_a \geq x_b > y_b$$

We now apply the Robin-Hood matrix

$$A = \lambda I + (1 - \lambda)P_\tau,$$

where $\tau = (a\ b)$ is the transposition that exchanges a and b , $\lambda := 1 - \frac{\varepsilon}{y_a - y_b}$, and $\varepsilon := \min(y_a - x_a, x_b - y_b)$. Clearly, $\varepsilon > 0$ and $\lambda \in (0, 1)$. Then, $z := Ay$ has the same components as y except for the a -th and the b -th entry, where we have

$$z_a = (Ay)_a = \lambda y_a + (1 - \lambda)y_b = y_a + (1 - \lambda)(y_b - y_a) = y_a - \varepsilon$$

and hence

$$z_b = y_b + \varepsilon.$$

Note that the definition of ε implies that either $z_a = x_a$ or $z_b = x_b$. Thus, z differs from x in fewer entries than y . We still need to verify that $x \prec z$. Note that z is still sorted nonincreasingly. Thus it suffices to prove that

$$\sum_{i=1}^k x_i \leq \sum_{i=1}^k z_i \tag{11.2}$$

for all $k \in \{1, \dots, n\}$ (we clearly have equality for $k = n$). For $k < a$, Eq. (11.2) follows from $x \prec y$ since $z_i = y_i$ for $i < a$. Using $z_a = y_a - \varepsilon \geq x_a$, Eq. (11.2) also holds for $k = a$. Our choice of a and b ensures that $x_k = y_k$ for all $k \in \{a + 1, \dots, b - 1\}$, hence Eq. (11.2) also follows for those k . Finally, for $k \in \{b, \dots, n\}$, using $y_a + y_b = z_a + z_b$ we see that $\sum_{i=1}^k y_i = \sum_{i=1}^k z_i$, so Eq. (11.2) again follows from $x \prec y$. This concludes the proof of the claim, and hence of the theorem. \square

Remark 11.8 (Robin Hood). The proof shows that the doubly stochastic matrix A in the statement of Theorem 11.7 can be taken to be a product of Robin-Hood matrices. You might be concerned that we assumed in the proof that the vectors were sorted, but this is not a problem. Indeed, sorting vectors amounts to applying permutation matrices, but any permutation can be written as a product of transposition, and transpositions correspond to special Robin-Hood matrices.

Example 11.9 (Extreme distributions). Let $u = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ denote the uniform distribution and d denote a deterministic distribution on n elements, e.g., $d = (1, 0, \dots, 0)$. Intuitively, they correspond to a maximally equally and maximally unequal way to distribute wealth, respectively. One can easily check that for any probability distribution p on n elements,

$$u \prec p \prec d.$$

Thus Theorem 11.7 shows that one can always perform the conversions

$$u \leftarrow p \leftarrow d$$

by applying doubly stochastic matrices (or a sequences of Robin-Hood matrices, see Remark 11.8).

Majorization is also tightly related to convexity:

Lemma 11.10. *Let p, q be probability distributions. Then, $p \prec q$ if and only if $\sum_{i=1}^n f(p_i) \leq \sum_{i=1}^n f(q_i)$ for all convex function $f: [0, 1] \rightarrow \mathbb{R}$.*

Proof. The “if” follows from part (c) of Exercise 11.2, since the function $f_t(x) := \max(x - t, 0)$ is convex for any $t \in \mathbb{R}$. For the “only if”, note that $p \prec q$ means that there exists a doubly stochastic matrix A such that $p = Aq$ by Theorem 11.7. Thus:

$$\sum_{i=1}^n f(p_i) = \sum_{i=1}^n f\left(\sum_{j=1}^n A_{ij}q_j\right) \leq \sum_{i=1}^n \sum_{j=1}^n A_{ij}f(q_j) = \sum_{j=1}^n \sum_{i=1}^n A_{ij}f(q_j) = \sum_{j=1}^n f(q_j),$$

where we first used that the columns of A are probability distributions and then that the row sums are equal to one. \square

As an application we have the following corollary, which is very intuitive in view of Example 11.9.

Corollary 11.11. *If p, q are probability distributions with $p \prec q$, then $H(p) \geq H(q)$.*

Proof. The function $f: [0, 1] \rightarrow \mathbb{R}$ defined by $f(t) = t \log t$ is convex and $H(p) = -\sum_{i=1}^n f(p_i)$. Thus the claim follows from Lemma 11.10. \square

You can practice majorization in Exercise 11.3.

11.2 Majorization for Hermitian operators

Any Hermitian operator can be unitarily diagonalized with real eigenvalues (Theorem 1.3). Thus one might wonder whether the theory of majorization has a “quantum” extension that deals with Hermitian operators instead of real vectors.

We first define majorization and then prove an analogue of Theorem 11.7. Let us denote by $\lambda(A)$ the vector of eigenvalues of a given operator A , repeated according to their multiplicity. We will not commit to any particular order of the eigenvalues (you can choose your favorite one for definiteness) but instead write $\lambda^\downarrow(A)$ or $\lambda^\uparrow(A)$ if a particular one is important.

Definition 11.12 (Majorization for Hermitian operators). Let A, B be Hermitian operators on some Hilbert space \mathcal{H} . Then we write $A \prec B$ if and only if $\lambda(A) \prec \lambda(B)$. We say that B majorizes A or that A is majorized by B .

Note that for diagonal matrices A , $\lambda(A)$ consists simply of the diagonal entries of A . Hence, majorization for diagonal operators reduces to majorization for vectors, thus recovering the classical notion.

What operations will play the role of permutations? Clearly we should generalize stochastic matrices (classical channels) to quantum channels. Among all stochastic matrices, permutations stand out as precisely those that are invertible and whose inverse is also stochastic, as discussed before. In the quantum case, unitary channels $\Phi[M] = U M U^\dagger$ are the channels that are invertible and whose inverse is also a channel. Since the doubly stochastic matrices are the convex hull of the permutation matrices (Theorem 11.6), this motivates the following definition:

Definition 11.13 (Mixed-unitary channel). A channel $\Phi \in \mathcal{C}(\mathcal{H})$ is called *mixed-unitary* if

$$\Phi[M] = \sum_{x \in \Omega} q_x U_x M U_x^\dagger,$$

for some finite set Ω , a probability distribution $(q_x)_{x \in \Omega}$, and unitaries $U_x \in \mathcal{U}(\mathcal{H})$ for $x \in \Omega$.

Equivalently, a channel is mixed-unitary if and only if it is in the convex hull of the unitary channels. Then the following theorem by Uhlmann (not to be confused with the more famous Uhlmann's Theorem 4.13) provides a quantum version of Theorem 11.7:

Theorem 11.14 (Uhlmann). Let A, B be Hermitian operators on a Hilbert space \mathcal{H} . Then, $A \prec B$ if and only if there exists a mixed unitary channel $\Phi \in \mathcal{C}(\mathcal{H})$ such that $A = \Phi[B]$.

You will prove it in Exercise 11.5.

Example 11.15 (Extreme states). Let \mathcal{H} be a Hilbert space of dimension n . Let $\tau = I/n$ be the maximally mixed state and let $|\psi\rangle\langle\psi|$ be an arbitrary pure state on \mathcal{H} . Then it follows directly from Example 11.9 that for any state $\rho \in \mathcal{D}(\mathcal{H})$,

$$\tau = \frac{I}{n} \prec \rho \prec |\psi\rangle\langle\psi|.$$

Accordingly, one can always perform the conversion

$$\tau = \frac{I}{n} \leftarrow \rho \leftarrow |\psi\rangle\langle\psi|$$

by applying mixed-unitary channels.

We also record the following easy but useful consequence:

Lemma 11.16. Let ρ, σ be quantum states. Then, $\rho \prec \sigma$ if and only if $\text{Tr } f(A) \leq \text{Tr } f(B)$ for all convex functions $f: [0, 1] \rightarrow \mathbb{R}$, where $f(\rho)$ and $f(\sigma)$ are defined as in Definition 1.6.

Proof. Let $n := \dim \mathcal{H}$. Then, clearly, $f(A) = \sum_{i=1}^n f(\lambda_i(A))$ and likewise $f(B) = \sum_{i=1}^n f(\lambda_i(B))$. Thus the claim follows from Lemma 11.10. \square

Corollary 11.17. If ρ, σ are quantum states with $\rho \prec \sigma$, then $H(\rho) \geq H(\sigma)$.

11.3 LOCC and Nielsen's theorem

As discussed before, majorization determines when we can convert density operators by mixed-unitary channels. We now discuss a second quantum interpretation: majorization determines precisely when one bipartite pure state can be converted into another by LOCC. This is the content of and made more precise in the following theorem:

Theorem 11.18 (Nielsen). *Let ρ_{AB}, σ_{AB} be two pure states on $\mathcal{H}_A \otimes \mathcal{H}_B$. Then the following are equivalent:*

- (a) $\rho_A \prec \sigma_A$.
- (a') $\rho_B \prec \sigma_B$.
- (b) *There exists a one-way LOCC channel Ξ from Alice to Bob such that $\Xi[\rho_{AB}] = \sigma_{AB}$.*
- (b') *There exists a one-way LOCC channel Ξ from Bob to Alice such that $\Xi[\rho_{AB}] = \sigma_{AB}$.*
- (c) *There exists an LOCC channel $\Xi \in \text{LOCC}(\mathcal{H}_A : \mathcal{H}_B)$ such that $\Xi[\rho_{AB}] = \sigma_{AB}$.*
- (d) *There exists a separable channel $\Xi \in \text{SepC}(\mathcal{H}_A : \mathcal{H}_B)$ such that $\Xi[\rho_{AB}] = \sigma_{AB}$.*

Before proving the theorem, we note that, unlike in Theorems 11.7 and 11.14 the direction of majorization is *opposite* to the direction in which the processing occurs: when $\rho_A \prec \sigma_A$ we are able to transform $\rho_{AB} \mapsto \sigma_{AB}$! This is in fact intuitive as the following example shows.

Example 11.19 (Product and maximally entangled states). Consider a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^n$. Let $|\Phi_{AB}^+\rangle$ be a maximally entangled state and $|\alpha\rangle \otimes |\beta\rangle$ be a pure product state. Then, we see from Example 11.15 that for any pure state $|\Omega_{AB}\rangle$, we can convert

$$|\Phi_{AB}^+\rangle\langle\Phi_{AB}^+| \mapsto |\Omega_{AB}\rangle\langle\Omega_{AB}| \mapsto |\alpha\rangle\langle\alpha| \otimes |\beta\rangle\langle\beta|.$$

In other words, a maximally entangled state can be converted to any bipartite state, and any bipartite state can be converted to a product state (the second claim is straightforward since Alice and Bob can simply discard their state and prepare their halves of the product state locally).

Proof of Theorem 11.18. It suffices to prove (a) \Rightarrow (b) and (d) \Rightarrow (a). Indeed, (b) \Rightarrow (c) holds by definition and (c) \Rightarrow (d) by Corollary 10.6, so we obtain that (a) to (d) are all equivalent. To obtain the equivalence to the primed statements, simply interchange the roles of systems A and B .

We first prove (a) \Rightarrow (b). The idea of the proof is to turn the mixed-unitary channel from Theorem 11.14, which we get thanks to the majorization condition, into a one-way LOCC channel. By Theorem 11.14, there exists a probability distribution $(q_x)_{x \in \Omega}$ and unitaries $U_{A,x}$ for $x \in \Omega$ such that

$$\sum_{x \in \Omega} q_x U_{A,x} \sigma_A U_{A,x}^\dagger = \rho_A. \quad (11.3)$$

We can use this data to define an instrument. For each $x \in \Omega$, define a completely positive map $\Psi_{A,x} \in \mathcal{C}(\mathcal{H}_A)$ by $\Psi_{A,x}[M] := Y_{A,x} M Y_{A,x}^\dagger$ for $M \in \mathcal{L}(\mathcal{H}_A)$, where

$$Y_{A,x} := \sqrt{q_x} \sqrt{\sigma_A} U_{A,x}^\dagger \sqrt{\rho_A}^{-1}.$$

Here we assumed for simplicity that ρ_A is invertible.² We claim that $\{\Psi_{A,x}\}_{x \in \Omega}$ is an instrument. Since each $\Psi_{A,x}$ is completely positive we only need to verify that $\sum_{x \in \Omega} \Psi_{A,x}$ is trace preserving

²The general case can be proved by a slight variation, where one replaces $\sqrt{\rho_A}^{-1}$ by the inverse of $\sqrt{\rho_A}$ on its support and adds an additional map to the instrument. Alternatively, one can employ a continuity argument.

(and hence a channel). Indeed, using Eq. (11.3) we find that, for all $M \in \mathcal{L}(\mathcal{H}_A)$,

$$\mathrm{Tr} \sum_{x \in \Omega} \Psi_{A,x}[M] = \mathrm{Tr} \sum_x Y_{A,x}^\dagger Y_{A,x} M = \mathrm{Tr} \sqrt{\rho_A}^{-1} \left(\sum_x q_x U_{A,x} \sigma_A U_{A,x}^\dagger \right) \sqrt{\rho_A}^{-1} M = \mathrm{Tr} M.$$

Now suppose that Alice applies this instrument to her part of the state ρ_{AB} and obtains some outcome $x \in \Omega$. Then the state is transformed into

$$\tilde{\omega}_{AB,x} := (\Psi_{A,x} \otimes \mathcal{I}_B) [\rho_{AB}] = q_x \left(\sqrt{\sigma_A} U_{A,x}^\dagger \sqrt{\rho_A}^{-1} \otimes I_B \right) \rho_{AB} \left(\sqrt{\rho_A}^{-1} U_{A,x} \sqrt{\sigma_A} \otimes I_B \right),$$

which is again a pure state (up to normalization). Moreover,

$$\tilde{\omega}_{A,x} = \mathrm{Tr}_B[\tilde{\omega}_{AB,x}] = q_x \left(\sqrt{\sigma_A} U_{A,x}^\dagger \sqrt{\rho_A}^{-1} \right) \rho_A \left(\sqrt{\rho_A}^{-1} U_{A,x} \sqrt{\sigma_A} \right) = q_x \sigma_A.$$

Thus, for any outcome $x \in \Omega$, the resulting state is a pure state whose reduced density matrix is exactly the desired one. By Lemma 2.18, there must exist a unitary $V_{B,x} \in \mathcal{U}(\mathcal{H}_B)$ such that

$$(I_A \otimes V_{B,x}) \tilde{\omega}_{AB,x} (I_A \otimes V_{B,x}^\dagger) = q_x \sigma_{AB}.$$

Thus, if we define unitary channels $\Psi_{B,x} \in \mathcal{C}(\mathcal{H}_B)$ by $\Psi_{B,x}[M] := V_{B,x} M V_{B,x}^\dagger$ for $M \in \mathcal{L}(\mathcal{H}_B)$, then $\Xi := \sum_{x \in \Omega} \Phi_{A,x} \otimes \Psi_{B,x}$ is a one-way LOCC channel from Alice to Bob and satisfies

$$\Xi[\rho_{AB}] = \sum_{x \in \Omega} (\mathcal{I}_A \otimes \Psi_{B,x})[\tilde{\omega}_{AB,x}] = \sum_{x \in \Omega} q_x \sigma_{AB} = \sigma_{AB}.$$

This concludes the proof that (a) \Rightarrow (b).

Next we show (d) \Rightarrow (a). The main idea is to relate the majorization inequalities to the variational characterization of eigenvalues. We may assume that $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^n$. In the following, it will be convenient to denote the standard basis by $|1\rangle, \dots, |n\rangle$ rather than $|0\rangle, \dots, |n-1\rangle$. By the Schmidt decomposition (Theorem 2.20), we may assume that

$$\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|, \quad \text{where} \quad |\Psi_{AB}\rangle = \sum_{i=1}^n \sqrt{\lambda_i^\uparrow(\rho_A)} |ii\rangle \quad (11.4)$$

and $\lambda^\uparrow(\rho_A)$ denotes the eigenvalues of ρ_A sorted *nondecreasingly*. Now let $\Xi \in \mathrm{SepC}(\mathcal{H}_A : \mathcal{H}_B)$ be a separable channel such that $\Xi[\rho_{AB}] = \sigma_{AB}$. By Lemma 10.7, we can find a Kraus representation of the form $\Xi[M_{AB}] = \sum_{x \in \Omega} (Y_x \otimes Z_x) M_{AB} (Y_x^\dagger \otimes Z_x^\dagger)$ for all $M_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $Y_x \in \mathcal{L}(\mathcal{H}_A)$ and $Z_x \in \mathcal{L}(\mathcal{H}_B)$. Thus,

$$\sigma_{AB} = \Xi[\rho_{AB}] = \sum_{x \in \Omega} (Y_x \otimes Z_x) \rho_{AB} (Y_x^\dagger \otimes Z_x^\dagger).$$

Since σ_{AB} is pure and pure states are the extreme points of the convex set of quantum states, it follows that each summand is proportional to σ_{AB} , i.e., there exist $q_x \geq 0$ such that, for all $x \in \Omega$,

$$q_x \sigma_{AB} = (Y_x \otimes Z_x) \rho_{AB} (Y_x^\dagger \otimes Z_x^\dagger).$$

Using Eq. (11.4), we obtain that, for all $x \in \Omega$,

$$q_x \sigma_A = \mathrm{Tr}_B[(Y_x \otimes Z_x) \rho_{AB} (Y_x^\dagger \otimes Z_x^\dagger)] = Y_x \sqrt{\rho_A} Z_x^T \overline{Z_x} \sqrt{\rho_A} Y_x^\dagger,$$

where the second equality follows from Lemma 3.7 (a). Therefore:

$$\lambda_i^\uparrow(\sigma_A) = \sum_{x \in \Omega} \lambda_i^\uparrow(q_x \sigma_A) = \sum_{x \in \Omega} \lambda_i^\uparrow(Y_x \sqrt{\rho_A} Z_x^T \bar{Z}_x \sqrt{\rho_A} Y_x^\dagger) \quad (11.5)$$

To prove that $\rho_A \prec \sigma_A$ we may, by Exercise 11.2, equivalently show that, for all $k \in \{1, \dots, n\}$,

$$\sum_{i=1}^k \lambda_i^\uparrow(\rho_A) \geq \sum_{i=1}^k \lambda_i^\uparrow(\sigma_A). \quad (11.6)$$

Note that using Eq. (11.5), the right-hand side can be written as

$$\sum_{i=1}^k \lambda_i^\uparrow(\sigma_A) = \sum_{x \in \Omega} \sum_{i=1}^k \lambda_i^\uparrow(Y_x \sqrt{\rho_A} Z_x^T \bar{Z}_x \sqrt{\rho_A} Y_x^\dagger) \leq \sum_{x \in \Omega} \text{Tr}[\Pi_x Y_x \sqrt{\rho_A} Z_x^T \bar{Z}_x \sqrt{\rho_A} Y_x^\dagger]$$

for any choice of orthogonal projections Π_x with $\text{rank } \Pi_x \geq k$. This follows by the variational characterization of sum of the smallest k eigenvalues, which you can prove in Exercise 11.6 in case you have never seen it. Let us choose Π_x as the projection onto the orthogonal complement of $V_x := \text{span} \{Y_x | i\rangle\}_{i=k+1}^n$ (since $\dim V_x \leq n - k$ we indeed have $\text{rank } \Pi_x \geq k$). Then, defining

$$\tilde{\rho}_{AB} = |\tilde{\Psi}_{AB}\rangle\langle\tilde{\Psi}_{AB}|, \quad \text{where} \quad |\tilde{\Psi}_{AB}\rangle = \sum_{i=1}^k \sqrt{\lambda_i^\uparrow(\rho_A)} |ii\rangle,$$

we find that

$$\begin{aligned} \sum_{x \in \Omega} \text{Tr}[\Pi_x Y_x \sqrt{\rho_A} Z_x^T \bar{Z}_x \sqrt{\rho_A} Y_x^\dagger] &= \sum_{x \in \Omega} \text{Tr}[\Pi_x Y_x \sqrt{\tilde{\rho}_A} Z_x^T \bar{Z}_x \sqrt{\tilde{\rho}_A} Y_x^\dagger] \\ &\leq \sum_{x \in \Omega} \text{Tr}[Y_x \sqrt{\tilde{\rho}_A} Z_x^T \bar{Z}_x \sqrt{\tilde{\rho}_A} Y_x^\dagger] \\ &= \sum_{x \in \Omega} \text{Tr}[(Y_x \otimes Z_x) \tilde{\rho}_{AB} (Y_x^\dagger \otimes Z_x^\dagger)] \\ &= \text{Tr} \Xi[\tilde{\rho}_{AB}] = \text{Tr} \tilde{\rho}_{AB} = \sum_{i=1}^k \lambda_i^\uparrow(\rho_A), \end{aligned}$$

where we left out the projections Π_x in the second step, the third step follows from Lemma 3.7 (a) (as before), and the penultimate step holds since Ξ is trace-preserving. Thus we have proved Eq. (11.6), thus $\rho_A \prec \sigma_A$. This concludes the proof of (d) \Rightarrow (a) and hence of the theorem. \square

Nielsen's theorem can be generalized to the situation where the two states can be on different Hilbert spaces. We formulate this as a corollary, which you get to prove in Exercise 11.8.

Corollary 11.20. *Let ρ_{AB} be a pure state on $\mathcal{H}_A \otimes \mathcal{H}_B$ and let $\sigma_{A'B'}$ be a pure state on $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. Then the following are equivalent:*

- (a) $(\lambda(\rho_A), 0, \dots, 0) \prec (\lambda(\sigma_{A'}), 0, \dots, 0)$, where the notation means that we pad the two vectors by any/some suitable number of zeros such that both vectors have the same length.
- (a') $(\lambda(\rho_B), 0, \dots, 0) \prec (\lambda(\sigma_{B'}), 0, \dots, 0)$, with the same notation as above.
- (b) There exists a one-way LOCC channel Ξ from Alice to Bob such that $\Xi[\rho_{AB}] = \sigma_{A'B'}$.
- (b') There exists a one-way LOCC channel Ξ from Bob to Alice such that $\Xi[\rho_{AB}] = \sigma_{A'B'}$.

- (c) There exists a channel $\Xi \in \text{LOCC}(\mathcal{H}_A:\mathcal{H}_B, \mathcal{H}_{A'}:\mathcal{H}_{B'})$ such that $\Xi[\rho_{AB}] = \sigma_{A'B'}$.
- (d) There exists a channel $\Xi \in \text{SepC}(\mathcal{H}_A:\mathcal{H}_B, \mathcal{H}_{A'}:\mathcal{H}_{B'})$ such that $\Xi[\rho_{AB}] = \sigma_{A'B'}$.

You can practice Nielsen's theorem (and its proof) in Exercises 11.7, 11.9, 11.10 and 11.12.

11.4 Exercises

11.1 Majorization warmup:

- (a) Find probability distributions p and q such that neither $p \prec q$ nor $q \prec p$.
- (b) If p and q are probability distributions such that $p \prec q$ and $q \prec p$, is it necessarily the case that $p = q$?

11.2 Alternative definitions of majorization: For a vector $x \in \mathbb{R}^n$, recall that x^\downarrow denotes the vector with the same entries but in nonincreasing order. Let us also denote by x^\uparrow denote the vector with the same entries but in nondecreasing order. Show that the following are equivalent for probability distributions p and q on $\{1, \dots, n\}$:

- (a) $\sum_{i=1}^k p_i^\downarrow \leq \sum_{i=1}^k q_i^\downarrow$ for all $k \in \{1, \dots, n-1\}$.
- (b) $\sum_{i=1}^k p_i^\uparrow \geq \sum_{i=1}^k q_i^\uparrow$ for all $k \in \{1, \dots, n-1\}$.
- (c) $\sum_{i=1}^n \max(p_i - t, 0) \leq \sum_{i=1}^n \max(q_i - t, 0)$ for all $t \in \mathbb{R}$.

As this exercise is used in the proof of Lemma 11.10, please give a self-contained proof that does not rely on Lemma 11.10.

11.3 Majorization examples: Let $p = (0.1, 0.7, 0.2)$ and $q = (0.3, 0.2, 0.5)$.

- (a) One of $p \prec q$ or $q \prec p$ is true. Determine which.
- (b) Find a sequence of Robin-Hood matrices that convert one distribution into the other.
- (c) Express this sequence as a single stochastic matrix and verify that this matrix is in fact doubly stochastic.
- (d) Express this matrix as a convex combination of permutations.

Hint: For (b), you can follow the proof of Theorem 11.7, but it is much easier to observe that p and q have a common entry and use this to reduce directly to the 2×2 case.

11.4 Majorization and rank:

- (a) Show that if $A, B \in \text{PSD}(\mathcal{H})$ are such that $A \prec B$, then $\text{rank } A \geq \text{rank } B$.
- (b) Is the same true for general Hermitian operators?

Note that part (a) together with Nielsen's theorem reproves Theorem 10.11 for pure states.

11.5 Majorization vs mixed-unitary channels: Prove Theorem 11.14.

Hint: You can reduce to the case that A and B are real diagonal matrices.

11.6 Variational characterization of eigenvalues: Let $H \in \text{L}(\mathcal{H})$ be a Hermitian operator and let $1 \leq k \leq \dim \mathcal{H}$. For any subspace $V \subseteq \mathcal{H}$, denote by Π_V the orthogonal projection.

- (a) Show that $\sum_{i=1}^k \lambda_i^\uparrow(H) = \min_{\dim V=k} \text{Tr}[\Pi_V H]$ and $\sum_{i=1}^k \lambda_i^\downarrow(H) = \max_{\dim V=k} \text{Tr}[\Pi_V H]$.
- (b) Show that $\sum_{i=1}^k \lambda_i^\uparrow(H) = \min_{\dim V \geq k} \text{Tr}[\Pi_V H]$ and $\sum_{i=1}^k \lambda_i^\downarrow(H) = \max_{\dim V \leq k} \text{Tr}[\Pi_V H]$ if H is PSD.

You can also use part (a) to prove the following *Schur-Horn inequalities*:

(c) Show that $\sum_{i=1}^k \lambda_i^\uparrow(H) \leq \sum_{i=1}^k \langle i|H|i \rangle \leq \sum_{i=1}^k \lambda_i^\downarrow(H)$.

In particular, the diagonal entries of a Hermitian matrix are majorized by its eigenvalues.

11.7 Entanglement entropy and LOCC channels: Show that, for pure states, the entanglement entropy is nonincreasing under separable channels (so in particular under LOCC).

11.8 Corollary of Nielsen's Theorem: Prove Corollary 11.20.

11.9 Nielsen action I: For any $p \in [0, 1]$, find a one-way LOCC channel from Alice to Bob that transforms the maximally entangled state $|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ into the pure state $|\Psi_{AB}\rangle = \sqrt{p}|00\rangle + \sqrt{1-p}|11\rangle$. Write down Alice's instrument and Bob's channels explicitly.

Hint: $\frac{1}{2} \begin{pmatrix} p \\ 1-p \end{pmatrix} + \frac{1}{2} X \begin{pmatrix} p \\ 1-p \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$, where $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

11.10 Nielsen action II: For any probability distribution $p \in P(\{1, \dots, n\})$, find a one-way LOCC channel from Alice to Bob that transforms the maximally entangled state $|\Phi_{AB}^+\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^n |ii\rangle$ into the pure state $|\Psi_{AB}\rangle = \sum_{i=1}^n \sqrt{p(i)} |ii\rangle$. Write down Alice's instrument and Bob's channels explicitly.

Hint: Try to generalize your solution to Exercise 11.9.

11.11 Converting maximally entangled states: Let $\rho_{AB}, \sigma_{A'B'}$ denote two *maximally entangled* states of local dimensions $d = \dim \mathcal{H}_A = \dim \mathcal{H}_B$, $d' = \dim \mathcal{H}_{A'} = \dim \mathcal{H}_{B'}$, respectively.

(a) When can ρ_{AB} be converted to $\sigma_{A'B'}$ by LOCC?

Now assume that $d' = 2$.

(b) How many copies of $\sigma_{A'B'}$ can be obtained from ρ_{AB} by LOCC?

(c) How many copies of $\sigma_{A'B'}$ that are required so that ρ_{AB} can be obtained by LOCC?

11.12 Practice: Implement a subroutine that given two probability distributions p and q determines whether $p \prec q$, $q \prec p$, or neither of the two statements hold. Then use your subroutine to investigate the following:

(a) The file `11-probabilities.txt` contains three probability distributions: p_X, q_X , and p_Y . Compare the distributions p_X and q_X using your subroutine and output "`p < q`", "`q < p`", or "`incomparable`".

(b) Now let p_{XY} and q_{XY} be the product distributions given by $p_{XY}(x, y) = p_X(x)p_Y(y)$ and $q_{XY}(x, y) = q_X(x)p_Y(y)$. Use your subroutine to compare the product distributions p_{XY} and q_{XY} . Output "`p_xy < q_xy`", "`q_xy < p_xy`", or "`incomparable`".

(c) How can you interpret this outcome?

(d) The files `11-psi1.txt` and `11-psi2.txt` contain bipartite pure states

$$|\Psi_{AB,1}\rangle \in \mathbb{C}^5 \otimes \mathbb{C}^7 \quad \text{and} \quad |\Psi_{AB',2}\rangle \in \mathbb{C}^5 \otimes \mathbb{C}^9,$$

where Alice's dimension is 5 and Bob's dimensions are 7 and 9, respectively. Output the eigenvalues of the reduced states on Alice's system A and determine whether $|\Psi_{AB,1}\rangle$ can be transformed by LOCC into $|\Psi_{AB',2}\rangle$.

Lecture 12

Distillable entanglement and entanglement cost

Last week we looked at majorization and Nielsen's theorem. Majorization provides a way to compare probability distributions in terms of how uniform they are. Intuitively, $p \prec q$ if p is more uniform than q . By Theorem 11.7, this is the case if and only if p can be obtained from q by a doubly stochastic matrix or a product of Robin-Hood matrices (taking away from the rich and giving to the poor). We saw that all these notions have natural quantum counterparts, where probability distributions are replaced by density matrices. Using this more general notion of majorization, which can be defined for arbitrary Hermitian operators, Nielsen's Theorem 11.18 provides an elegant answer to the following problem: when can one pure state ρ_{AB} be converted to another pure state σ_{AB} by LOCC? Nielsen's theorem asserts that this is possible if and only if $\rho_A \prec \sigma_A$. This condition is easy to check by computing the eigenvalues of the reduced states and checking whether they obey the majorization conditions. Corollary 11.20 generalizes this result to states on different Hilbert spaces.

When ρ_{AB} can be converted to σ_{AB} by LOCC then we can interpret this as saying that ρ_{AB} is more entangled than σ_{AB} , because local operations and classical communication should not be able to create more entanglement. Thus, one can think of Nielsen's theorem as a way to compare the amount of entanglement in different states. However, not every pair of states is comparable, as follows from the analogous fact for probability distributions (Exercise 11.1). Moreover, even when two states are comparable, Nielsen's theorem is not quantitative in the sense that it does not tell us "how much more entangled" one state is compared to the other.

Ideally, we would like to assign a single number to every state telling us how much entanglement the state has. This number should be easy to compute and also have some precise operational interpretation. In this lecture we will see how this can be done for bipartite pure states.

12.1 Conversion, distillable entanglement, entanglement cost

A convenient way to measure the amount of entanglement for bipartite states is to choose a "gold standard" state and ask how many of these states can be obtained from the given state by LOCC. The canonical maximally entangled two-qubit state is a natural choice of such a "gold standard". Throughout today's lecture we will denote it by

$$\phi_{A'B'}^+ = |\Phi_{A'B'}^+\rangle\langle\Phi_{A'B'}^+|, \quad \text{where} \quad |\Phi_{A'B'}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad \text{and} \quad \mathcal{H}_{A'} = \mathcal{H}_{B'} = \mathbb{C}^2.$$

Thus we might ask how many copies of ϕ^+ can be created by LOCC when Alice and Bob start out by sharing a given state ρ_{AB} . In other words, we might try to quantify the amount of

entanglement in a given state ρ_{AB} as the maximal number N such that

$$\rho_{AB} \xrightarrow{\text{LOCC}} (\phi_{A'B'}^+)^{\otimes N}.$$

However, since the number of copies N is always an integer, this would only give a rather coarse measure of entanglement. For example, it is not hard to check using Nielsen's theorem that if ρ_{AB} is an arbitrary two-qubit pure state then $N \in \{0, 1\}$, with $N = 1$ if and only if ρ_{AB} is maximally entangled. To get a more nuanced quantity, it is useful to look at *conversion rates*. Namely, we might ask for the optimal rate $R > 0$ such that we can convert

$$\rho_{AB}^{\otimes n} \xrightarrow{\text{LOCC}} (\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}$$

for large n . Note that R no longer needs to be a integer! We need to make one more modification to develop a good theory. Namely, instead of asking for exact conversion, we will only require that we can by LOCC obtain a state that is arbitrary close to the target state as $n \rightarrow \infty$.¹ The quantity defined in this way called the *distillable entanglement*, since “distillation” refers to a process by which a large amount of an impure substance is refined to a smaller amount of a more concentrated and pure substance. We now define it formally:

Definition 12.1 (Distillable entanglement). The *distillable entanglement* $E_D(A : B)_\rho$ of a state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is defined as the supremum over all $R \geq 0$ for which there are LOCC channels

$$\Xi_n \in \text{LOCC}(\mathcal{H}_A^{\otimes n} : \mathcal{H}_B^{\otimes n}, \mathcal{H}_A^{\otimes \lfloor Rn \rfloor} : \mathcal{H}_B^{\otimes \lfloor Rn \rfloor})$$

for sufficiently large $n \in \mathbb{N}$ such that

$$\lim_{n \rightarrow \infty} F(\Xi_n[\rho_{AB}^{\otimes n}], (\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}) = 1.$$

When the state is clear from context, we will often write $E_D(A : B)$ in place of $E_D(A : B)_\rho$ (just like for entropy and mutual information).

Note that formally $\rho_{AB}^{\otimes n}$ is a state on the Hilbert space

$$(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes n} = (\mathcal{H}_A \otimes \mathcal{H}_B) \otimes \cdots \otimes (\mathcal{H}_A \otimes \mathcal{H}_B).$$

However, we group together all A systems and all B systems and think of $\rho_{AB}^{\otimes n}$ as a state on

$$\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n} = \underbrace{(\mathcal{H}_A \otimes \cdots \otimes \mathcal{H}_A)}_{\text{Alice}} \otimes \underbrace{(\mathcal{H}_B \otimes \cdots \otimes \mathcal{H}_B)}_{\text{Bob}}.$$

In principle, this identification amounts to a unitary, but we will not write it explicitly since it will always be clear from context and our subscript notation helps avoid any ambiguities. The same discussion applies to $(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}$.

Instead of distilling maximal entanglement, one can also ask the reverse question – how many copies of the maximally entangled state ϕ^+ are required to approximately produce some large number of copies of some desired state? This is called the *entanglement cost* and is formally defined as follows:

¹In Lectures 6 and 7 we saw something completely analogous in the context of compression. There, we found that a good theory can be developed if we compress blocks of n symbols at a time and allow for a small probability of error (that can be taken to go to zero as n becomes large).

Definition 12.2 (Entanglement cost). The *entanglement cost* $E_C(A : B)_\rho$ of a state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is defined as the infimum over all $R \geq 0$ for which there are LOCC channels

$$\Gamma_n \in \text{LOCC}(\mathcal{H}_{A'}^{\otimes \lfloor Rn \rfloor} : \mathcal{H}_{B'}^{\otimes \lfloor Rn \rfloor}, \mathcal{H}_A^{\otimes n} : \mathcal{H}_B^{\otimes n})$$

for sufficiently large $n \in \mathbb{N}$ such that

$$\lim_{n \rightarrow \infty} F(\Gamma_n[(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}], \rho_{AB}^{\otimes n}) = 1.$$

When the state is clear from context, we will often write $E_C(A : B)$ in place of $E_C(A : B)_\rho$.

How do distillable entanglement and entanglement cost compare in general? Think of the following analogy: if you go to a currency exchange to exchange money, the buying rate is always lower than the selling rate. If this were not the case, you could make money by repeatedly exchanging it back and forth. But there is no such thing as a free lunch! Similarly, one should not be able to obtain an increasingly large amount of entanglement by repeatedly distilling and then recreating a state by LOCC. This should be as impossible as constructing a perpetual motion machine that keeps generating energy for free. However, proving this formally is actually not so trivial and relies on the following lemma, which you will prove in Exercise 12.1. We will formulate it in a slightly more general situation than need be:

Lemma 12.3. Let $\mathcal{H}_C = \mathcal{H}_D = \mathbb{C}^d$, let σ_{CD} be a state of entanglement rank r , and let ω_{CD} be a maximally entangled state. Then, $F(\sigma_{CD}, \omega_{CD}) \leq \sqrt{\frac{r}{d}}$.

We can use this to prove that the entanglement cost is indeed at least as large as the distillable entanglement:

Corollary 12.4 (No free lunch). $E_C(A : B)_\rho \geq E_D(A : B)_\rho$ for any state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

Proof. Let R be such that there exist LOCC channels Γ_n such that

$$\lim_{n \rightarrow \infty} F(\Gamma_n[(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}], \rho_{AB}^{\otimes n}) = 1.$$

Likewise, let R' be such that there exist LOCC channels Ξ_n such that

$$\lim_{n \rightarrow \infty} F(\Xi_n[\rho_{AB}^{\otimes n}], (\phi_{A'B'}^+)^{\otimes \lfloor R'n \rfloor}) = 1.$$

Since E_C is the infimum over all rates R as above and E_D is the supremum over all rates R' as above, it suffices to prove that $R \geq R'$. To prove the latter, note that, by Exercise 5.9,

$$\lim_{n \rightarrow \infty} F((\Xi_n \circ \Gamma_n)[(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}], (\phi_{A'B'}^+)^{\otimes \lfloor R'n \rfloor}) = 1. \quad (12.1)$$

On the other hand, the entanglement rank of $(\Xi_n \circ \Gamma_n)[(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}]$ is at most $2^{\lfloor Rn \rfloor}$, since this is the entanglement (Schmidt) rank of $\phi_{A'B'}^+$ and $\Xi_n \circ \Gamma_n$ is LOCC. Therefore, using Lemma 12.3,

$$F((\Xi_n \circ \Gamma_n)[(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}], (\phi_{A'B'}^+)^{\otimes \lfloor R'n \rfloor})^2 \leq 2^{\lfloor Rn \rfloor - \lfloor R'n \rfloor} \leq 2^{Rn - (R'n - 1)} = 2^{(R - R')n + 1}.$$

Since this holds for sufficiently large n , Eq. (12.1) implies that $R \geq R'$, concluding the proof. \square

By definition, E_C and E_D are meaningful quantities, but how can we compute them explicitly for a given state ρ_{AB} ? In general there is no closed formula, but we will prove in Section 12.2 that for bipartite pure states they are both equal to the entanglement entropy!

In light of Corollary 12.4, it is tempting to ask if E_C and E_D are also equal for mixed states. Surprisingly, the answer is “No!” – there are mixed states for which $E_C > E_D$. In fact, there even exist mixed states such that $E_C > 0$ but $E_D = 0$. Such states are called *bound entangled* because the entanglement in them is bound or confined within them and cannot be extracted back. This is what makes mixed state entanglement so much more interesting and also difficult! For example, bound entangled states are responsible for such strange phenomena as *superactivation* of quantum channels – the counterintuitive fact that two zero-capacity quantum channels can have positive capacity when used together in parallel. Unfortunately, we will not have time to go into these interesting but more advanced topics.

12.2 Distillable entanglement equals entanglement cost for pure states

We will now prove that, for pure states, distillable entanglement and entanglement cost are the same, and equal to entanglement entropy (Definition 8.2).

Theorem 12.5. *For any pure state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we have*

$$E_D(A : B)_\rho = E_C(A : B)_\rho = H(A)_\rho = H(B)_\rho.$$

Proof. We already know that $E_D(A : B) \leq E_C(A : B)$, so it suffices to prove that

$$E_C(A : B) \leq H(A) \leq E_D(A : B).$$

The key idea is very simple: we replace $\rho_{AB}^{\otimes n}$ by the pure state

$$\tilde{\rho}_{A^n B^n} = \frac{(\Pi_{n,\varepsilon} \otimes I_{B^n}) \rho_{AB}^{\otimes n} (\Pi_{n,\varepsilon} \otimes I_{B^n})}{p_{n,\varepsilon}},$$

where $\Pi_{n,\varepsilon}$ denotes a typical projector for ρ_A (Definition 7.11) and $p_{n,\varepsilon} := \text{Tr}[\Pi_{n,\varepsilon} \rho_A^{\otimes n}]$, and apply Nielsen’s theorem. For fixed $\varepsilon > 0$ and large n , part (c) of the quantum AEP (Lemma 7.12) states that $p_{n,\varepsilon} \rightarrow 1$, so using the gentle measurement lemma (Exercise 4.9) it follows that

$$\lim_{n \rightarrow \infty} F(\tilde{\rho}_{A^n B^n}, \rho_{AB}^{\otimes n}) = 1. \quad (12.2)$$

Moreover, part (a) asserts that the nonzero eigenvalue of $\tilde{\rho}_{A^n}$ are within $\frac{2^{-n(H(A) \pm \varepsilon)}}{p_{n,\varepsilon}}$, which makes it easy to apply Nielsen’s theorem. We now make this idea precise and prove the two inequalities.

Let us first show that $E_C(A : B) \leq H(A)$. Let $R > H(A)$ and choose $\varepsilon > 0$ such that $R \geq H(A) + 2\varepsilon$. Then we have for all $n \geq \frac{1}{\varepsilon}$ that

$$\lfloor nR \rfloor \geq nR - 1 \geq n(H(A) + 2\varepsilon) - 1 \geq n(H(A) + \varepsilon)$$

and hence

$$2^{-\lfloor nR \rfloor} \leq 2^{-n(H(A) + \varepsilon)} \leq \frac{2^{-n(H(A) + \varepsilon)}}{p_{n,\varepsilon}} \leq \lambda$$

for any nonzero eigenvalue λ of $\tilde{\rho}_{A^n}$. This implies that the eigenvalues of $(\phi_{A'}^+)^{\otimes \lfloor nR \rfloor} = \frac{I}{2^{\lfloor nR \rfloor}}$ are majorized by the eigenvalues of $\tilde{\rho}_{A^n}$ (padded by sufficiently many zeros). Thus Corollary 11.20 shows that there exists an LOCC channel Γ_n such that $\Gamma_n[(\phi_{A'B'}^+)^{\otimes \lfloor nR \rfloor}] = \tilde{\rho}_{A^n B^n}$. In view of Eq. (12.2), this shows that $E_C(A : B) \leq R$. Since $R > H(A)$ was arbitrary, we conclude that $E_C(A : B) \leq H(A)$.

We now show that $H(A) < E_D(A : B)$. Let $R < H(A)$ and choose $\varepsilon > 0$ such that $R \leq H(A) - 2\varepsilon$. Then we have for all $n \geq -\frac{1}{\varepsilon} \log(1 - \varepsilon)$ that

$$\lfloor nR \rfloor \leq nR \leq n(H(A) - 2\varepsilon) \leq n(H(A) - \varepsilon) + \log(1 - \varepsilon)$$

and hence

$$\lambda \leq \frac{2^{-n(H(A) - \varepsilon)}}{p_{n,\varepsilon}} \leq \frac{2^{-n(H(A) - \varepsilon)}}{1 - \varepsilon} \leq 2^{-\lfloor nR \rfloor}$$

for any eigenvalue of $\tilde{\rho}_{A^n}$ provided n is sufficiently large since we have $p_{n,\varepsilon} \rightarrow 1$ for fixed $\varepsilon > 0$. This implies that the eigenvalues of $\tilde{\rho}_{A^n}$ are majorized by the eigenvalues of $(\phi_{A'}^+)^{\otimes \lfloor nR \rfloor} = \frac{I}{2^{\lfloor nR \rfloor}}$ (padded by sufficiently many zeros). Thus Corollary 11.20 shows that there exists an LOCC channel Ξ_n such that $\Xi_n[\tilde{\rho}_{A^n B^n}] = (\phi_{A'B'}^+)^{\otimes \lfloor nR \rfloor}$. Then we have

$$\liminf_{n \rightarrow \infty} F(\Xi_n[\rho_{AB}^{\otimes n}], (\phi_{A'B'}^+)^{\otimes \lfloor nR \rfloor}) = \liminf_{n \rightarrow \infty} F(\Xi_n[\rho_{AB}^{\otimes n}], \Xi_n[\tilde{\rho}_{A^n B^n}]) \geq \lim_{n \rightarrow \infty} F(\rho_{AB}^{\otimes n}, \tilde{\rho}_{A^n B^n}) = 1,$$

where the last equality is Eq. (12.2). Since fidelities are also upper bounded by 1, it follows that

$$\lim_{n \rightarrow \infty} F(\Xi_n[\rho_{AB}^{\otimes n}], (\phi_{A'B'}^+)^{\otimes \lfloor nR \rfloor}) = 1.$$

This shows that $E_D(A : B) \geq R$. Since $R < H(A)$ was arbitrary, we conclude that $E_D(A : B) \geq H(A)$. \square

12.3 Exercises

12.1 Entanglement rank and fidelity: Prove Lemma 12.3.

Hint: First show the claim assuming σ_{CD} is pure.

12.2 From any pure state to any other: Let $\rho \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ and $\sigma \in D(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ be two arbitrary pure states. What is the optimal rate of conversion from copies of ρ_{AB} to copies of $\sigma_{A'B'}$ by LOCC (if we allow for arbitrarily small error)?

12.3 Entanglement cost using compression and teleportation:

In this exercise you can find an alternative proof of the fact that $E_C(A : B) \leq H(A)$ for any pure state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$.

- (a) Let $R > H(A) = H(B)$ and $\delta > 0$. Use Schumacher's Theorem 7.10 and teleportation to show that there exists n_0 such that, for all $n \geq n_0$, there is an LOCC protocol which converts $(\phi_{A'B'}^+)^{\otimes \lfloor Rn \rfloor}$ into a state $\tilde{\rho}_{A^n B^n}$ with $F(\rho_{AB}^{\otimes n}, \tilde{\rho}_{A^n B^n}) \geq 1 - \delta$.
- (b) Use part (a) to show that $E_C(A : B)_\rho \leq H(A)_\rho$.

Lecture 13

Monogamy of entanglement

Last week, we looked at ways to quantify entanglement. Using the canonical two-qubit maximally entangled state as a “golden standard”, we asked how many copies of it can be extracted from a given state, or how many copies are needed to produce a given state by LOCC. In Theorem 12.5, we showed that for pure states these two quantities – distillable entanglement and entanglement cost – coincide and are equal to entanglement entropy.

In this lecture, we will take a different approach to entanglement. Instead of the usual bipartite setting, we will consider multiple parties and observe a curious property of entanglement known as monogamy – namely, one cannot simultaneously share a large amount of entanglement with multiple parties. Using this observation, we will draw a non-trivial conclusion about bipartite entanglement. Namely, a bipartite state that admits a symmetric extension to a multipartite setting cannot be too entangled, otherwise the extended state would violate monogamy.

13.1 Sharing classical vs quantum correlations

You showed in Exercise 2.10 (a) that if ρ_{ABC} is a state such that ρ_{AB} is pure, then $\rho_{ABC} = \rho_{AB} \otimes \rho_C$. In particular, this implies that $\rho_{AC} = \rho_A \otimes \rho_C$ and $\rho_{BC} = \rho_B \otimes \rho_C$, meaning that A and C are not correlated, and neither are B and C . Hence, one cannot share a *pure* entangled state with more than one system – this is known as *monogamy of entanglement*.

$\rho_{ABC} =$

Diagram illustrating a quantum state ρ_{ABC} . The state is represented by three nodes: A , B , and C . Node A is connected to node B by a horizontal line, with the label $|\Psi_{AB}\rangle$ above the line. Node C is positioned below the line connecting A and B , and is associated with the label ρ_C .

In contrast, a classical state that is maximally correlated can be shared with an arbitrary number of parties. To see this, consider the tripartite classical state

$$\rho_{ABC} = \frac{1}{2}(|000\rangle\langle 000| + |111\rangle\langle 111|), \quad (13.1)$$

which corresponds to flipping an unbiased coin and telling the outcome to all three parties. Note that any two parties are maximally correlated since

$$\rho_{AB} = \rho_{BC} = \rho_{AC} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|). \quad (13.2)$$

Moreover, one can easily distribute this correlation to additional parties by attaching a fresh qubit in state $|0\rangle$ and performing a CNOT operation with this qubit as a target.

Can we also do this in the quantum case? Let's see what happens if we try to use the same approach. The natural equivalent of Eq. (13.1) for pure states is

$$|\Psi_{ABC}\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle).$$

Are all pairs of parties in this state maximally entangled? For this to be the case, all two-party reduced states of $|\Psi_{ABC}\rangle$ should be $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ or, written as a density matrix,

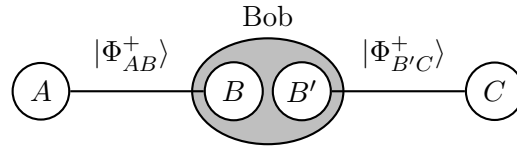
$$|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|).$$

However, the actual two-party reduced states of $|\Psi_{ABC}\rangle$ are exactly the same as in Eq. (13.2):

$$\rho_{AB} = \rho_{BC} = \rho_{AC} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|),$$

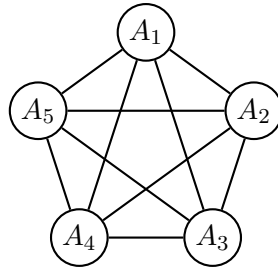
meaning that each pair of parties shares a maximal *classical* correlation, not a maximally entangled state. This is a very stark manifestation of monogamy of entanglement – we were hoping for all pairs of parties to be maximally entangled, while in reality each pair shares a separable state that has no entanglement whatsoever! This means the state $|\Psi_{ABC}\rangle$ possesses global tripartite entanglement while having no bipartite entanglement.

The only way for one party to simultaneously share a maximally entangled state with two other parties is by increasing the dimension of its system from one to two qubits. For example, if Bob has two qubits B and B' , he can share a maximally entangled with A and C as follows:



Note that in this case the B system is completely uncorrelated with C , and B' is completely uncorrelated with A . Another interesting observation is that if we increase the dimension of system A , Bob cannot share more entanglement with it while still maintaining a maximally entangled state with C . For example, if we increased the system A to two qubits, Bob could now share two qubits of entanglement with A , however he would have to completely give up all entanglement with C .

Let us now consider a more complicated situation with n parties denoted by A_1, \dots, A_n . Assume their joint state $\rho_{A_1 \dots A_n}$ is such that all two-party reduced states $\rho_{A_i A_j}$ are the same for all $i \neq j$. Let us denote this reduced state by ρ_{AB} and call $\rho_{A_1 \dots A_n}$ its *symmetric extension*.



Intuitively, ρ_{AB} should not be too entangled because each party of its symmetric extension shares this state with the remaining $n - 1$ other parties.

The goal of this lecture is to prove Theorem 13.13 which establishes a quantitative bound on how close ρ_{AB} is to the set of separable states, given that it has such a highly symmetric n -party

extension $\rho_{A_1 \dots A_n}$ (results of this form are known as *de Finetti* theorems). More specifically, we will show that the distance between ρ_{AB} and the set of separable states is upper bounded by

$$\sqrt{\frac{2d}{n+2}},$$

where d is the dimension of each system. Hence, the larger symmetric extension of ρ_{AB} we can find (i.e., the more parties n it has), the closer ρ_{AB} must be to a separable state. This provides a way to study bipartite entanglement through the lens of multipartite states.

13.2 The symmetric subspace

Since the setup of our de Finetti theorem involves a state with a high degree of symmetry, we first need to develop the mathematical machinery for dealing with such states. These states live in the so-called *symmetric subspace*.

Definition 13.1 (Symmetric subspace). Let \mathcal{H} be a Hilbert space, let $n \geq 1$, and let S_n denote the set of all permutations acting on $\{1, \dots, n\}$. For every $\pi \in S_n$, let $R_\pi \in U(\mathcal{H}^{\otimes n})$ denote the operator that acts on n systems and permutes them according to π :

$$R_\pi(|\psi_1\rangle \otimes \dots \otimes |\psi_n\rangle) := |\psi_{\pi^{-1}(1)}\rangle \otimes \dots \otimes |\psi_{\pi^{-1}(n)}\rangle, \quad (13.3)$$

for all $|\psi_1\rangle, \dots, |\psi_n\rangle \in \mathcal{H}$. The *symmetric subspace* of $\mathcal{H}^{\otimes n}$ is then defined as

$$\text{Sym}^n(\mathcal{H}) := \{|\Phi\rangle \in \mathcal{H}^{\otimes n} : R_\pi|\Phi\rangle = |\Phi\rangle, \forall \pi \in S_n\}.$$

Remark 13.2. The reason for using π^{-1} instead of π on the right-hand side of Eq. (13.3) is so that $R_\pi R_\tau = R_{\pi\tau}$ for all $\pi, \tau \in S_n$, which makes the map $\pi \mapsto R_\pi$ a unitary representation of the symmetric group S_n . You will show this and also the fact that $R_\pi^\dagger = R_{\pi^{-1}}$ for all $\pi \in S_n$ in Exercise 13.1 (c).

Remark 13.3. Make sure to not confuse R_π with the permutation matrix P_π from Definition 11.3! The distinction is that R_π permutes the systems while P_π permutes the standard basis states according to π . In particular, R_π is of size $d^n \times d^n$, where $d = \dim \mathcal{H}$, while P_π is of size $n \times n$.

Here are some basic observations about $\text{Sym}^n(\mathcal{H})$:

- For any $|\psi\rangle \in \mathcal{H}$, $|\psi\rangle^{\otimes n} \in \text{Sym}^n(\mathcal{H})$ since $R_\pi|\psi\rangle^{\otimes n} = |\psi\rangle^{\otimes n}$.
- For any $\pi \in S_n$, $|\Phi\rangle \in \text{Sym}^n(\mathcal{H})$ iff $R_\pi|\Phi\rangle \in \text{Sym}^n(\mathcal{H})$.
- If $|\Phi_1\rangle, |\Phi_2\rangle \in \text{Sym}^n(\mathcal{H})$ then $|\Phi_1\rangle + |\Phi_2\rangle \in \text{Sym}^n(\mathcal{H})$.

In other words, all tensor power states are in the symmetric subspace, the order in which the systems are arranged does not affect whether a state is symmetric or not, and the symmetric subspace is indeed a subspace.

Example 13.4 (Two qubits). When $n = 2$ and $d = 2$,

$$\text{Sym}^2(\mathbb{C}^2) = \text{span}\left\{|0,0\rangle, \frac{|0,1\rangle + |1,0\rangle}{\sqrt{2}}, |1,1\rangle\right\}.$$

The remaining vector $(|0,1\rangle - |1,0\rangle)/\sqrt{2}$ (also known as the *singlet state*) is anti-symmetric.

Since $\text{Sym}^n(\mathcal{H})$ is a subspace of $\mathcal{H}^{\otimes n}$, we can write down a projector onto this subspace. You will show in Exercise 13.1 (d) that the following is an orthogonal projection onto $\text{Sym}^n(\mathcal{H})$:

$$\Pi_n := \frac{1}{n!} \sum_{\pi \in S_n} R_\pi. \quad (13.4)$$

In particular, $\Pi_n^\dagger = \Pi_n$ and $\Pi_n^2 = \Pi_n$, i.e., Π_n is Hermitian and a projector. Moreover, $\Pi_n R_\pi = R_\pi \Pi_n = \Pi_n$ for any $\pi \in S_n$. Intuitively, applying Π_n to a state corresponds to “symmetrizing” it:

$$\Pi_n |\Phi\rangle = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi |\Phi\rangle.$$

As part of your argument in Exercise 13.1 (d) you will show that $\Pi_n |\Phi\rangle \in \text{Sym}^n(\mathcal{H})$, for any $|\Phi\rangle \in \mathcal{H}^{\otimes n}$. Moreover, if $|\Phi\rangle \in \text{Sym}^n(\mathcal{H})$ then $\Pi_n |\Phi\rangle = |\Phi\rangle$. In fact, for any $k \geq 0$ and $|\Phi\rangle \in \text{Sym}^{k+n}(\mathcal{H})$, the following more general identity holds:

$$(I_k \otimes \Pi_n) |\Phi\rangle = |\Phi\rangle, \quad (13.5)$$

where I_k denotes the identity operator on the first k systems.

Example 13.5 (Projector Π_2). For $n = 2$, it follows immediately from Eq. (13.4) that

$$\Pi_2 = \frac{1}{2}(I + F) = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

where $F \in \text{U}(\mathcal{H}^{\otimes 2})$ is the swap operator: $F(|\alpha\rangle \otimes |\beta\rangle) = |\beta\rangle \otimes |\alpha\rangle$, for all $|\alpha\rangle, |\beta\rangle \in \mathcal{H}$.

Recall from Example 13.4 that the symmetric subspace for two qubits is spanned by $|00\rangle$, $|11\rangle$, and $(|01\rangle + |10\rangle)/\sqrt{2}$. How can we find all states in $\text{Sym}^n(\mathbb{C}^d)$, for any $n \geq 1$ and $d \geq 1$? We can symmetrize the standard basis states, thus projecting them to the symmetric subspace!

Let $\Lambda_{n,d}$ denote the set of all integers $t_1, \dots, t_d \geq 0$ such that $\sum_{i=1}^d t_i = n$:

$$\Lambda_{n,d} := \left\{ (t_1, \dots, t_d) \in \mathbb{Z}^d : t_1, \dots, t_d \geq 0, \sum_{i=1}^d t_i = n \right\}.$$

For any $(t_1, \dots, t_d) \in \Lambda_{n,d}$, let $|T_{t_1, \dots, t_d}\rangle \in (\mathbb{C}^d)^{\otimes n}$ denote the following state:

$$|T_{t_1, \dots, t_d}\rangle := \underbrace{\overbrace{|1\rangle \otimes \dots \otimes |1\rangle}^{t_1} \otimes \overbrace{|2\rangle \otimes \dots \otimes |2\rangle}^{t_2} \otimes \dots \otimes \overbrace{|d\rangle \otimes \dots \otimes |d\rangle}^{t_d}}_n, \quad (13.6)$$

where t_i denotes the number of terms $|i\rangle$ occurring in the tensor product.

Example 13.6 (Two qubits). The set $\Lambda_{n,d}$ and the corresponding basis states for two qubits are

$$\Lambda_{2,2} = \{(2, 0), (1, 1), (0, 2)\}, \quad |T_{2,0}\rangle = |0, 0\rangle, \quad |T_{1,1}\rangle = |0, 1\rangle, \quad |T_{0,2}\rangle = |1, 1\rangle.$$

To match with the case of general d , one should use $\{|1\rangle, |2\rangle\}$ instead of $\{|0\rangle, |1\rangle\}$ for the qubit standard basis here. However, we used $|0\rangle$ and $|1\rangle$ to emphasize the correspondence with the states in Example 13.4.

We can get a basis for the symmetric subspace by symmetrizing the states $|T_{t_1, \dots, t_d}\rangle$.

Lemma 13.7 (Basis of symmetric subspace). *The following is an orthogonal basis for $\text{Sym}^n(\mathbb{C}^d)$:*

$$\text{Sym}^n(\mathbb{C}^d) = \text{span}\{\Pi_n |T_{t_1, \dots, t_d}\rangle : (t_1, \dots, t_d) \in \Lambda_{n,d}\}.$$

Proof. Since Π_n projects onto the symmetric subspace, we need to find the image of $(\mathbb{C}^d)^{\otimes n}$ under Π_n . For this, it suffices to apply Π_n to all standard basis vectors $|\Phi\rangle$ of $(\mathbb{C}^d)^{\otimes n}$. Since $\Pi_n R_\pi |\Phi\rangle = \Pi_n |\Phi\rangle$ for all $\pi \in S_n$, we can first permute the systems and sort the basis vectors in the tensor product expansion of $|\Phi\rangle$ to obtain one of the states $|T_{t_1, \dots, t_d}\rangle$. Since t_i counts the number of appearances of $|i\rangle$, you can think of the resulting sequence t_1, \dots, t_d as a “generalized Hamming weight” of the original string of basis vectors. To obtain a basis of $\text{Sym}^n(\mathbb{C}^d)$, it suffices to apply Π_n to all vectors $|T_{t_1, \dots, t_d}\rangle$ with $(t_1, \dots, t_d) \in \Lambda_{n,d}$. Note that all terms in the expansion of $\Pi_n |T_{t_1, \dots, t_d}\rangle$ have the same Hamming weight, and for different choices of t_1, \dots, t_d the Hamming weights are different. The resulting basis is orthogonal since all vectors have disjoint supports, i.e., their standard basis expansions do not contain a single common term. \square

Remark 13.8. While the states $\Pi_n |T_{t_1, \dots, t_d}\rangle$ with $(t_1, \dots, t_d) \in \Lambda_{n,d}$ are mutually orthogonal, they are not normalized in general since Π_n is a projector.

As a practice, you can work out the bases for $\text{Sym}^2(\mathbb{C}^d)$ and $\text{Sym}^3(\mathbb{C}^2)$ in Exercise 13.1 (b). Using Lemma 13.7, we can easily find the dimension of the symmetric subspace.

Lemma 13.9. *The dimension of the symmetric subspace is*

$$\dim(\text{Sym}^n(\mathbb{C}^d)) = |\Lambda_{n,d}| = \binom{n+d-1}{n} = \frac{(n+d-1)!}{n!(d-1)!}$$

Proof. Recall from Lemma 13.7 that the states $\Pi_n |T_{t_1, \dots, t_d}\rangle$ with $(t_1, \dots, t_d) \in \Lambda_{n,d}$ are mutually orthogonal since they have disjoint supports. Hence, the dimension of $\text{Sym}^n(\mathbb{C}^d)$ is equal to $|\Lambda_{n,d}|$. Note that $|\Lambda_{n,d}|$ is the number of ways of grouping n elements into d (possibly empty) groups. Using the method of [stars and bars](#), this can be determined by separating n stars with $d-1$ bars. This corresponds to choosing $d-1$ out of $n+d-1$ elements to be the bars and the remaining n to be stars, yielding the desired binomial coefficient. \square

We will need the following result, which we state without proof.

Lemma 13.10. *Let $A \in \mathcal{L}(\mathcal{H}^{\otimes n})$ for some Hilbert space \mathcal{H} and $n \geq 1$. Then*

$$U^{\otimes n} A U^{\dagger \otimes n} = A, \quad \forall U \in \mathcal{U}(\mathcal{H}),$$

iff $A = \sum_{\pi \in S_n} c_\pi R_\pi$, for some $c_\pi \in \mathbb{C}$.

Using this, we can provide an alternative expression for the projector Π_n defined in Eq. (13.4). Instead of a discrete sum over permutations, this expression involves a continuous integral with the uniform measure over pure quantum states. We will denote this measure by $d\psi$. Since it is uniform, the states $|\psi\rangle$ and $U|\psi\rangle$ have the same probability density, for any unitary U . This high degree of symmetry alone implies, for example, that $\int d\psi |\psi\rangle\langle\psi| = I/d$ (indeed, $M = I/d$ is the only matrix such that $UMU^\dagger = M$ for all $U \in \mathcal{U}(d)$ and $\text{Tr } M = 1$). A similar uniform measure can also be defined on the set $\mathcal{U}(d)$ of all unitary matrices (see Exercise 13.3).

Lemma 13.11. For any $n \geq 1$ and $d \geq 2$,

$$\Pi_n = \binom{n+d-1}{n} \int d\psi (|\psi\rangle\langle\psi|)^{\otimes n},$$

where $d\psi$ is the uniform probability measure on pure states in \mathbb{C}^d .

Proof. You will prove this in Exercise 13.2. □

Example 13.12 (Integral for Π_n when $d = 2$ and $n = 2$). The uniform measure for pure qubit states is the same as for the points on the unit sphere in \mathbb{R}^3 (a.k.a. the Bloch sphere, see Section 1.4):

$$d\psi = \frac{1}{4\pi} \sin \theta \, d\theta \, d\varphi,$$

where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ are the angles in the spherical coordinates. The corresponding point on the unit sphere in \mathbb{R}^3 has coordinates

$$(x, y, z) := (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta) \in \mathbb{R}^3.$$

This corresponds to the pure state

$$|\psi(\theta, \varphi)\rangle := \begin{pmatrix} \cos \frac{\theta}{2} \\ e^{i\varphi} \sin \frac{\theta}{2} \end{pmatrix} \in \mathbb{C}^2,$$

as can be seen by comparing the density matrix $\rho(\theta, \varphi) := |\psi(\theta, \varphi)\rangle\langle\psi(\theta, \varphi)|$ with

$$\rho(x, y, z) := \frac{1}{2}(I + xX + yY + zZ) = \frac{1}{2} \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}.$$

By explicitly evaluating the integral from Lemma 13.11 with $n = d = 2$ we get

$$\binom{2+2-1}{2} \frac{1}{4\pi} \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} \rho(\theta, \varphi)^{\otimes 2} \sin \theta \, d\theta \, d\varphi = \frac{1}{2} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} = \Pi_2,$$

which agrees with the matrix in Example 13.4.

13.3 The quantum de Finetti theorem

If $|\Phi_{A_1 \dots A_n}\rangle \in \text{Sym}^n(\mathcal{H})$ then all its two-party reduced density matrices $\Phi_{A_i A_j}$, $i \neq j$, are identical. If a given mixed state $\rho_{AA'}$ can be extended to such symmetric pure state $|\Phi_{A_1 \dots A_n}\rangle$, for some large value of n , then $\rho_{AA'}$ must be very close to separable (in fact, the distance goes to zero as $n \rightarrow \infty$). This is made rigorous by the quantum de Finetti theorem.

Theorem 13.13 (Quantum de Finetti theorem). Let $k \geq 1$, $n \geq 0$, and consider $k+n$ systems $A_1 \dots A_{k+n}$, each of dimension $d \geq 2$. For any $|\Phi\rangle \in \text{Sym}^{k+n}(\mathbb{C}^d)$, there exists a probability

density function p on pure states in \mathbb{C}^d such that

$$\frac{1}{2} \left\| \Phi_{A_1 \dots A_k} - \int d\psi p(\psi) (|\psi\rangle\langle\psi|)^{\otimes k} \right\|_1 \leq \sqrt{\frac{dk}{k+n}}, \quad (13.7)$$

where $\Phi_{A_1 \dots A_k} = \text{Tr}_{A_{k+1} \dots A_{k+n}} [|\Phi\rangle\langle\Phi|]$.

Remark 13.14. While the definition of bipartite separable states involves a finite sum, not an integral (see Definition 3.1), by [Carathéodory's theorem](#) an integral $\int dx p(x) \rho_A(x) \otimes \rho_B(x)$ can always be converted to a finite sum (as long as the dimensions of both systems are finite). In particular, one can always write it as $\sum_{i \in I} p_i \rho_{A,i} \otimes \rho_{B,i}$ where $|I| \leq r^2$ and r is the rank of the separable state that is being represented. A similar argument can also be used for multi-partite states to replace the integral in Eq. (13.7) by a finite sum.

Proof. Recall from Eq. (13.5) that $(I_{A_1 \dots A_k} \otimes \Pi_n)|\Phi\rangle = |\Phi\rangle$, so

$$\begin{aligned} \Phi_{A_1 \dots A_k} &= \text{Tr}_{A_{k+1} \dots A_{k+n}} [|\Phi\rangle\langle\Phi|] \\ &= \text{Tr}_{A_{k+1} \dots A_{k+n}} [(I_{A_1 \dots A_k} \otimes \Pi_n)|\Phi\rangle\langle\Phi|] \\ &= \binom{n+d-1}{n} \int d\psi \text{Tr}_{A_{k+1} \dots A_{k+n}} \left[(I_{A_1 \dots A_k} \otimes (|\psi\rangle^{\otimes n} \langle\psi|^{\otimes n})_{A_{k+1} \dots A_{k+n}}) |\Phi\rangle\langle\Phi| \right] \\ &= \binom{n+d-1}{n} \int d\psi (I_{A_1 \dots A_k} \otimes \langle\psi|^{\otimes n}) |\Phi\rangle\langle\Phi| (I_{A_1 \dots A_k} \otimes |\psi\rangle^{\otimes n}), \end{aligned} \quad (13.8)$$

where we substituted the integral formula for Π_n from Lemma 13.11 and then used the following cyclic property of the partial trace:

$$\text{Tr}_B [(I_A \otimes |\psi\rangle\langle\psi|_B) \Phi_{AB}] = (I_A \otimes \langle\psi|_B) \Phi_{AB} (I_A \otimes |\psi\rangle_B), \quad (13.9)$$

which holds for any $\Phi_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and unit vector $|\psi\rangle_B \in \mathcal{H}_B$ (this identity follows from Definition 2.8 by computing the partial trace in any basis that contains $|\psi\rangle$).

Let us rewrite the integral in Eq. (13.8) as follows:

$$\Phi_{A_1 \dots A_k} = \int d\psi p(\psi) |\Phi_\psi\rangle\langle\Phi_\psi|, \quad (13.10)$$

where $|\Phi_\psi\rangle_{A_1 \dots A_k} \in (\mathbb{C}^d)^{\otimes k}$ and p are such that

$$\sqrt{p(\psi)} |\Phi_\psi\rangle := \sqrt{\binom{n+d-1}{n}} (I_{A_1 \dots A_k} \otimes \langle\psi|^{\otimes n}) |\Phi\rangle. \quad (13.11)$$

If we rescale $|\Phi_\psi\rangle$ to a unit vector, p becomes a probability density function on pure states in \mathbb{C}^d , as can be seen by taking trace on both sides of Eq. (13.10). More specifically, $p(\psi)$ is given by

$$p(\psi) := \binom{n+d-1}{n} \|(I_{A_1 \dots A_k} \otimes \langle\psi|^{\otimes n}) |\Phi\rangle\|^2.$$

Let us compare the integral in Eq. (13.10) with $\tilde{\Phi}_{A_1 \dots A_k} = \int d\psi p(\psi) |\psi\rangle^{\otimes k} \langle\psi|^{\otimes k}$, where p is the same probability density function. Recall from Eq. (4.10) the formula $\frac{1}{2} \|\alpha\rangle\langle\alpha| - |\beta\rangle\langle\beta|\|_1 = \sqrt{1 - |\langle\alpha|\beta\rangle|^2}$ for the trace distance between pure states $|\alpha\rangle$ and $|\beta\rangle$. Using triangle inequality and then this formula,

$$\frac{1}{2} \left\| \Phi_{A_1 \dots A_k} - \tilde{\Phi}_{A_1 \dots A_k} \right\|_1 \leq \int d\psi p(\psi) \frac{1}{2} \left\| |\Phi_\psi\rangle\langle\Phi_\psi| - |\psi\rangle^{\otimes k} \langle\psi|^{\otimes k} \right\|_1$$

$$\begin{aligned}
&= \int d\psi p(\psi) \sqrt{1 - |\langle \psi |^{\otimes k} | \Phi_\psi \rangle|^2} \\
&\leq \sqrt{\int d\psi p(\psi) (1 - |\langle \psi |^{\otimes k} | \Phi_\psi \rangle|^2)} \\
&= \sqrt{1 - \int d\psi p(\psi) |\langle \psi |^{\otimes k} | \Phi_\psi \rangle|^2},
\end{aligned}$$

where we used Jensen's inequality [Eq. (6.4)] to bring the integral underneath the square root, a concave function.

For the rest of the proof, let us focus on bounding the integral. Note from Eq. (13.11) that

$$\begin{aligned}
\sqrt{p(\psi)} \langle \psi |^{\otimes k} | \Phi_\psi \rangle &= \sqrt{\binom{n+d-1}{n}} (\langle \psi |^{\otimes k} \otimes \langle \psi |^{\otimes n}) | \Phi \rangle \\
&= \sqrt{\binom{n+d-1}{n}} \langle \psi |^{\otimes k+n} | \Phi \rangle.
\end{aligned}$$

Hence,

$$\begin{aligned}
\int d\psi p(\psi) |\langle \psi |^{\otimes k} | \Phi_\psi \rangle|^2 &= \binom{n+d-1}{n} \int d\psi \langle \Phi | (|\psi\rangle \langle \psi|^{\otimes k+n}) | \Phi \rangle \\
&= \binom{n+d-1}{n} \binom{k+n+d-1}{k+n}^{-1} \int d\psi \langle \Phi | \Pi_{k+n} | \Phi \rangle \\
&= \binom{n+d-1}{n} \binom{k+n+d-1}{k+n}^{-1},
\end{aligned}$$

where we used the integral formula from Lemma 13.11 and then the assumption $|\Phi\rangle \in \text{Sym}^{k+n}(\mathbb{C}^d)$ which implies that $\Pi_{k+n}|\Phi\rangle = |\Phi\rangle$.

The ratio of the two binomial coefficients can be expressed as follows:

$$\begin{aligned}
\binom{n+d-1}{n} \binom{k+n+d-1}{k+n}^{-1} &= \frac{(n+d-1)!}{n!(d-1)!} \cdot \frac{(k+n)!(d-1)!}{(k+n+d-1)!} \\
&= \frac{(n+d-1)!}{n!} \cdot \frac{(k+n)!}{(k+n+d-1)!} \\
&= \frac{(n+d-1) \cdots (n+1)}{(k+n+d-1) \cdots (k+n+1)}.
\end{aligned}$$

Note that $\frac{a+1}{b+1} - \frac{a}{b} = \frac{b-a}{b(b+1)} \geq 0$ when $b \geq a$, so $\frac{a+1}{b+1} \geq \frac{a}{b}$ and hence

$$\begin{aligned}
\frac{n+d-1}{k+n+d-1} \cdots \frac{n+1}{k+n+1} &\geq \left(\frac{n+1}{k+n+1} \right)^{d-1} \\
&= \left(1 - \frac{k}{k+n+1} \right)^{d-1} \\
&\geq 1 - (d-1) \frac{k}{k+n+1} \\
&\geq 1 - \frac{dk}{k+n},
\end{aligned}$$

where we used $(1 - \alpha)^x \geq 1 - \alpha x$ for $\alpha \in (0, 1)$ and $x \geq 1$. Putting everything together,

$$\begin{aligned} \frac{1}{2} \left\| \Phi_{A_1 \dots A_k} - \tilde{\Phi}_{A_1 \dots A_k} \right\|_1 &\leq \sqrt{1 - \int d\psi p(\psi) |\langle \psi |^{\otimes k} | \Phi_\psi \rangle|^2} \\ &= \sqrt{1 - \binom{n+d-1}{n} \binom{k+n+d-1}{k+n}^{-1}} \\ &\leq \sqrt{\frac{dk}{k+n}}, \end{aligned}$$

which is the desired bound. \square

13.4 Exercises

13.1 Symmetric subspace:

- (a) Write out Π_2 and Π_3 .
- (b) Example 13.4 gives a basis for $\text{Sym}^2(\mathbb{C}^2)$. Write down bases of $\text{Sym}^2(\mathbb{C}^d)$ and $\text{Sym}^3(\mathbb{C}^2)$.
- (c) Verify that $R_\pi R_\tau = R_{\pi\tau}$ and $R_\pi^\dagger = R_{\pi^{-1}}$, for all $\pi, \tau \in S_n$.
- (d) Verify that $\Pi_n = \frac{1}{n!} \sum_{\pi \in S_n} R_\pi$ is the orthogonal projection onto the symmetric subspace.

13.2 Integral formula: In this exercise you can prove the integral formula:

$$\Pi_n = \binom{n+d-1}{n} \int |\psi\rangle^{\otimes n} \langle \psi|^{\otimes n} d\psi =: \tilde{\Pi}_n$$

- (a) Show that $\tilde{\Pi}_n = \Pi_n \tilde{\Pi}_n$.
- (b) Use the very important Lemma 13.10 to show that $\tilde{\Pi}_n = \sum_{\pi} c_\pi R_\pi$ for suitable $c_\pi \in \mathbb{C}$.
- (c) Use parts (a) and (b) to prove the integral formula. That is, show that $\tilde{\Pi}_n = \Pi_n$.

13.3 First moment of Haar measure: There is a unique probability measure dU on the unitary operators $U(\mathcal{H})$ that is invariant under $U \mapsto VUV$ for every pair of unitaries V, W . It is called the *Haar measure*. Its defining property can be stated as follows: For every continuous function f on $U(\mathcal{H})$ and for all unitaries $V, W \in U(\mathcal{H})$, it holds that $\int f(U) dU = \int f(VUV) dU$. Now let $M \in L(\mathcal{H})$.

- (a) Argue that $\int U M U^\dagger dU$ commutes with all unitaries.
- (b) Deduce that $\int U M U^\dagger dU = \text{Tr}[M] \frac{I_d}{d}$, where $d = \dim \mathcal{H}$.
- (c) Generalize this to $\int (U_A \otimes I_B) M_{AB} (U_A \otimes I_B)^\dagger dU_A = \text{Tr}_A[M_{AB}] \frac{I_A}{d_A} \otimes M_B$, where $M_{AB} \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$.

13.4 De Finetti theorem and quantum physics: Given a Hermitian operator h on $\mathbb{C}^d \otimes \mathbb{C}^d$, consider the operator $H = \frac{1}{n-1} \sum_{i \neq j} h_{i,j}$ on $(\mathbb{C}^d)^{\otimes n}$, where $h_{i,j}$ acts by h on subsystems i and j and by the identity on the remaining subsystems (e.g., $h_{1,2} = h \otimes I^{\otimes (n-2)}$).

- (a) Show that $\frac{E_0}{n} \leq \frac{1}{n} \langle \psi^{\otimes n} | H | \psi^{\otimes n} \rangle = \langle \psi^{\otimes 2} | h | \psi^{\otimes 2} \rangle$ for every pure state ψ on \mathbb{C}^d .

Let E_0 denote the smallest eigenvalue of H and $|E_0\rangle$ a corresponding eigenvector. If the eigenspace is one-dimensional and $n > d$ then $|E_0\rangle \in \text{Sym}^n(\mathbb{C}^d)$ (you do not need to show this).

- (b) Use the de Finetti theorem to show that $\frac{E_0}{n} \approx \min_{\|\psi\|=1} \langle \psi^{\otimes 2} | h | \psi^{\otimes 2} \rangle$ for large n .

Interpretation: The Hamiltonian H describes a mean-field system. Your result shows that in the thermodynamic limit the ground state energy density can be computed using states of form $\psi^{\otimes n}$.

13.5 Rényi-2 entropy: In this problem you will study a new entropy measure called the *Rényi-2 entropy*. It is defined by $H_2(\rho) := -\log \text{Tr}[\rho^2]$ for any quantum state $\rho \in \mathcal{D}(\mathbb{C}^d)$.

- (a) Find a formula for $H_2(\rho)$ in terms of the eigenvalues of ρ .
- (b) Show that $H_2(\rho) \leq H(\rho)$ by using Jensen's inequality.
- (c) Show that $\text{Tr}[\rho^2] = \text{Tr}[F\rho^{\otimes 2}]$, where $F : |i\rangle \otimes |j\rangle \mapsto |j\rangle \otimes |i\rangle$ for all $i, j \in \{1, \dots, d\}$, is the *swap operator*.

13.6 Average entanglement: In this exercise you will study the average entanglement of a random pure state in $\mathcal{H}_A \otimes \mathcal{H}_B$ drawn from the uniform distribution $d\psi_{AB}$ discussed in class. Recall that the entanglement entropy of a pure state $|\psi_{AB}\rangle$ is given by $H(\rho_A) = H(\rho_B)$, where ρ_A and ρ_B are the reduced states of $|\psi_{AB}\rangle$.

- (a) Let F_{AA}, F_{BB} denote the swap operators on $\mathcal{H}_A^{\otimes 2}, \mathcal{H}_B^{\otimes 2}$ and let $d_A = \dim \mathcal{H}_A, d_B = \dim \mathcal{H}_B$. Use the integral formula for the symmetric subspace to deduce that

$$\int |\psi_{AB}\rangle^{\otimes 2} \langle \psi_{AB}|^{\otimes 2} d\psi_{AB} = \frac{1}{d_A d_B (d_A d_B + 1)} (I_{AA} \otimes I_{BB} + F_{AA} \otimes F_{BB}).$$

- (b) Verify that $\int \text{Tr}[\rho_A^2] d\psi_{AB} = \frac{d_A + d_B}{d_A d_B + 1}$.
- (c) Show that the average Rényi-2 entropy $H_2(\rho_A)$ for a random pure state $|\psi_{AB}\rangle$ is at least $\log(\min(d_A, d_B)) - 1$. Conclude that the same holds for the entanglement entropy.

Hint: Use Exercise 13.5 and Jensen's inequality.

13.7 Second moment of Haar measure: In Exercise 13.3, we discussed the Haar measure on $U(\mathcal{H})$, which is the unique probability measure dU with the following property: For every continuous function f on $U(\mathcal{H})$ and for all unitaries $V, W \in U(\mathcal{H})$, it holds that $\int f(U) dU = \int f(VUW) dU$.

- (a) Argue that, for any operator $A \in L(\mathcal{H}^{\otimes n})$, the so-called *twirl* $\int U^{\otimes n} A U^{\dagger \otimes n} dU$ can always be written as a linear combination of permutation operators $R_\pi, \pi \in S_n$.
- (b) Deduce that $\int U^{\otimes 2} A U^{\dagger \otimes 2} dU = \alpha I + \beta F$ for every $A \in L(\mathcal{H}^{\otimes 2})$, where F is the swap operator on $\mathcal{H}^{\otimes 2}$, $\alpha = \frac{d}{d^2 - d} \text{Tr}[A] - \frac{1}{d^2 - d} \text{Tr}[FA]$, and $\beta = \frac{d}{d^2 - d} \text{Tr}[FA] - \frac{1}{d^2 - d} \text{Tr}[A]$.

13.8 Practice: Let $|\psi_{AB}\rangle \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ be a pure state and $\rho_A = \text{Tr}_B[|\psi_{AB}\rangle\langle\psi_{AB}|]$ its reduced state on system A . We would like to understand how the entropy $H(\rho_A)$ behaves when $|\psi_{AB}\rangle$ is chosen uniformly at random. More specifically, let us fix $d_A = 3$ and try to understand how $H(\rho_A)$ depends on the dimension d_B of system B that is discarded.

- (a) Compute the average value of $H(\rho_A)$ over $n = 50,000$ uniformly random samples of $|\psi_{AB}\rangle$ when $d_B = 3$ and when $d_B = 5$.
- (b) Produce a histogram for the values of $H(\rho_A)$ over $n = 50,000$ uniformly random samples of $|\psi_{AB}\rangle$ when $d_B = 3$ and when $d_B = 5$ (use bars of width 0.05 in the histogram).

To generate a uniformly random unit vector $|\psi\rangle$ in \mathbb{C}^d , write $|\psi\rangle = \sum_{j=1}^d \psi_j |j\rangle$ and set $\psi_j = a_j + ib_j$ where each a_j and b_j is a real random variable chosen independently from the normal distribution of mean value zero and variance one. Once all amplitudes are chosen, you simply normalize the state so that it is a unit vector.