Lectures Notes on Quantum Information Theory

Michael Walter and Maris Ozols, University of Amsterdam

Spring 2020

Summary

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

Notation

- $\mathcal{H}$ – Hilbert space, Section 1.1
- $L(\mathcal{H},\mathcal{H}')$ – linear operators from $\mathcal{H}$ to $\mathcal{H}'$, Eq. (1.4)
- $\mathrm{PSD}(\mathcal{H})$ – positive semidefinite operators on $\mathcal{H}$, Eq. (1.6)
- $D(\mathcal{H})$ – quantum states on $\mathcal{H}$, Definition 1.6
- $|\Phi^+_{AB}\rangle$ – maximally entangled state, Eq. (2.17)
- $U(\mathcal{H})$ – unitary operators on $\mathcal{H}$, Eq. (2.20)
- $U(\mathcal{H},\mathcal{K})$ – isometries from $\mathcal{H}$ to $\mathcal{K}$, Eq. (2.21)
- $\|x\|_1$ – ℓ^1-norm of vectors, Definition 6.2
- $\|M\|_1$ – trace norm of operator M, Eq. (3.1)
- $\|M\|_2$ – Frobenius norm of operator M, Eq. (3.2)
- $\langle M,N\rangle_{HS}$ – Hilbert-Schmidt inner product, Eq. (3.3)
- $\|M\|_\infty$ – operator norm of M, Eq. (3.4)
- $T(p,q)$ – trace distance between distributions p and q, Definition 6.2
- $T(\rho,\sigma)$ – trace distance between states ρ and σ, Definition 3.1
- $F(\rho,\sigma)$ – fidelity between states ρ and σ, Definition 3.2
- $\mathrm{CP}(\mathcal{H}_A,\mathcal{H}_B)$ – completely positive maps from $\mathcal{H}_A$ to $\mathcal{H}_B$, Definition 3.8
- $C(\mathcal{H}_A,\mathcal{H}_B)$ – quantum channels from $\mathcal{H}_A$ to $\mathcal{H}_B$, Definition 3.8
- J^Φ_{AB} – Choi operator of $\Phi_{A\to B}$, Eq. (4.1)
- Δ – completely dephasing channel, Eq. (4.3)
- $P(\Sigma)$ – probability distributions on Σ, Eq. (5.1)
- $H(p)$, $H(X)$, $H(X)_p$ – Shannon entropy, Definitions 5.1 and 5.3
- $T_{n,\varepsilon}(p)$ – typical set, Definition 5.8
- $H(\rho)$, $H(A)$, $H(A)_\rho$ – von Neumann entropy, Definitions 6.1 and 7.1
- $F(\mathcal{T},\rho)$ – channel fidelity of channel $\mathcal{T}$ and state ρ, Definition 6.5
- $S_{n,\varepsilon}(\rho)$ – typical subspace, Definition 6.9
- $I(A:B)_\rho$ – mutual information of state ρ_{AB}, Definition 7.2
- $\chi(\{p_x,\rho_x\})$ – Holevo χ-quantity of ensemble $\{p_x,\rho_x\}$, Definition 8.1
- $D(p\|q)$ – relative entropy of distribution p with respect to distribution q, Definition 8.4
- $D(\rho\|\sigma)$ – quantum relative entropy of state ρ with respect to state σ, Definition 8.5
- $|\Phi^{zx}\rangle$ – Bell states, Eq. (9.1)
- $\mathrm{Sep}(\mathcal{H}_A\otimes\mathcal{H}_B)$ – separable operators on $\mathcal{H}_A\otimes\mathcal{H}_B$, Definition 9.4
- $\mathrm{SepD}(\mathcal{H}_A\otimes\mathcal{H}_B)$ – separable states on $\mathcal{H}_A\otimes\mathcal{H}_B$, Definition 9.4
- $\mathrm{SepCP}(\mathcal{H}_A,\mathcal{H}_C:\mathcal{H}_B,\mathcal{H}_D)$ – separable completely positive maps from $A:B$ to $C:D$, Definition 10.1
- $\mathrm{SepC}(\mathcal{H}_A,\mathcal{H}_C:\mathcal{H}_B,\mathcal{H}_D)$ – separable quantum channels, Definition 10.1
- $|M_{AB}\rangle$ – vectorization of $M\in L(\mathcal{H}_A,\mathcal{H}_B)$, Definition 10.2
- $\mathrm{Ent}_r(\mathcal{H}_A:\mathcal{H}_B)$ – operators of entanglement rank at most r on $\mathcal{H}_A\otimes\mathcal{H}_B$, Definition 10.4
- $\mathrm{LOCC}(\mathcal{H}_A,\mathcal{H}_C:\mathcal{H}_B,\mathcal{H}_D)$ – LOCC channels from $A:B$ to $C:D$, Definition 10.8
- $u\succ v$ – majorization of vectors, Definition 11.3
- $A\succ B$ – majorization of Hermitian operators, Definition 11.7
- $E_D(\rho)$ – distillable entanglement of ρ, Definition 12.1
- $E_C(\rho)$ – entanglement cost of ρ, Definition 12.2
- $\mathrm{Sym}^n(\mathcal{H})$ – symmetric subspace of $\mathcal{H}^{\otimes n}$, Definition 13.1

Acknowledgements

We would like to thank Harold Nieuwboer, Alvaro Piedrafita, and Freek Witteveen for all their help, feedback, and corrections. Thanks also to Khallil Berrekkal, Robert Cañellas, Egor Cickovskij, Jasper van Egeraat, Titas Geryba, Dimitrios Loupas, Andrea Mazzocco, and Thijs Jenneskens for spotting typos.

Last updated: June 12, 2020.

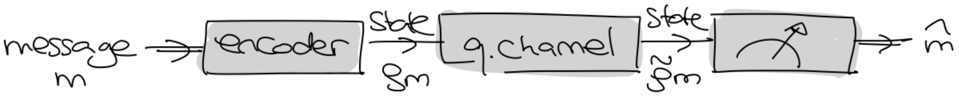
Contents

Lecture 1

Introduction to the formalism of quantum information theory

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.

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. The following compares Dirac notation with the corresponding expression in coordinates.

$$|\psi\rangle = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_d \end{pmatrix}, \qquad \langle\psi| = \begin{pmatrix} \overline{\psi_1} & \cdots & \overline{\psi_d} \end{pmatrix},$$

$$\langle\phi|\psi\rangle = \sum_{i=1}^d \overline{\phi_i}\psi_i, \qquad |\psi\rangle\langle\phi| = \begin{pmatrix} \psi_1\overline{\phi_1} & \dots & \psi_1\overline{\phi_d} \\ \vdots & & \vdots \\ \psi_d\overline{\phi_1} & \dots & \psi_d\overline{\phi_d} \end{pmatrix}$$

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

$$P = |\psi\rangle\langle\psi| \tag{1.1}$$

is the *orthogonal projection* ('projector') onto the one-dimensional space $\mathbb{C}|\psi\rangle$. For this, 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

$$\sum_i |e_i\rangle\langle e_i| = I \tag{1.2}$$

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:

$$\mathrm{Tr}[X] = \sum_i \langle e_i|X|e_i\rangle. \tag{1.3}$$

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

In quantum information, it is often useful to work with Hilbert spaces that have a privileged basis. Given a finite set Σ, we denote by $\mathbb{C}^\Sigma$ the Hilbert space with orthonormal basis $\{|x\rangle\}_{x\in\Sigma}$. Thus, $\langle x|y\rangle = \delta_{x,y}$. Note that the inner product is fully specified by this requirement. We will call the basis $\{|x\rangle\}$ the *standard basis* of $\mathbb{C}^\Sigma$.

Remark 1.2. *You may also can think of $\mathbb{C}^\Sigma$ as the vector space of functions $\Sigma \to \mathbb{C}$, 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 $f_x \colon \Sigma \to \mathbb{C}$, $f_x(y) = \delta_{x,y}$ which sends x to 1 and all other $y \in \Sigma \setminus \{x\}$ to 0.*

The simplest quantum system is the *qubit* – short for *quantum bit*. It corresponds to the two-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^\Sigma$, where $\Sigma = \{0, 1\}$. We will always identify $\mathcal{H} \cong \mathbb{C}^2$ using the standard basis, so that

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

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

1.2 Operators on Hilbert space

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

$$L(\mathcal{H}, \mathcal{K}) := \{A \colon \mathcal{H} \to \mathcal{K} \text{ linear}\}, \quad L(\mathcal{H}) := L(\mathcal{H}, \mathcal{H}) = \{A \colon \mathcal{H} \to \mathcal{H} \text{ linear}\}. \tag{1.4}$$

Recall that any operator $A \in L(\mathcal{H}, \mathcal{K})$ has an *adjoint*. This is the operator $A^\dagger \in L(\mathcal{K}, \mathcal{H})$ defined by the property that

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

If you write A as a matrix with respect to an orthonormal basis, then the adjoint is given by the conjugate transpose matrix: $A^\dagger = \overline{A^T} = (\overline{A})^T$.

Remark 1.3. *Note that this is the same rule that we used to go from a 'ket' to the corresponding 'bra'. Indeed, if we think of* $|\psi\rangle \in \mathcal{H}$ *as an operator* $\mathbb{C} \to \mathcal{H}$ *then it is not hard (but slightly confusing) to verify that* $\langle \psi | = |\psi\rangle^\dagger$ *– so this makes perfect sense!*

Next, we say that an operator $A \in L(\mathcal{H})$ is *Hermitian* if $A = A^\dagger$. Hermitian operators satisfy the important *spectral theorem*, which asserts that they are diagonalizable with real eigenvalues and orthonormal eigenvectors. Using Eq. (1.1), we can write the eigendecomposition in the following way:

$$A = \sum_i a_i |\psi_i\rangle\langle\psi_i|, \tag{1.5}$$

where the a_i are real numbers (the eigenvalues) and the $|\psi_i\rangle$ form an orthonormal basis (of eigenvectors). Conversely, any operator of this form is necessarily Hermitian with eigenvectors $|\psi_i\rangle$ and eigenvalues a_i. This can be seen by verifying that

$$A|\psi_j\rangle = \sum_i a_i |\psi_i\rangle\langle\psi_i|\psi_j\rangle = \sum_i a_i |\psi_i\rangle \delta_{i,j} = a_j |\psi_j\rangle.$$

The set of Hermitian operators forms a *real* vector space of dimension d^2, where $d = \dim \mathcal{H}$.

We now come to a central definition. We say that an operator A is *positive semidefinite (PSD)* if A is Hermitian and its eigenvalues are nonnegative. Thus A can be written as in Eq. (1.5) with $a_i \geqslant 0$. It is not easy to verify this definition directly, since in general it can be difficult to compute the eigenvalues of a given matrix. To this end, the following criterion is useful:

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

1. A *is PSD.*
2. $A = B^\dagger B$ *for some arbitrary* $B \in L(\mathcal{H}, \mathcal{K})$ *and Hilbert space* $\mathcal{K}$.
3. $\langle \psi | A | \psi \rangle \geqslant 0$ *for all* $|\psi\rangle \in \mathcal{H}$.

You can prove this in Practice Problem 1.2, which gives some further useful critera for positive semidefiniteness.

Positive semidefinite operators are so important that we will give them their own notation and define

$$\mathrm{PSD}(\mathcal{H}) = \{A \in L(\mathcal{H}) : A \text{ positive semidefinite}\}. \tag{1.6}$$

This is a convex set – see Practice Problem 1.3.

We will moreover write $A \geqslant B$ or $B \leqslant A$ iff $A - B \in \mathrm{PSD}(\mathcal{H})$. This defines a partial order on $L(\mathcal{H})$. For example, $A \geqslant 0$ means simply that A is positive semidefinite, while $A \leqslant I$ states that $I - A$ is positive semidefinite, i.e., A is Hermitian and has eigenvalues less or equal to one. See Practice Problem 1.4 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 $\mathbb{R}_{\geqslant 0}$. 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: $a \in \mathbb{R}_{\geqslant 0}$ iff $a = \bar{b}b = |b|^2$ for some $b \in \mathbb{C}$.*

1.3 States

We will now discuss the state space of quantum systems.

> **Definition 1.6** (State). *A* (quantum) state, density operator, *or* density 'matrix' *is by definition a semidefinite operator with trace one. We denote by $D(\mathcal{H}) = \{\rho \in \mathrm{PSD}(\mathcal{H}), \mathrm{Tr}[\rho] = 1\}$ the set of all quantum states.*

> **Axiom 1.7** (State space). *The state space of a quantum system with Hilbert space $\mathcal{H}$ is given by $D(\mathcal{H})$.*

By the spectral theorem for Hermitian operators [Eq. (1.5)], any quantum state can be written in the form

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \tag{1.7}$$

with eigenvalues p_i and orthonormal eigenvectors $|\psi_i\rangle$. Since ρ is positive semidefinite, all $p_i \geqslant 0$, and since $\mathrm{Tr}[\rho] = 1$, $\sum_i p_i = 1$. Thus, the eigenvalues (p_i) form a probability distribution! It is useful to distinguish some special classes of states:

- We say that ρ is a *pure* quantum state if it is of the form $\rho = |\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle \in \mathcal{H}$. Those are precisely the states of rank one – equivalently, the states that have one nonzero eigenvalue (which is then necessarily equal to 1). Thus, Eq. (1.7) shows that any state can be written as a mixture of pure states.

  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). You may remember this from your physics class. Mathematically, the space of pure states is a projective space.

- Quantum states that are not pure are often called *mixed*. In particular, on every Hilbert space $\mathcal{H}$ we always have a *maximally mixed state*

  $$\tau = \frac{I}{d}, \quad d = \dim \mathcal{H}.$$

  Its eigenvalues form a uniform probability distribution $(1/d, \ldots, 1/d)$.

- Given an arbitrary probability distribution $(p_x)_{x \in \Sigma}$ on some finite set Σ, we can define a corresponding quantum state on $\mathcal{H} = \mathbb{C}^\Sigma$:

$$\rho = \sum_{x \in \Sigma} p_x |x\rangle\langle x|. \tag{1.8}$$

  Such quantum states are called *classical*. With respect to the standard basis, classical states corresponding precisely to *diagonal* matrices.

The set of quantum states $D(\mathcal{H})$ is convex. This follows easily from the convexity of $\mathrm{PSD}(\mathcal{H})$, and you can prove this in Practice Problem 1.3. Convexity means that for any two states σ, ω and $p \in [0,1]$, the operator $\rho = p\sigma + (1-p)\omega$ is again a state. More generally, for any ensemble $\{q_j, \rho_j\}$ of quantum states – i.e., (q_j) is a probability distribution and the ρ_j are states – the mixture

$$\rho = \sum_j q_j \rho_j \tag{1.9}$$

is again a quantum state. We caution that, in general, Eq. (1.9) has nothing to do with the eigendecomposition (the ρ_j need neither be pure nor pairwise orthogonal).

The fact that the state space is convex is a very useful property. For example, suppose that we have a machine that emits the state ρ_j with probability q_j. Then it is natural to describe its average state by Eq. (1.9).

The following picture shows three convex sets:

The first has a round boundary, while the latter two have some 'flat' boundary pieces. What does the convex set of quantum states look like? To get more intuition we consider the case of a qubit.

1.4 Qubit and Bloch sphere

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

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.10}$$

are linearly independent and form a basis of the real vector space of Hermitian 2×2-matrices. 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 $\mathrm{Tr}[I] = 2$. As a consequence, we see that

$$\rho = \frac{1}{2}(I + xX + yY + zZ) = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}, \qquad \text{where } \vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3, \tag{1.11}$$

is the most general form of a Hermitian 2×2-matrix with $\mathrm{Tr}[\rho] = 1$. So far, the *Bloch vector* $\vec{r}$ is completely arbitrary. When is ρ a quantum state? We will need to ensure that ρ is PSD, i.e., has nonnegative eigenvalues. Since $\operatorname{Tr}[\rho] = 1$, its eigenvalues are given by $p, 1-p$ for some $p \in \mathbb{R}$. A moments thought shows that $p \geqslant 0$ and $1-p \geqslant 0$ if and only if $p(1-p) \geqslant 0$ (since p and $1-p$ cannot both be negative). But precisely this product is computed by the determinant:

$$p(1-p) = \det(\rho) = \frac{1}{4}\left((1+z)(1-z) - (x+iy)(x-iy)\right) = \frac{1}{4}\left(1 - x^2 - y^2 - z^2\right) = \frac{1}{4}\left(1 - \|\vec{r}\|^2\right).$$

Thus, ρ is a quantum state if and only if $\|\vec{r}\| \leqslant 1$. Thus we have shown that the state space of a qubit can be identified with the unit ball in $\mathbb{R}^3$ – which is known as the *Bloch ball* and clearly convex.

When is ρ a *pure* state? This is the case precisely when $p(1-p) = 0$ (one eigenvalue is zero and the other is one), i.e., when $\|\vec{r}\| = 1$, i.e., when $\vec{r}$ is in the *Bloch sphere* of radius one.

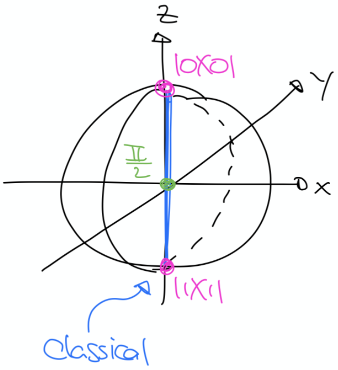
We summarize our findings in the following lemma.

Lemma 1.8 (Bloch ball)**.** *Any qubit state $\rho \in D(\mathbb{C}^2)$ can be written in the form*

$$\rho = \frac{1}{2}\left(I + xX + yY + zZ\right),$$

where $\vec{r} = \left(\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right)$ is an arbitrary vector of norm $\|\vec{r}\| \leqslant 1$. Moreover, ρ is pure if and only if $\|\vec{r}\| = 1$.

The following picture gives a rough sketch of the situation:



The north and south poles have Bloch vectors

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

which correspond [via Eq. (1.11)] to the pure states

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

More generally, the Bloch vectors on the blue line segment between north and south pole correspond to the classical states

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}. \tag{1.12}$$

In particular, the origin of the Bloch ball corresponds to the maximally mixed qubit state $\tau = I/2$, with Bloch vector $\vec{r} = 0$. Can you figure out the pure states that corresponding to the 'east' and 'west poles', as well as to the 'front' and 'back poles'? You can learn more about the Bloch sphere in Practice Problem 1.5. In particular, you may prove there that the components of the Bloch vector $\vec{r}$ can be calculated by

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

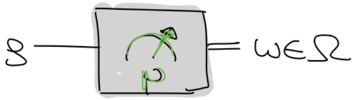
1.5 Measurements

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 1.9 (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*

$$\mu \colon \Omega \to \mathrm{PSD}(\mathcal{H}) \quad \textit{such that} \quad \sum_{\omega \in \Omega} \mu(\omega) = I. \tag{1.13}$$

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



Importantly, the measurement outcome ω will in general be *random* (even if we know μ and ρ precisely). In this sense, quantum mechanics is a *probabilistic* theory. How can we calculate the probability of measurement outcomes? For this we use the following axiom, which is often referred to as *Born's rule*.

Axiom 1.10 (Born's rule). *If we measure a quantum system in state $\rho \in D(\mathcal{H})$ using a measurement μ, then the probability of outcome $\omega \in \Omega$ is given by* Born's rule*:*

$$\Pr(\textit{outcome } \omega | \textit{state } \rho) = \mathrm{Tr}[\mu(\omega)\rho] \tag{1.14}$$

Let us verify that Born's rule in Eq. (1.14) defines a probability distribution. Indeed, $\mathrm{Tr}[\mu(\omega)\rho] \geqslant 0$ (since the trace of a product of two PSD operators is always nonnegative) and

$$\sum_{\omega \in \Omega} \mathrm{Tr}[\mu(\omega)\rho] = \mathrm{Tr}[\sum_{\omega \in \Omega} \mu(\omega)\rho] = \mathrm{Tr}[\rho] = 1,$$

where we first use that the elements $\mu(\omega)$ sum to the identify operator [Eq. (1.13)] and then that quantum state have trace one. Thus, Born's rule makes sense.

Remark 1.11 (Measurements vs. observables). *If you have attended a course in quantum mechanics, you may know the notion of an* observable*, which is another way to think about measurements. On Practice Problem 2.6 you can explore how these two notions are related. Namely, observables correspond precisely to* projective *measurements (as defined below) that take outcomes in the reals (i.e., $\Omega \subseteq \mathbb{R}$).*

Remark 1.12 (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 post-measurement state 'collapses' 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 – as in the figure above.*

Just like we did for states, it is useful to single out some special classes of measurement:

- We say that a measurement is *projective* if all $\mu(\omega)$ are orthogonal projections. Recall that this means $\mu(\omega)^2 = \mu(\omega) = \mu(\omega)^\dagger$ for each $\omega \in \Omega$. It is not directly obvious, but true that this implies that $\mu(\omega)\mu(\omega') = 0$ for all $\omega \neq \omega' \in \Omega$.

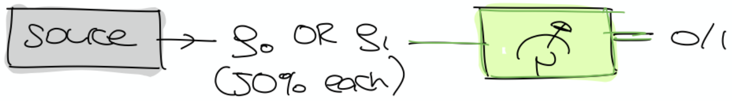
- Given an orthonormal basis $\{|\psi_\omega\rangle\}_{\omega\in\Omega}$, we can always define a projective measurement as follows. Define the *basis measurement* in basis $\{|\psi_\omega\rangle\}_{\omega\in\Omega}$ as the following measurement,
$$\mu \colon \Omega \to \mathrm{PSD}(\mathcal{H}), \quad \mu(\omega) = |\psi_\omega\rangle\langle\psi_\omega|,$$
with outcomes in Ω. Note that Born's rule can be rewritten as follows:
$$\Pr(\text{outcome } \omega|\text{state } \rho) = \mathrm{Tr}[\mu(\omega)\rho] = \mathrm{Tr}[|\psi_\omega\rangle\langle\psi_\omega|\rho] = \langle\psi_\omega|\rho|\psi_\omega\rangle.$$

- In particular, we can always consider the *standard basis measurement* on $\mathcal{H} = \mathbb{C}^\Sigma$:
$$\mu \colon \Sigma \to \mathrm{PSD}(\mathcal{H}), \quad \mu(x) = |x\rangle\langle x|,$$
with outcomes in Σ. The probabilites of measurement outcomes are given by
$$\Pr(\text{outcome } x|\text{state } \rho) = \langle x|\rho|x\rangle.$$
In particular, if $\rho = |\Psi\rangle\langle\Psi|$ is a pure state then
$$\Pr(\text{outcome } x|\text{state } \Psi) = \langle x|\Psi\rangle\langle\Psi|x\rangle = |\langle x|\Psi\rangle|^2 = |\Psi_x|^2,$$
where we note that $\langle x|\Psi\rangle$ is the same as the component Ψ_x when expanding $|\Psi\rangle = \sum_x \Psi_x|x\rangle$ with respect to the standard basis. If you attend Ronald de Wolf's quantum computing course then this formula will look very familiar to you!

For a qubit, $\mathbb{C}^2$, the standard basis is $\{|0\rangle, |1\rangle\}$, so the *standard basis measurement* reads
$$\mu_{\mathrm{Std}} \colon \{0,1\} \to \mathrm{PSD}(\mathbb{C}^2), \quad x \mapsto |x\rangle\langle x|.$$
Another basis of $\mathbb{C}^2$ is the so-called *Hadamard basis* $\{|+\rangle, |-\rangle\}$, where
$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ \pm 1\end{pmatrix}.$$
The Hadamard basis measurement is defined by
$$\mu_{\mathrm{Had}} \colon \{0,1\} \to \mathrm{PSD}(\mathbb{C}^2), \quad \mu_{\mathrm{Had}}(0) = |+\rangle\langle +|, \quad \mu_{\mathrm{Had}}(1) = |-\rangle\langle -|.$$
Suppose for example that our qubit is in state $\rho = |0\rangle\langle 0|$ and we carry out the standard basis measurement. Then the probability of outcome '0' is given by
$$p_{\mathrm{Std}}(0) = \langle 0|\rho|0\rangle = |\langle 0|0\rangle|^2 = 1,$$
i.e., the measurement yields outcome '0' with certainty (as one might expect). In contrast, if we perform a Hadamard basis measurement then the probability of outcome '0' (corresponding to state $|+\rangle\langle +|$) is given by
$$p_{\mathrm{Had}}(0) = \langle +|\rho|+\rangle = |\langle +|0\rangle|^2 = \frac{1}{2},$$
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 Homework Problem 1.2, you will show an *uncertainty relation* which states that there exists no quantum state for which both the standard and the Hadamard measurement are certain.

1.6 Distinguishing quantum states

Now we have learned all the quantum information formalism required to prove a first interesting result. Suppose that we have a source that emits either of two states $\rho_0, \rho_1 \in D(\mathcal{H})$ with 50% probability each, as in the following picture:



Our goal is to design a measurement $\mu\colon \{0,1\} \to \mathrm{PSD}(\mathcal{H})$ that identifies the correct state as best as possible. That is, we want to maximize the success probability

$$p_{\text{success}} = \frac{1}{2}\Pr(\text{outcome } 0|\text{state } \rho_0) + \frac{1}{2}\Pr(\text{outcome } 1|\text{state } \rho_1) = \frac{1}{2}\operatorname{Tr}[\mu(0)\rho_0] + \frac{1}{2}\operatorname{Tr}[\mu(1)\rho_1]$$

over all possible measurements. How can we calculate the optimal success probability and find the corresponding measurement? We will come back to this next week (see end of Lecture 2 and Homework Problem 2.2)!

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. (1.14)]. 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. Why did we care about mixed states? One reason is that they allow us to model probability distributions. Indeed, if $(p_x)_{x\in\Sigma}$ is a probability distribution then we can associate with it the classical state $\rho = \sum_x p_x|x\rangle\langle x|$ on $\mathcal{H} = \mathbb{C}^\Sigma$ [Eq. (1.8)]. More generally, we can use mixed states to model the average state of an ensemble. For example, if a quantum device outputs a system in state ρ_j with probability p_j (the states ρ_j need not be pure or orthogonal), then we might describe the average output of the device by the state $\rho = \sum_j p_j \rho_j$, which is generically mixed [Eq. (1.9)]. (If you have taken a course on quantum statistical physics then you will also have seen that equilibrium states such as Gibbs states are naturally mixed states.)

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. (2.15)). This phenomenon is closely related to the notion of *entanglement*, which will be discussed in more detail next month.

Remark 2.1 (Why not restrict to pure states?). *There is a more general philosophical point that is worth mentioning. In quantum computing, we usually start out with a pure initial state, apply unitary operations, and only at the very end carry out a basis measurement. This allows us to work with vectors $|\psi\rangle$ rather than with density operators ρ. In contrast, in information theory we often deal with uncertainty and noise. For this, it is more natural to work with mixed states. 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, which we will introduce next week. However, it is important to point out that both formalisms are equivalent. For example, one of this lecture's key points will be that we can always think of mixed states in terms of subsystems of pure states (see Section 2.3 below), and we will see that quantum channels can similarly be reduced to unitary operations on a larger system.*

2.1 Joint or composite systems

Axiom 2.2 (Composing systems). *For a quantum system composed of n subsystems with Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$, the overall Hilbert space is given by the tensor product $\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$.*

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

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

This space has a natural product basis $|x_1, \ldots, x_n\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle$, indexed by bitstrings $x = (x_1, \ldots, x_n) \in \{0,1\}^n$. We will often leave out the commas and simply write, e.g., $|010\rangle = |0\rangle \otimes |1\rangle \otimes |0\rangle$.

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

$$\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$$

has a natural product basis $|x_1, \ldots, x_n\rangle = |x_1\rangle \otimes \cdots \otimes |x_n\rangle$, indexed by n-tuples $x = (x_1, \ldots, x_n) \in \Sigma = \Sigma_1 \times \cdots \times \Sigma_n$. (Thus, $\mathcal{H} \cong \mathbb{C}^{\Sigma}$.)

What are possible states on a tensor-product Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$?

- Given states $\rho_1, \ldots, \rho_n$, where $\rho_i \in D(\mathcal{H}_i)$, we can always form a so-called *product state*

  $$\rho = \rho_1 \otimes \cdots \otimes \rho_n \tag{2.1}$$

  Here we use the tensor product of *operators*, rather than of vectors (see Remark 2.3 below for a reminder).

  You can think of the product states as the quantum generalization of joint probability distribtions where the random variables are *independent* (which means that the probability distribution factors as $p(x_1, \ldots, x_n) = p(x_1) \ldots p(x_n)$).

- Not all states are product states [i.e., of the form Eq. (2.1)]. States that are not product states are called *correlated*. Here is an example of a correlated two-qubit state:

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

  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). *Let us recall the definition of the tensor product of operators [already used in Eq. (2.1) above]. If $X \in L(\mathcal{H}_1, \mathcal{K}_1)$ and $Y \in L(\mathcal{H}_2, \mathcal{K}_2)$ are linear operators, then their tensor product $X \otimes Y$ is a linear operator in $L(\mathcal{H}_1 \otimes \mathcal{H}_2, \mathcal{K}_1 \otimes \mathcal{K}_2)$ defined as follows:*

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

Note that this definition is not *circular – we define the tensor product of operators in terms of the tensor product of vectors. Note that Eq. (2.3) in particular implies that 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 the matrices 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, the operators $Y \in L(\mathbb{C}, \mathcal{K}_2)$ can be identified with a vector $|\chi\rangle \in \mathcal{K}_2$ (in coordinates, this means that a matrix with a single column is the same as a column vector), and the operator $X \otimes |\chi\rangle$ in $L(\mathcal{H}_1, \mathcal{K}_1 \otimes \mathcal{K}_2)$ simply acts as

$$\big(X \otimes |\chi\rangle\big)|\psi\rangle = X|\psi\rangle \otimes |\chi\rangle. \tag{2.4}$$

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

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

$$\big(X \otimes \langle\chi|\big)\big(|\psi\rangle \otimes |\phi\rangle\big) = X|\psi\rangle\,\langle\chi|\phi\rangle = \langle\chi|\phi\rangle\,X|\psi\rangle \tag{2.5}$$

Note that $\langle\chi|\phi\rangle \in \mathbb{C}$. In the middle formula, we right-multiply the vector $X|\psi\rangle$ by this number, while on the right we left-multiply it. (If also $\mathcal{K}_1 = \mathbb{C}$ then 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 between an operator and a vector or covector.

Remark 2.4. *It is not hard to see that generic quantum states are correlated. Here is a simple dimension counting argument. 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 $\operatorname{Tr} \rho = 1$ reduces the dimension by one). Now suppose for simplicity that each $\mathcal{H}_i$ is d-dimensional, so that $\mathcal{H} = \mathcal{H}_1 \otimes \ldots \otimes \mathcal{H}_n$ has dimension d^n. Then the space of density operators has dimension $d^{2n} - 1$, which grows exponentially with n, while the space of product states has dimension $n(d^2 - 1)$, so 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 ubiquitious in the quantum information literature.

Definition 2.5 (Subscripts for subsystems). *Fom 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 $D(\mathcal{H}_{AB})$, X_B for linear operators on $\mathcal{H}_B$, and so forth.*

2.2 Partial trace and reduced states

Suppose 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, say?

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

To answer this question we need more input from quantum theory:

Axiom 2.6 (Born's rule for measuring a subsystem). *If the quantum system AB is in state ρ_{AB} and we want to measure $\mu_A : \Omega \to \mathrm{PSD}(\mathcal{H}_A)$ on subsystem A, then the probability of measurement outcomes is calculated as follows:*

$$\Pr(\textit{outcome } \omega) = \operatorname{Tr}\big[\rho_{AB}(\mu_A(\omega) \otimes I_B)\big] \tag{2.6}$$

Note that this is precisely Born's rule [Eq. (1.14)] for the following measurement on AB*:*

$$\mu_A \otimes I_B \colon \Omega \to \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B), \quad \omega \mapsto \mu_A(\omega) \otimes I_B$$

We can now phrase our initial question more precisely: Given a state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$, can we find a state $\rho_A \in D(\mathcal{H}_A)$ such that

$$\mathrm{Tr}\big[\rho_{AB}(\mu_A(\omega) \otimes I_B)\big] = \mathrm{Tr}\big[\rho_A \mu_A(\omega)\big] \tag{2.7}$$

for all possible Ω, measurements $\mu_A \colon \Omega \to \mathrm{PSD}(\mathcal{H}_A)$, and outcomes $\omega \in \Omega$?

Note that Eq. (2.7) means that the state ρ_A reproduces the statistics of all possible measurements on A – but contains no information about B (since it is a state on A alone). This is exactly the kind of object that we are looking for. How can we find such a ρ_A? We first give the solution and then verify that it does the job.

Definition 2.7 (Partial trace)**.** *The* partial trace *over* B *is the linear map* $\mathrm{Tr}_B \colon L(\mathcal{H}_A \otimes \mathcal{H}_B) \to L(\mathcal{H}_A)$ *defined as follows: For every* $M_{AB} \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$*,*

$$\mathrm{Tr}_B[M_{AB}] := \sum_b \big(I_A \otimes \langle b|\big) M_{AB} \big(I_A \otimes |b\rangle\big), \tag{2.8}$$

where $|b\rangle$ *is an arbitrary orthonormal basis of* $\mathcal{H}_B$*. Note that* $\mathrm{Tr}_B[M_{AB}] \in L(\mathcal{H}_A)$*. Concretely, its matrix entries with respect to an arbitrary orthonormal basis* $|a\rangle$ *of* $\mathcal{H}_A$ *are given by:*

$$\langle a|\,\mathrm{Tr}_B[M_{AB}]|a'\rangle = \sum_b \langle a, b|M_{AB}|a', b\rangle \tag{2.9}$$

See Eqs. (2.4) and (2.5) to remind yourself of the meaning of $I_A \otimes \langle b|$ and $I_A \otimes |b\rangle$. Note that, not only does the partial trace send operators to operators – but it is itself a linear operator! Sometimes such maps are called *superoperators*. We now list some useful properties of the partial trace.

1. For any operator of tensor product form, $M_{AB} = X_A \otimes Y_B$, where $X_A \in L(\mathcal{H}_A)$ and $Y_B \in L(\mathcal{H}_B)$, we have

$$\mathrm{Tr}_B[X_A \otimes Y_B] = X_A\,\mathrm{Tr}[Y_B] = \mathrm{Tr}[Y_B]\,X_A. \tag{2.10}$$

   This justifies the name *partial trace* (we take the trace of Y_B but leave X_A untouched). To prove Eq. (2.10), use Eq. (2.8) to see that

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

   (If the second step confuses you, apply $|a\rangle$ from the left and $\langle a'|$ from the right.) □

2. For any two operators M_{AB} and X_A, where $M_{AB} \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$, $X_A \in L(\mathcal{H}_A)$, we have

$$\mathrm{Tr}\big[M_{AB}(X_A \otimes I_B)\big] = \mathrm{Tr}\big[\mathrm{Tr}_B[M_{AB}]\,X_A\big] \tag{2.11}$$

Let us give a careful proof of this crucial identity:

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

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. (2.8). □

We now recognize that the partial trace indeed solves our problem. Simply define $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$. Then Eq. (2.7) as a direct consequence of Eq. (2.11) [choose $M_{AB} = \rho_{AB}$ and $X_A = \mu_A(\omega)$]. This calls for its own definition and notation:

> **Definition 2.8** (Reduced states). *Given a state ρ_{AB} on AB, we define its* reduced state *on subsystem A by $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$. Similarly, we define the reduced state on subsystem B by $\rho_B := \mathrm{Tr}_A[\rho_{AB}]$.*
>
> *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 $\rho_{AB} := \mathrm{Tr}_C[\rho_{ABC}]$, $\rho_{AC} := \mathrm{Tr}_B[\rho_{ABC}]$, $\rho_A := \mathrm{Tr}_{BC}[\rho_{ABC}]$, etc.*

It is a pleasant exercise to verify that reduced states are again states, i.e., that ρ_A is PSD and has trace one (etc). This follows by combining the following two facts:

3. $\mathrm{Tr}[M_{AB}] = \mathrm{Tr}[\mathrm{Tr}_B[M_{AB}]]$ for any operator M_{AB}.

   This follows directly from Eq. (2.11) by choosing $X_A = I_A$ (the identity operator). □

4. If M_{AB} is positive semidefinite, then so is $\mathrm{Tr}_B[M_{AB}]$.

   To see this, note that if $X_A \in \mathrm{PSD}(\mathcal{H}_A)$ then $X_A \otimes I_B \in \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$, so

   $$\mathrm{Tr}\big[\mathrm{Tr}_B[M_{AB}] \, X_A\big] = \mathrm{Tr}\big[M_{AB}(X_A \otimes I_B)\big] \geqslant 0.$$

   The equality is Eq. (2.11), and the inequality holds since M_{AB} is PSD [see Practice Problem 1.2 (e)]. This in turn implies that $\mathrm{Tr}_B[M_{AB}]$ is PSD (by the same criterion). □

Here is another useful observation.

5. ρ_{AB} is a product state iff $\rho_{AB} = \rho_A \otimes \rho_B$ (i.e., ρ_{AB} is a product of its reduced states).

   Clearly, the right-hand side is stronger than the left-hand side one. Conversely, suppose that ρ_{AB} is a product state, i.e., $\rho_{AB} = \sigma_A \otimes \omega_B$ for some arbitrary states σ_A and ω_B. Then, necessarily $\rho_A = \sigma_A$ and $\rho_B = \sigma_B$, as follows from Eq. (2.10). □

In Practice Problem 2.5 and Homework Problem 2.3 (a) you can establish further useful properties of the partial trace.

Let us discuss a concrete example. Let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$ and consider the unit vector

$$|\Phi^+_{AB}\rangle := \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right). \tag{2.12}$$

(the $^+$ means nothing in particular, it is just a symbol to indicate this particular vector). The corresponding pure state is known as a *maximally entangled state* of two qubits,

$$\rho_{AB} := |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}| = \frac{1}{2}\Big(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|\Big) \tag{2.13}$$

$$= \frac{1}{2}\Big(|0\rangle\langle 0| \otimes \underbrace{|0\rangle\langle 0|}_{\mathrm{Tr}=1} + |1\rangle\langle 0| \otimes \underbrace{|1\rangle\langle 0|}_{\mathrm{Tr}=0} + |0\rangle\langle 1| \otimes \underbrace{|0\rangle\langle 1|}_{\mathrm{Tr}=0} + |1\rangle\langle 1| \otimes \underbrace{|1\rangle\langle 1|}_{\mathrm{Tr}=1}\Big) \tag{2.14}$$

(The second line follows from the first by the same reasoning as in Eq. (2.2).) To compute its reduced state on A, we can simply use linearity and Eq. (2.10) for each of the four terms. Using the traces indicated in Eq. (2.14), the result is

$$\rho_A = \frac{1}{2}\Big(|0\rangle\langle 0| + |1\rangle\langle 1|\Big). \tag{2.15}$$

By symmetry, the reduced state ρ_B is given by the same formula.

Remark 2.9. *It is also instructive to write down the above objects with respect to the product basis $|00\rangle$, $|01\rangle$, $|10\rangle$, $|11\rangle$:*

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \qquad \rho_{AB} = \frac{1}{2}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 0 & 0 & 1\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\1 & 0 & 0 & 1\end{pmatrix} \tag{2.16}$$

The right-hand side matrix can also be read off directly from Eq. (2.13). Compared with the correlated classical state Eq. (2.2), the maximally entangled state also has 1s in the top right and bottom left corners. This is a crucial difference! For example, ρ_{AB} is a pure state (has rank one), while Eq. (2.2) is mixed (its rank is two).

More generally, for any finite set Σ, we can let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^\Sigma$ and define the maximally entangled state

$$|\Phi^+_{AB}\rangle := \frac{1}{\sqrt{|\Sigma|}}\sum_{x\in\Sigma}|xx\rangle. \tag{2.17}$$

Then we can check, analogously to the qubit example, that its reduced state ρ_A is given by

$$\begin{aligned}\rho_A &:= \mathrm{Tr}_B\Big[|\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|\Big]\\ &= \frac{1}{|\Sigma|}\mathrm{Tr}_B\Big[\sum_{x,y\in\Sigma}|xx\rangle\langle yy|\Big]\\ &= \frac{1}{|\Sigma|}\sum_{x\in\Sigma}|x\rangle\langle x|\end{aligned}$$

using that $\mathrm{Tr}_B\Big[|xx\rangle\langle yy|\Big] = 0$ if $x \neq y$ and $\mathrm{Tr}_B\Big[|xx\rangle\langle xx|\Big] = |x\rangle\langle x|$. Hence, ρ_A is the maximally mixed state on A, and by symmetry ρ_B is also the maximally mixed state on B.

The above example has a remarkable feature. We started with a global pure state ρ_{AB}, but nevertheless its reduced states ρ_A and ρ_B were mixed. This is an important reason for allowing general density operators – they naturally arise when describing the states of subsystems. We alluded to this in the discussion at the beginning of the lecture.

2.3 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, as is the content of the following lemma:

Lemma 2.10 (Existence of purifications)**.** *Let* $\sigma_A \in D(\mathcal{H}_A)$ *be a state and* $\mathcal{H}_B$ *a Hilbert space of dimension* $\dim \mathcal{H}_B \geqslant \operatorname{rank} \sigma_A$. *Then there exists a pure state* $|\Psi_{AB}\rangle$ *such that*

$$\mathrm{Tr}_B \left[|\Psi_{AB}\rangle\langle\Psi_{AB}|\right] = \sigma_A.$$

A pure state with this property is called a purification *of* σ_A.

Proof. Consider a spectral decomposition $\sigma_A = \sum_{i=1}^r p_i |e_i\rangle\langle e_i|$, where $r = \operatorname{rank}(\sigma_A)$, the p_i are the nonzero eigenvalues of σ_A (so $p_i > 0$), and $|e_i\rangle$ corresponding orthonormal eigenvectors. Since $\dim \mathcal{H}_B \geqslant r$, we can choose orthonormal vectors $|f_1\rangle, \dots, |f_r\rangle \in \mathcal{H}_B$. Then,

$$|\Psi_{AB}\rangle := \sum_{i=1}^r \sqrt{p_i} |e_i\rangle \otimes |f_i\rangle \tag{2.18}$$

is a purification of σ_A. Indeed,

$$\mathrm{Tr}_B \left[|\Psi_{AB}\rangle\langle\Psi_{AB}|\right] = \sum_{i,j} \sqrt{p_i p_j} |e_i\rangle\langle e_j| \underbrace{\mathrm{Tr}\left[|f_i\rangle\langle f_j|\right]}_{=\delta_{i,j}} = \sum_i p_i |e_i\rangle\langle e_i| = \sigma_A$$

by virtually the same calculation that we used to deduce Eq. (2.15) from Eq. (2.13). □

In case you were suspicious of why we consider arbitrary density operator in Axiom 1.7 and not just pure states, Lemma 2.10 should alleviate your concerns!

Are purifications unique? In the proof of Lemma 2.10 we chose an arbitrary orthonormal basis of $\mathcal{H}_B$, so clearly they are not unique. However, this turns out to be the only source of ambiguity:

Lemma 2.11 (Uniqueness of purifications)**.** *Let* $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *and* $|\Phi_{AC}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_C$ *be two purifications of* $\sigma_A \in D(\mathcal{H}_A)$. *If* $\dim \mathcal{H}_B \leqslant \dim \mathcal{H}_C$ *then there exists an isometry* $V_{B\to C} \colon \mathcal{H}_B \to \mathcal{H}_C$ *such that*

$$|\Phi_{AC}\rangle = (I_A \otimes V_{B\to C})|\Psi_{AB}\rangle. \tag{2.19}$$

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

Recall that an operator $U \in L(\mathcal{H}, \mathcal{K})$, where $\dim \mathcal{H} = \dim \mathcal{K}$, is called a *unitary* if $U^\dagger U = I_{\mathcal{H}}$ and $UU^\dagger = I_{\mathcal{K}}$, the two conditions being equivalent. We denote the set of all unitary operators on $\mathcal{H}$ by

$$U(\mathcal{H}) := \{U \in L(\mathcal{H}) : U^\dagger U = I_{\mathcal{H}}\}. \tag{2.20}$$

More generally, an operator $V \in L(\mathcal{H}, \mathcal{K})$ is called an *isometry* if $V^\dagger V = I_{\mathcal{H}}$. This implies that $\dim \mathcal{H} \leqslant \dim \mathcal{K}$, so the target space can generally have a larger dimension than the domain. We denote the set of all isometries from $\mathcal{H}$ to $\mathcal{K}$ by

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

If $\dim \mathcal{H} = \dim \mathcal{K}$ then any isometry is a unitary. Isometries (in particular unitaries) preserve inner products, so they map orthonormal sets to orthonormal sets.

See Remark 2.15 below for a proof sketch of Lemma 2.11. In Practice Problem 3.5 you get to fill in the details.

There is a particularly convenient way to construct a purification: Given σ_A, choose $\mathcal{H}_B = \mathcal{H}_A$ and define the *standard purification*

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

where $|x\rangle$ is some arbitrary orthonormal basis of $\mathcal{H}_A = \mathcal{H}_B$ (since this involves a choice, the term 'standard purification' is a bit of a misnomer). The square root $\sqrt{\sigma_A}$ is the PSD operator defined by taking the square roots of the eigenvalues of σ_A, while keeping the eigenvectors the same. Clearly, $\sqrt{\sigma_A}\sqrt{\sigma_A} = \sigma_A$, which justifies the notation. To see that Eq. (2.22) is a purification, simply compute the partial trace:

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

To go from the first to the second line, use Practice Problem 2.5 (a). One can verify that $|\Psi^{\text{std}}_{AB}\rangle$ is indeed a valid quantum state (i.e., a unit vector) for any state σ_A – we leave it as an exercise for you to show that $\langle\Psi^{\text{std}}_{AB}|\Psi^{\text{std}}_{AB}\rangle = \operatorname{Tr}[\sigma] = 1$.

2.4 Schmidt decomposition

States of the form Eq. (2.18) are quite pleasant to work with, since 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.

Lemma 2.12 (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

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

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

Corollary 2.13 (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 $\{s_i^2\}$).*

Proof. This is immediate from Eq. (2.23). □

Corollary 2.14 (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 iff ρ_B is pure iff ρ_{AB} is a product state.*

Proof. Clearly, ρ_A is pure iff ρ_B is pure since both have the same rank (Corollary 2.13). If ρ_A is pure then there is only one nonzero Schmidt coefficients ($s_1 = 1$), 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 $\rho_{AB} = \rho_A \otimes \rho_B$ (see Property 5 on p. 23). Since ρ_{AB} is pure, we have $1 = \operatorname{rank} \rho_{AB} = \operatorname{rank} \rho_A \operatorname{rank} \rho_B$. Thus, both ρ_A and ρ_B have rank one, hence are pure states. □

It is crucially important in Corollary 2.14 that the global state ρ_{AB} is pure. For mixed ρ_{AB}, it is still holds that if ρ_A is pure then ρ_{AB} is a product state – you will prove this in Homework Problem 2.3 – 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.

Remark 2.15 (On the uniqueness of purifications). *We can also use the Schmidt decomposition to see why Lemma 2.11 should hold. Consider a Schmidt decomposition of the first purification, say,*

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

By Corollary 2.13, both r and the $s_i > 0$ are uniquely determined by σ_A. For simplicity, assume that the eigenvalues of σ_A are distinct. In this case, the $|e_i\rangle$ are likewise uniquely determined up to phases (namely, by the property of being a norm-one eigenvector of σ_A with eigenvalue s_i^2). This means that the Schmidt decomposition of the second purification necessarily reads

$$|\Phi_{AC}\rangle = \sum_{i=1}^{r} s_i (e^{i\phi_i}|e_i\rangle) \otimes |h_i\rangle = \sum_{i=1}^{r} s_i |e_i\rangle \otimes \underbrace{e^{i\phi_i}|h_i\rangle}_{=:|h_i'\rangle},$$

The $|f_i\rangle$ and $|h_i'\rangle$ each consist of r many orthonormal vectors. Since $\dim \mathcal{H}_B \leqslant \dim \mathcal{H}_C$, we can find an isometry $V_{B\to C}$ that sends $|f_i\rangle \mapsto |h_i'\rangle$ for $i = 1, \dots, r$. Clearly, this means that $|\Phi_{AC}\rangle = (I_A \otimes V_{B\to C})|\Psi_{AB}\rangle$, which is what we wanted to show.

If some eigenvalues of σ_A are degenerate, then we have some more freedom in the Schmidt decompositions. But just like we pushed the phases $e^{i\phi_i}$ from the first to the second tensor factor, we can always find two Schmidt decomposition that are 'aligned' as above (i.e., have the same s_i and $|e_i\rangle$). You may prove this in Practice Problem 3.5.

The Schmidt decomposition is a mild restatement of the singular value decomposition of operators, which we recall in the following.

Lemma 2.16 (Singular value decomposition). *Any operator $M \in L(\mathcal{H}, \mathcal{K})$ has a* singular value decomposition (SVD)*: That is, we can write*

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

> *where* $r = \operatorname{rank} M$*, the* $s_i > 0$*, the* $|e_i\rangle$ *are orthonormal in* $\mathcal{K}$*, and the* $|g_i\rangle$ *are orthonormal in* $\mathcal{H}$*. The* s_i *are called the* singular values *of* M.

In Practice Problem 2.4 you can prove precisely that the Schmidt decomposition follows from the singular value decomposition.

Proof of Lemma 2.16. 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). Consider the operator $MM^\dagger$, which is always positive semidefinite (Practice Problem 1.2 (c)), 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

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

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

$$r = \operatorname{rank} M^\dagger M \leqslant \operatorname{rank} M \leqslant r,$$

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. □

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 $M^\dagger M$).
- If $M = M^\dagger$, then the singular values are simply the *absolute* nonzero eigenvalues of M.

Let us conclude with some outlook. Next week, we will discuss distance measures between quantum states. The singular values will be an important tool for this. For example, if $M \in L(\mathcal{H}, \mathcal{K})$ then we can define its *trace norm* by $\|M\|_1 := \sum_i s_i$. The subscript reminds us that this is nothing but the ℓ^1-norm of the singular values (recall that $s_i > 0$). If M is Hermitian then this the same as $\|M\|_1 := \sum_i |m_i|$, where the m_i are the eigenvalues of M.

In particular, we may use this to define the *(normalized) trace distance* of two states ρ, σ:

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

(Note that $\rho - \sigma$ is a difference of Hermitian operators, so itself Hermitian.) You already saw this distance measure in Homework Problem 1.1. In this week's Homework Problem 2.2, you will derive a useful variational expression for it and show that it has a pleasant interpretation. Namely, the trace distance is directly related to how well we can distinguish ρ and σ by an arbitrary measurement (a result known as *Helstrom's theorem*). This answers the problem that we raised in Section 1.6 at the end of Lecture 1.

Lecture 3

Trace distance and fidelity, introduction to quantum channels

Today, we will be concerned with two separate topics. First, we will discuss ways of quantifying to what extent two quantum states are similar. One way to do this is by using last week's trace distance – and we will recall its most important properties below. Next, we will define the *fidelity*, which generalizes the overlap $|\langle\phi|\psi\rangle|$ between pure states and is often a convenient tool.

We will then switch gears for the remainder of the lecture and work towards the definition of a *quantum channel*. Roughly speaking, quantum channels describe the most general way by which we can modify (or 'process') a quantum state. We will first discuss the classical situation and then turn towards the quantum case – culminating in the definition of a quantum channel as a completely positive and trace-preserving superoperator. We will continue the discussion of channels next week. After this, we will have fully developed the basic formalism of quantum information theory.

3.1 Interlude: Norms of operators

Since quantum states are operators, we can in principle use any norm on $L(\mathcal{H})$ to define a distance measure. What are useful norms on operators? Let us briefly discuss this more generally. Given Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, we can define norms on $L(\mathcal{H}, \mathcal{K})$ by using the singular values. For $p \in [1, \infty)$, define the *Schatten* p*-norm* of an operator $M \in L(\mathcal{H}, \mathcal{K})$ by

$$\|M\|_p := \left(\sum_{i=1}^{r} s_i^p \right)^{1/p},$$

where $s_1, \ldots, s_r > 0$ denote the singular values of M. In other words, $\|M\|_p$ is the ℓ^p-norm of the singular values of M. If M is Hermitian, then the singular values are the absolute eigenvalues – so $\|M\|_p$ is the ℓ^p-norm of the eigenvalues. We will mostly be concerned with the following important special cases:

- $\|\cdot\|_1$ is called the *trace norm* (or nuclear norm):

$$\|M\|_1 := \sum_{i=1}^{r} s_i = \operatorname{Tr} \sqrt{M^\dagger M}. \tag{3.1}$$

On the right-hand side, we take the trace of the square root of the PSD operator $M^\dagger M$. In general, if Q is PSD then its *(positive semidefinite) square root* $\sqrt{Q}$ is the operator with the same eigenvectors but eigenvalues the square root of those of Q. That is, if $Q = \sum_i \lambda_i |e_i\rangle\langle e_i|$ is an eigendecomposition then $\sqrt{Q} = \sum_i \sqrt{\lambda_i} |e_i\rangle\langle e_i|$. Clearly, $\sqrt{Q}\sqrt{Q} = Q$. We used this notation already in Eq. (2.22).

Warning: In general, it is *not* true that $\sqrt{QR} = \sqrt{Q}\sqrt{R}$ for PSD Q and R. Indeed, QR will not even be PSD in general.

- $\|\cdot\|_2$ is called the *Frobenius norm* (or Hilbert-Schmidt norm):

$$\|M\|_2 := \left(\sum_{i=2}^{r} s_i^2 \right)^{1/2} = \sqrt{\operatorname{Tr} M^\dagger M} \tag{3.2}$$

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

$$\langle M, N \rangle_{HS} := \operatorname{Tr}[M^\dagger N] \qquad \forall M, N \in L(\mathcal{H}, \mathcal{K}). \tag{3.3}$$

Thus, $L(\mathcal{H}, \mathcal{K})$ is itself a Hilbert space if we use this inner product.

We can also extend the definition of $\|\cdot\|_p$ to $p = \infty$ by continuity. The result is the following:

- $\|\cdot\|_\infty$ is the *operator norm* (or spectral norm), defined by

$$\|M\|_\infty := \max_{i=1,\ldots,r} s_i = \max_{\||\phi\rangle\|=1} \|M|\phi\rangle\| \tag{3.4}$$

Let us discuss some useful properties, which hold for $p \in [1, \infty]$:

- The norms are invariant under taking the adjoint, as well as under conjugation and transposition (w.r.t. any orthonormal basis): $\|M\|_p = \|M^\dagger\|_p = \|\overline{M}\|_p = \|M^T\|_p$

- Since the norms $\|\cdot\|_p$ only depend on the singular values, they are *invariant under isometries* $V\colon \mathcal{K} \to \mathcal{K}'$, $W\colon \mathcal{H} \to \mathcal{H}'$ (i.e., $V^\dagger V = I_\mathcal{K}$, $W^\dagger W = I_\mathcal{H}$):

$$\|M\|_p = \|VMW^\dagger\|_p \tag{3.5}$$

In particular, they are invariant under left and right multiplication by unitaries.

- Next, the norms are *monotonically decreasing* in the parameter p as a direct consequence of the same property for the ordinary ℓ^p-norms. In particular:

$$\|M\|_1 \geqslant \|M\|_2 \geqslant \|M\|_\infty$$

- We also have a version of the *Hölder inequality*: For $\frac{1}{p} + \frac{1}{q} = 1$, it holds that $|\operatorname{Tr}[M^\dagger N]| \leqslant \|M\|_p \|N\|_q$ for all $M, N \in L(\mathcal{H}, \mathcal{K})$. In fact, $\|\cdot\|_p$ is the dual norm of $\|\cdot\|_q$. Again, we record the two most important special cases – the *Cauchy-Schwarz inequality*

$$|\operatorname{Tr}[M^\dagger N]| \leqslant \|M\|_2 \|N\|_2 \tag{3.6}$$

[which holds for any inner product, so in particular for Eq. (3.3)], and the *Hölder inequality for the trace and operator norm*:

$$|\operatorname{Tr}[M^\dagger N]| \leqslant \|M\|_1 \|N\|_\infty. \tag{3.7}$$

Proof. To see that the latter holds, let $M = \sum_i s_i |e_i\rangle\langle g_i|$ be an SVD. Then

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

to estimate the underbraced inner product, first use the Cauchy-Schwarz inequality (for vectors) and then the definition of the operator norm in Eq. (3.4). □

How about duality? Any Hilbert space is self-dual, so this is clear for $p = q = 2$. For $p = 1$ and $q = \infty$, duality is the first of the following two equations:

$$\|N\|_1 = \max_{M \in L(\mathcal{H},\mathcal{K}), \|M\|_\infty \leqslant 1} |\mathrm{Tr}[M^\dagger N]| \tag{3.8}$$

Proof. The direction '$\geqslant$' is just Eq. (3.7), while '$\leqslant$' can be seen by taking an SVD $N = \sum_i s_i |e_i\rangle\langle g_i|$ and evaluating the right-hand side for $M = \sum_i |e_i\rangle\langle g_i|$. □

The case that $\mathcal{H} = \mathcal{K}$ is so important that we re-state Eq. (3.8) in this case and add a slight variation. For $N \in L(\mathcal{H})$,

$$\|N\|_1 = \max_{M \in L(\mathcal{H}), \|M\|_\infty \leqslant 1} |\mathrm{Tr}[MN]| = \max_{U \in U(\mathcal{H})} |\mathrm{Tr}[MU]| \geqslant |\mathrm{Tr}[N]| \tag{3.9}$$

Proof. The first equality is just Eq. (3.8) for $\mathcal{H} = \mathcal{K}$. For the second equality, note that '$\geqslant$' holds since $\|U\|_\infty = 1$ for any unitary, while for '$\leqslant$' we merely note that if $\mathcal{H} = \mathcal{K}$ then we can choose U as a unitary that maps the left singular vectors onto the right singular vectors (while acting arbitrarily on the orthogonal complement). The inequality is obvious – simply choose $M = I$ or $U = I$. □

By combining the above, we can also prove variants of the Hölder inequalities for $M \in L(\mathcal{H}', \mathcal{K})$, $N \in L(\mathcal{H}, \mathcal{K})$ (with $\mathcal{H}$ and $\mathcal{H}'$ not necessarily the same space), such as the following:

$$\|M^\dagger N\|_1 \leqslant \|M\|_1 \|N\|_\infty. \tag{3.10}$$

Proof. By Eq. (3.8),

$$\|M^\dagger N\|_1 = \max_{X \in L(\mathcal{H},\mathcal{H}'), \|X\|_\infty \leqslant 1} |\mathrm{Tr}[X^\dagger M^\dagger N]|,$$

but

$$|\mathrm{Tr}[X^\dagger M^\dagger N]| = |\mathrm{Tr}[M^\dagger N X^\dagger]| \leqslant \|M\|_1 \|N X^\dagger\|_\infty \leqslant \|M\|_1 \|N\|_\infty \|X^\dagger\|_\infty \leqslant \|M\|_1 \|N\|_\infty,$$

where we first used Eq. (3.7), then submultiplicativity for the operator norm, and then that $\|X\|_\infty \leqslant 1$. □

If $\mathcal{H} = \mathcal{H}'$ then above strengthens Eqs. (3.6) and (3.7) [since the trace norm is never smaller than the trace, viz the inequality in Eq. (3.9)].

- Lastly, we note that the Schatten p-norms are all *submultiplicative*, which means that $\|MN\|_p \leqslant \|M\|_p \|M\|_p$ for all $N \in L(\mathcal{H}, \mathcal{K})$, $M \in L(\mathcal{K}, \mathcal{L})$.

3.2 Trace distance and fidelity

We can use the norms defined in Section 3.1 to define distance measures (metrics) between quantum states. One particular useful definition is the trace distance.

Definition 3.1 (Trace distance)**.** *The* (normalized) trace distance *between two states* $\rho, \sigma \in D(\mathcal{H})$ *is defined as follows:*

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

We already know several useful properties:

- $T(\rho, \sigma) \in [0, 1]$, and $T(\rho, \sigma) = 0$ if and only if $\rho = \sigma$.
- For pure $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$, the trace distance is directly related to the *overlap* $|\langle\phi|\psi\rangle|$:

  $$T(\rho, \sigma) = \sqrt{1 - |\langle\phi|\psi\rangle|^2} \tag{3.11}$$

  You proved this in Homework Problem 1.1.
- In general, we have the following variational formula:

  $$T(\rho, \sigma) = \max_{0 \leqslant Q \leqslant I} \operatorname{Tr}[Q(\rho - \sigma)] \tag{3.12}$$

  This directly implies Helstrom's theorem, which shows that the trace distance has the following *operational interpretation*: The optimal probability of distinguishing two equiprobable states ρ and σ is precisely $\frac{1}{2} + \frac{1}{2}T(\rho, \sigma)$. You proved this in Homework Problem 2.2.

On Practice Problem 3.2, you will furthermore show that:

- *Invariance under isometries*: That is, $T(\rho, \sigma) = T(V\rho V^\dagger, V\sigma V^\dagger)$ for any isometry V. [This follows directly from Eq. (3.5)].
- *Monotonicity:* $T(\rho_A, \sigma_A) \leqslant T(\rho_{AB}, \sigma_{AB})$ for all states ρ_{AB}, σ_{AB}. This is an intuitive property, since it means that two states can only get closer if we discard a subsystem. [This follows from Eq. (3.12), or also as a consequence of Helstrom's theorem! Can you see how?]

Is there also something like an overlap for mixed states? Yes, there is! The correct definition is as follows (although this will only become clear in view of Theorem 3.5 below).

Definition 3.2 (Fidelity)**.** *Given states* $\rho, \sigma \in D(\mathcal{H})$*, define their* fidelity *by*

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

For the second equality, see Eq. (3.1). Let us discuss some properties:

- $F(\rho, \sigma) \in [0, 1]$, and $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$. Note that the fidelity is a *similarity* measure rather than a distance measure (i.e., it is maximized if the two states are the same) – just like the overlap.

- *Symmetry:* $F(\rho, \sigma) = F(\sigma, \rho)$.
- If $\sigma = |\psi\rangle\langle\psi|$ is pure then $\sqrt{\sigma} = \sigma$, so
$$F(\rho, \sigma) = \mathrm{Tr}[\sqrt{|\psi\rangle \underbrace{\langle\psi|\rho|\psi\rangle}_{\geqslant 0} \langle\psi|}] = \sqrt{\langle\psi|\rho|\psi\rangle}\, \underbrace{\mathrm{Tr}[\sqrt{|\psi\rangle\langle\psi|}]}_{=\mathrm{Tr}|\psi\rangle\langle\psi|=1} = \sqrt{\langle\psi|\rho|\psi\rangle}. \tag{3.13}$$
If both states are pure, $\sigma = |\psi\rangle\langle\psi|$ and $\rho = |\phi\rangle\langle\phi|$, then
$$F(\rho, \sigma) = \sqrt{\langle\psi|\phi\rangle\langle\phi|\psi\rangle} = |\langle\psi|\phi\rangle|.$$
Thus, the fidelity indeed generalizes the overlap of pure states.
- Just like the trace distance, the fidelity is *invariant under isometries* V: $F(\rho, \sigma) = F(V\rho V^\dagger, V\sigma V^\dagger)$. (Can you see why?)

Remark 3.3 (Why so complicated?). *Why don't we simply use* $\sqrt{\mathrm{Tr}[\rho\sigma]}$ *to generalize the overlap? Whatever the definition, we would like that our quantity is maximized when both states are the same. But note that* $\mathrm{Tr}[\rho^2]$ *can be any number in* $[1/d, 1]$, $d = \dim \mathcal{H}$*, so the above is not a good definition.*

Remark 3.4 (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).*

The following central result gives a nice interpretation of the fidelity – it is simply the maximal overlap between any pair of purifications!

Theorem 3.5 (Uhlmann). *Let* $\rho_A, \sigma_A \in D(\mathcal{H}_A)$ *be states and* $\mathcal{H}_B$ *a Hilbert space such that both states admit purifications on* $\mathcal{H}_A \otimes \mathcal{H}_B$*. Then,*
$$F(\rho_A, \sigma_A) = \max\Big\{ |\langle\Psi_{AB}|\Phi_{AB}\rangle| \;:\; |\Psi_{AB}\rangle, |\Phi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \textit{ purifications of } \rho_A, \sigma_A \Big\} \tag{3.14}$$

Since any two purifications are related by a unitary [Lemma 2.11], Uhlmann's theorem can equivalently be stated as follows:
$$F(\rho_A, \sigma_A) = \max\Big\{ |\langle\Psi^{\text{fixed}}_{AB}|(I_A \otimes U_B)|\Phi^{\text{fixed}}_{AB}\rangle| \;:\; U_B \in U(\mathcal{H}_B) \Big\}, \tag{3.15}$$
where $|\Psi^{\text{fixed}}_{AB}\rangle$, $|\Phi^{\text{fixed}}_{AB}\rangle$ are arbitrary fixed purifications of ρ_A and σ_A, respectively.

Proof of Theorem 3.5 (if $\mathcal{H}_A = \mathcal{H}_B$*).* 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). Then we can use the standard purifications of ρ_A and σ_A, respectively. Recall from Eq. (2.22) that these are given by
$$|\Psi^{\text{std}}_{AB}\rangle := \left(\sqrt{\rho_A} \otimes I_B\right) \sum_x |x\rangle \otimes |x\rangle, \qquad |\Phi^{\text{std}}_{AB}\rangle := \left(\sqrt{\sigma_A} \otimes I_B\right) \sum_x |x\rangle \otimes |x\rangle, \tag{3.16}$$
where $|x\rangle$ is an arbitrary orthonormal basis of $\mathcal{H}_A = \mathcal{H}_B$. We will now prove Eq. (3.15), using these purifications as the 'fixed' purifications.
$$\begin{aligned}
|\langle\Psi^{\text{std}}_{AB}|(I_A \otimes U_B)|\Phi^{\text{std}}_{AB}\rangle| &= \sum_{x,y} \big(\langle x| \otimes \langle x|\big)\big(\sqrt{\rho_A}\sqrt{\sigma_A} \otimes U_B\big)\big(|y\rangle \otimes |y\rangle\big) \\
&= \sum_{x,y} \langle x|\sqrt{\rho_A}\sqrt{\sigma_A}|y\rangle\langle x|U|y\rangle \\
&= \sum_{x,y} \langle x|\sqrt{\rho_A}\sqrt{\sigma_A}|y\rangle\langle y|U^T|x\rangle \\
&= \sum_{x} \langle x|\sqrt{\rho_A}\sqrt{\sigma_A}U_A^T|x\rangle = \mathrm{Tr}\left[\sqrt{\rho_A}\sqrt{\sigma_A}U_A^T\right].
\end{aligned}$$

In the first step we inserted Eq. (3.16), next we used Eq. (2.3), then we perform the transpose (in the basis $|x\rangle$), and then we finally use Eqs. (1.2) and (1.3). (The unitaries $U_B = U = U_A$ are all the same objects – the subscripts are just notation to help us remember on which quantum system the operator acts.) By maximizing the left and right hand side of this equality, we obtain

$$\begin{aligned}\max_{U_B} |\langle\Psi^{\text{std}}_{AB}|(I_A \otimes U_B)|\Phi^{\text{std}}_{AB}\rangle| &= \max_{U_A} \operatorname{Tr}\left[\sqrt{\rho_A}\sqrt{\sigma_A}U_A^T\right] \\ &= \max_{U_A} \operatorname{Tr}\left[\sqrt{\rho_A}\sqrt{\sigma_A}U_A\right] = \|\sqrt{\rho_A}\sqrt{\sigma_A}\|_1 = F(\rho_A, \sigma_A).\end{aligned}$$

The second step uses that $U \mapsto U^T$ is a permutation of the set of unitaries and the third equality is precisely Eq. (3.12). Thus we have proved Eq. (3.15), and thereby the theorem. □

Proof of Theorem 3.5 (general case). 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. Now we can no longer use the standard purification. Instead we consider the following purifications:

$$|\Psi^{\text{fixed}}_{AB}\rangle := \left(\sqrt{\rho_A}V \otimes X\right)\sum_{x=1}^{r}|x\rangle \otimes |x\rangle, \qquad |\Phi^{\text{fixed}}_{AB}\rangle := \left(\sqrt{\sigma_A}W \otimes X\right)\sum_{x=1}^{r}|x\rangle \otimes |x\rangle. \tag{3.17}$$

Here, $r = \max\{\operatorname{rank}(\rho_A), \operatorname{rank}(\sigma_A)\}$ and we think of $\sum_{x=1}^{r}|x\rangle \otimes |x\rangle$ in $\mathcal{K} \otimes \mathcal{K}$, where $\mathcal{K} = \mathbb{C}^r$ is an auxiliary Hilbert space. The operator V is an isometry that the standard basis $|x\rangle$ of $\mathcal{K}$ to a subset of an orthonormal eigenbasis of ρ_A, such that the first $\operatorname{rank}(\rho_A)$ vectors correspond to the nonzero eigenvalues (this is possible since $\operatorname{rank}(\rho_A) \leqslant r \leqslant \dim \mathcal{H}_A$). Likewise, the operator W is an isometry that maps the standard basis $|x\rangle$ to a subset of an orthonormal eigenbasis of σ_A, again such that the first $\operatorname{rank}(\rho_B)$ vectors correspond to the nonzero eigenvalues. Finally, the operator X is an arbitrary isometry $\mathcal{K} \to \mathcal{H}_B$ (this exists since $r \leqslant \dim \mathcal{H}_B$, since we assumed that both ρ_A and σ_A have purifications to $\mathcal{H}_A \otimes \mathcal{H}_B$.) It is easy to verify that Eq. (3.17) defines purifications of ρ_A and σ_A, respectively.

We now proceed as before and consider Eq. (3.15), but now using Eq. (3.17) as our 'fixed' purifications. Now,

$$\begin{aligned}|\langle\Psi^{\text{fixed}}_{AB}|(I_A \otimes U_B)|\Phi^{\text{fixed}}_{AB}\rangle| &= \sum_{x,y=1}^{r}\left(\langle x| \otimes \langle x|\right)\left(V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}W \otimes X^\dagger U_B X\right)\left(|y\rangle \otimes |y\rangle\right) \\ &= \sum_{x,y=1}^{r}\langle x|V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}W|y\rangle\langle x|X^\dagger U_B X|y\rangle \\ &= \sum_{x,y=1}^{r}\langle x|V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}W|y\rangle\langle y|(X^\dagger U_B X)^T|x\rangle \\ &= \operatorname{Tr}[V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}W(X^\dagger U_B X)^T]\end{aligned}$$

What kind of object are $X^\dagger U_B X$ and its transpose? This is an operator on $\mathcal{K} = \mathbb{C}^r$ which we can think of as the restriction of the unitary U_B to a subspace (namely the image $\operatorname{im}(X)$ of the isometry X). As such, it is clear that $\|X^\dagger U_B X\|_\infty \leqslant 1$, 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

$$\max_{U_B} |\langle\Psi^{\text{fixed}}_{AB}|(I_A \otimes U_B)|\Phi^{\text{fixed}}_{AB}\rangle| \leqslant \max_{\|Y\|_\infty \leqslant 1} \operatorname{Tr}[V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}WY] = \|V^\dagger\sqrt{\rho_A}\sqrt{\sigma_A}W\|_1 \tag{3.18}$$

using the first characterization in Eq. (3.9). On the other hand, we can write any unitary matrix in $U(\mathcal{K})$ as $X^\dagger U_B X$ for some unitary $U_B \in U(\mathcal{H}_B)$ (simply choose U_B to be a direct sum of the desired unitary on $\mathcal{K} \cong \operatorname{im}(X)$ and, say, an identity matrix on the orthogonal complement). Hence,

$$\max_{U_B} |\langle \Psi_{AB}^{\text{fixed}}|(I_A \otimes U_B)|\Phi_{AB}^{\text{fixed}}\rangle| \geqslant \max_{Z \in U(\mathcal{K})} \operatorname{Tr}[V^\dagger \sqrt{\rho_A}\sqrt{\sigma_A} W Z] = \|V^\dagger \sqrt{\rho_A}\sqrt{\sigma_A} W\|_1 \tag{3.19}$$

Combining Eqs. (3.18) and (3.19), we find that

$$\max_{U_B} |\langle \Psi_{AB}^{\text{fixed}}|(I_A \otimes U_B)|\Phi_{AB}^{\text{fixed}}\rangle| = \|V^\dagger \sqrt{\rho_A}\sqrt{\sigma_A} W\|_1. \tag{3.20}$$

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

$$\|\sqrt{\rho_A}\sqrt{\sigma_A}\|_1 = \|VV^\dagger \sqrt{\rho_A}\sqrt{\sigma_A} WW^\dagger\|_1 \leqslant \|V^\dagger \sqrt{\rho_A}\sqrt{\sigma_A} W\|_1 \leqslant \|\sqrt{\rho_A}\sqrt{\sigma_A}\|_1 \tag{3.21}$$

The equality holds, because $VV^\dagger$ is projects onto the orthogonal complement 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 the Hölder inequality [Eq. (3.10)] 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 Eq. (3.5).) Since the left and the right hand side of Eq. (3.21) 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. (3.20) this concludes the proof. Phew! □

On Homework Problem 3.2, you will use Uhlmann's theorem to prove the following two important properties:

- Just like the trace distance, the fidelity satisfies a *monotonicity property*: $F(\rho_{AB}, \sigma_{AB}) \leqslant F(\rho_A, \sigma_A)$ for any two states $\rho_{AB}, \sigma_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ Note that the inequality goes 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.

- *Joint concavity:* $\sum_{i=1}^n p_i F(\rho_i, \sigma_i) \leqslant F(\sum_{i=1}^n p_i \rho_i, \sum_{i=1}^n p_i \sigma_i)$, where $(p_i)_{i=1}^n$ is an arbitrary probability distribution and $\rho_1, \dots, \rho_n$ and $\sigma_1, \dots, \sigma_n$ are arbitrary states.

Finally, we mention (without proof) that the trace distance and fidelity are related by the so-called *Fuchs-van de Graaf inequalities*: For all $\rho, \sigma \in D(\mathcal{H})$,

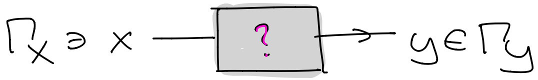
$$1 - F(\rho, \sigma) \leqslant T(\rho, \sigma) \leqslant \sqrt{1 - F^2(\rho, \sigma)}.$$

You can prove the upper bound in Practice Problem 5.1. For pure states, the upper bound is an equality, see Eq. (3.11). Moreover, it is sometimes useful to know that the function $P(\rho, \sigma) := \sqrt{1 - F^2(\rho, \sigma)}$ is a metric – called the *purified distance*.

3.3 Motivation: Channels in probability theory

So far, the only way to manipulate a quantum state has been to measure it – but we have not discussed at all how quantum states can be evolved or manipulated. (For example, you may know from a previous quantum mechanics class, or the ongoing quantum computing course, that we can always apply unitary operators to any quantum state.) In the remainder of today's lecture we will start developing the mathematical formalism that describes the most general ways that quantum states can be manipulated. Before considering the quantum situation, it is instructive to consider the classical situation.

Suppose we have 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 even if we use it many times. Then the most straightforward description might be to assume that there exists a function, $f\colon \Sigma_X \to \Sigma_Y$, 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 how about there is 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 are given an assignment

$$p(y|x) \quad \text{such that} \quad \begin{cases} p(y|x) \geqslant 0 & \forall x, y \\ \sum_y p(y|x) = 1 & \forall x. \end{cases} \tag{3.22}$$

The right-hand side conditions mean 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 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. In the context of Markov chains, $p(y|x)$ is often known as a *transition operator* or *Markov operator*.

Remark 3.6 (Functions as channels). *Note that given a function $f\colon \Sigma_X \to \Sigma_Y$, we can always define*

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

Then, if x is the input then $y = f(x)$ is the output with certainty. This shows that we can use channels to describe 'boxes' that simply apply a deterministic function.

What is 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,

$$p(y) = \sum_x p(y|x)p(x). \tag{3.23}$$

Here we used the slightly terrible (but concise) 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. (3.23) reads

$$p_Y(y) = \sum_x P_{Y|X}(y|x)p_X(x). \tag{3.24}$$

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. (3.22) mean that all entries are nonnegative and each column sums to one – such matrices are often called *(column) stochastic*. Then, the formula Eq. (3.24) for computing the output distribution given a channel and input distribution can be succinctly written as follows:

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

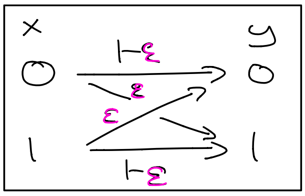
The mapping $p_X \mapsto p_Y$ is evidently linear (since it is implemented by left multiplication with the channel matrix $P_{Y|X}$). Conversely, any linear mapping that sends probability distributions to probability distributions must be of this form, with $P_{Y|X}$ a channel.

Let us discuss two families of channels that are very important from (classical) information theory.

1. 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:

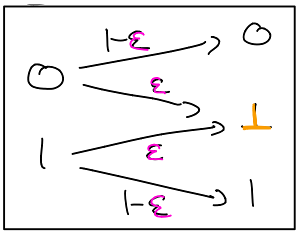


Note that output y does not contain any information about whether the bit has been flipped. This is perhaps the most straightforward way of modeling an unreliable (digital) information transmission line.

2. 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:

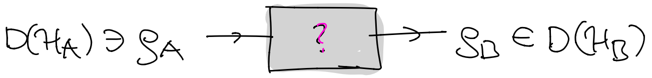


You could for example use this to describe a situation where you send a (physical or digital) packet from a sender to a receiver which sometimes gets lost.

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' evolutions. One of the central goals of information theory is to understand how to communicate reliably in the presense of uncertainty and noise.

3.4 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:



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

$$\Phi\colon\ L(\mathcal{H}_A) \to L(\mathcal{H}_B).$$

What additional properties should this maps 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 then

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

Remark 3.7. *Precisely speaking, this condition only justifies that Φ should be a* convex *map from $D(\mathcal{H}_A)$ to $D(\mathcal{H}_B)$. But any such map has a unique extension to linear map from $L(\mathcal{H}_A)$ to $L(\mathcal{H}_B)$.*

The fact that Φ is supposed to be linear can be succinctly written as follows:

$$\Phi \in L(L(\mathcal{H}_A), L(\mathcal{H}_B)). \tag{3.25}$$

Thus, Φ is an operator that maps operators to operators! Such maps are called *superoperator*. As for states and operators, we will use subscripts to indicate the labels of systems. Thus, we will write $\Phi_{A\to B}$ for a superoperator as in Eq. (3.25) and

$$\Phi_{A\to B}[\rho_A]$$

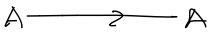
to apply a superoperator to a state $\rho_A \in L(\mathcal{H}_A)$, the result of which is an operator in $L(\mathcal{H}_B)$. We will consistently use square brackets $[\dots]$ to apply superoperators to operators.

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

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

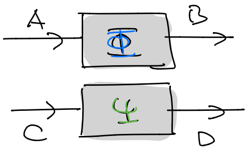
  This naturally describes the situation where our box does not change the input at all (or there is no box). We will visualize $\mathcal{I}_A$ as follows:



- Second, given two superoperators $\Phi_{A\to B}$ and $\Psi_{C\to D}$, we can always form their *tensor product*. This is a superoperator $\Phi_{A\to B} \otimes \Psi_{C\to D}$ from AC to BD, i.e., an element in $L(L(\mathcal{H}_A \otimes \mathcal{H}_C), L(\mathcal{H}_B \otimes \mathcal{H}_D))$. It is defined as follows on tensor product operators,

$$(\Phi_{A\to B} \otimes \Psi_{C\to D})[M_A \otimes N_C] := \Phi_{A\to B}[M_A] \otimes \Psi_{C\to D}[N_C], \tag{3.26}$$

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 we apply the first to one subsystem and the second to the other, as in the following picture:



What conditions to 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:

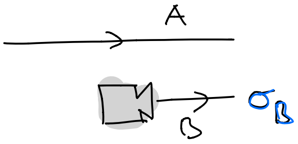
$$\rho_A \in D(\mathcal{H}_A) \quad \Rightarrow \quad \Phi_{A\to B}[\rho_A] \in D(\mathcal{H}_B)$$

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

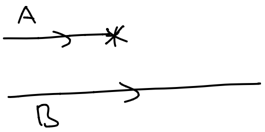
1. *Positive*, meaning it maps PSD operators to PSD operators: $\Phi[M_A] \geqslant 0$ for all $M_A \geqslant 0$,
2. *Trace-preserving:* $\mathrm{Tr}[\Phi[M_A]] = \mathrm{Tr}[M_A]$ for all M_A.

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

- *Basis change:* $\Phi[\rho] = U\rho U^\dagger$ for a fixed unitary or isometry U
- *Add state:* $\Phi_{A\to AB}[\rho_A] = \rho_A \otimes \sigma_B$ for a fixed state σ_B. This superoperator corresponds to a source that emits a quantum system in state σ_B – as in the figure:



- *Partial trace:* $\Phi_{AB\to A} = \mathrm{Tr}_A$. Indeed, the partial trace is a superoperator that maps states to states, as we discussed on p. 23. This corresponds to the situation where we simply discard a subsystem A:

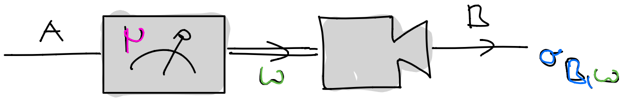


It is very instructive to note that we can write $\mathrm{Tr}_A = (\mathrm{Tr} \otimes \mathcal{I}_B)$, as can be seen by comparing Eqs. (2.10) and (3.26).

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

$$\Phi_{A\to B}[\rho_A] = \sum_{\omega\in\Omega} \operatorname{Tr}[\rho_A \mu_A(\omega)]\sigma_{B,\omega},$$

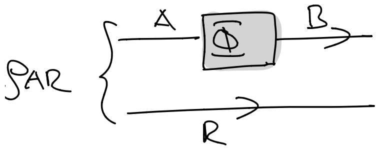
  where $\mu_A : \Omega \to \operatorname{PSD}(\mathcal{H}_A)$ is an arbitrary measurement on A and $\sigma_{B,\omega}$ a state on B for each possible outcome $\omega \in \Omega$. By Born's rule, $\operatorname{Tr}[\rho_A \mu_A(\omega)]$ is the probability of outcome ω using the measurement μ_A. Thus, the above superoperator corresponds to performing a measurement and then preparing a state labeled by the measurement outcome:



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

This is encouraging – we found many superoperators that are 'obviously reasonable' and map states to states.

However, we will now see that there is a *problem* – the two conditions above are not sufficient to single out 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 $(\Phi_{A\to B} \otimes \mathcal{I}_R)[\rho_{AR}]$ is *not* a state!



This is clearly nonsensical, since we want to interpret $\Phi \otimes \mathcal{I}_R$ 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):

$$\mathcal{T}[M] = M^T.$$

For concreteness, think of this as a qubit superoperator, i.e., from $L(\mathbb{C}^2)$ to $L(\mathbb{C}^2)$.

- 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 [Eq. (2.13)]:

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

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

$$\begin{aligned}(\mathcal{T} \otimes \mathcal{I}_R)[\rho_{AR}] &= \frac{1}{2}\left(|0\rangle\langle 0|^T \otimes |0\rangle\langle 0| + \underline{|0\rangle\langle 1|}^T \otimes |0\rangle\langle 1| + \underline{|1\rangle\langle 0|}^T \otimes |1\rangle\langle 0| + |1\rangle\langle 1|^T \otimes |1\rangle\langle 1|\right)\\ &= \frac{1}{2}\left(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + \underline{|1\rangle\langle 0|} \otimes |0\rangle\langle 1| + \underline{|0\rangle\langle 1|} \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|\right) = \frac{1}{2}\begin{pmatrix} 1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}\end{aligned}$$

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 $-1/2$.

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.

Definition 3.8 (Quantum channel). *A* (quantum) channel *is a superoperator* $\Phi_{A\to B} \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$ *that is both*

1. Completely positive*: For all* $\mathcal{H}_R$ *and* $M_{AR} \geqslant 0$*, it holds that* $(\Phi_{A\to B} \otimes \mathcal{I}_R)[M_{AR}] \geqslant 0$*,*
2. Trace-preserving*:* $\mathrm{Tr}[\Phi_{A\to B}[M_A]] = \mathrm{Tr}[M_A]$ *for all* M_A*.*

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

Remark 3.9. *Note that the second condition is unchanged. Indeed, unlike for positivity, it holds automatically that if* $\Phi_{A\to B}$ *is trace-preserving then so is* $\Phi_{A\to B} \otimes \mathcal{I}_R$ *for any system* R*.*

What are some examples of quantum channels?

- Clearly, the identity channel is a quantum channel according to this definition.
- All examples given above – except for the transpose map – are quantum channels. You will show this in Practice Problem 3.3 and Homework Problem 3.3.
- We can also build old channels from new ones. E.g., it follows almost by definition that if $\Phi_{A\to B}$ is a channel then so is $\Phi_{A\to B} \otimes \mathcal{I}_R$. More generally, channels can be composed in parallel (by $\Psi_{A\to B} \otimes \Phi_{C\to D}$), but also sequentially (by $\Phi_{B\to C} \circ \Psi_{A\to B}$). You can prove this in Practice Problem 3.4.

The set of quantum channels is a convex set.

Perhaps you might still feel a bit uneasy with this definition – perhaps there is another problem that we might have missed? 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 theory tells us that these three building blocks are all 'physical', this justifies the mathematical definition. See the discussion surrounding Axiom 4.5 for more details.

Remark 3.10 (Complete positivity for classical channels?). *In probability theory this problem does not appear. If* $p(y|x)$ *is a conditional probability distribution then so is* $p(yz'|xz) = p(y|x)\delta_{z,z'}$ *(in fact, this is the same as tensoring the transition matrix with* I_Z*). Thus, in a classical world, 'complete positivity' is automatic.*

Lecture 4

Structure of quantum channels

Last week, we defined the notion of a *quantum channel* as a completely positive and trace-preserving superoperator. 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.

4.1 Superoperators and complete positivity

Let $\Phi_{A\to B} \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$ 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 $\Phi_{A\to B}$ 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 $\Phi_{A\to B}$ 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.

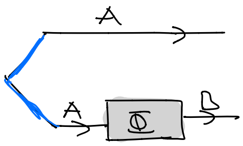
To start, let us define the *Choi operator* associated with the superoperator $\Phi_{A\to B}$ by

$$J^{\Phi}_{AB} := \sum_{x,y} |x\rangle\langle y| \otimes \Phi_{A\to B}[|x\rangle\langle y|] \in L(\mathcal{H}_A \otimes \mathcal{H}_B), \tag{4.1}$$

where $|x\rangle$ denotes an arbitrary orthonormal basis of $\mathcal{H}_A$. We can also write

$$J^{\Phi}_{AB} = \sum_{x,y} (\mathcal{I}_A \otimes \Phi_{A\to B})[|xx\rangle\langle yy|], \tag{4.2}$$

which makes it clear that J^{Φ}_{AB} is the result of applying $\Phi_{A\to B}$ to half of an *unnormalized* maximally entangled state $\sum_x |xx\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A$, as defined in Eq. (2.17). 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. (4.2):



The ‘blue bracket’ on the left-hand side is standard notation for a maximally entangled state.

For example, taking $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^\Sigma$ and the standard basis, the so-called *completely dephasing channel*

$$\Delta[\rho] = \sum_{x \in \Sigma} \langle x|\rho|x\rangle\, |x\rangle\langle x| \tag{4.3}$$

has the following Choi operator:

$$J^\Delta_{AB} = \sum_x |x\rangle\langle x| \otimes |x\rangle\langle x|, \tag{4.4}$$

an unnormalized maximally correlated state. You can verify this and more in Practice Problem 4.2.

In fact, the mapping $\Phi \mapsto J^\Phi$ defines an isomorphism, known as the *Choi-Jamiołkowski isomorphism*:

Lemma 4.1 (Choi-Jamiołkowski isomorphism). *The following map is an isomorphism,*

$$L(L(\mathcal{H}_A), L(\mathcal{H}_B)) \to L(\mathcal{H}_A \otimes \mathcal{H}_B), \quad \Phi_{A\to B} \mapsto J^\Phi_{AB},$$

with inverse given by

$$\Phi_{A\to B}[M_A] = \operatorname{Tr}_A\left[(M_A^T \otimes I_B) J^\Phi_{AB}\right] \qquad \forall M_A \in L(\mathcal{H}_A), \tag{4.5}$$

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. (4.5) by a direct calculation:

$$\begin{aligned}
\operatorname{Tr}_A\left[(M_A^T \otimes I_B) J^\Phi_{AB}\right] &= \sum_{x,y} \operatorname{Tr}_A\left[(M_A^T \otimes I_B)(|x\rangle\langle y| \otimes \Phi_{A\to B}[|x\rangle\langle y|])\right] \\
&= \sum_{x,y} \operatorname{Tr}_A\left[M_A^T |x\rangle\langle y| \otimes \Phi_{A\to B}[|x\rangle\langle y|]\right] \\
&= \sum_{x,y} \underbrace{\operatorname{Tr}[M_A^T |x\rangle\langle y|]}_{=\langle y|M_A^T|x\rangle = \langle x|M_A|y\rangle} \Phi_{A\to B}[|x\rangle\langle y|] \\
&= \sum_{x,y} \Phi_{A\to B}[|x\rangle\langle x|M_A|y\rangle\langle y|] = \Phi_{A\to B}[M_A].
\end{aligned}$$

□

It is a nice exercise to verify that this formula indeed recovers Eq. (4.3) from Eq. (4.4).

We now state the central theorem that gives four equivalent ways of characterizing when a superoperator is completely positive.

Theorem 4.2 (When is a superoperator completely positive?). *For a superoperator* $\Phi_{A\to B} \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$, *the following statements are equivalent:*

1. $\Phi_{A\to B}$ *is completely positive (i.e., for all* $\mathcal{H}_R$ *and* $M_{AR} \geqslant 0$ *it holds that* $(\Phi_{A\to B} \otimes \mathcal{I}_R)[M_{AR}] \geqslant 0$*).*

2. $\Phi_{A\to B} \otimes \mathcal{I}_{A'}$ *is positive (i.e., for all* $M_{AA'} \geqslant 0$ *it holds that* $(\Phi_{A\to B} \otimes \mathcal{I}_{A'})[M_{AA'}] \geqslant 0$*).*

3. $J^\Phi_{AB} \geqslant 0$*, i.e, the Choi operator of* $\Phi_{A\to B}$ *is positive semidefinite.*

4. Kraus representation: *There exist operators* $X_1, \dots, X_r \in L(\mathcal{H}_A, \mathcal{H}_B)$ *such that*

$$\Phi[M] = \sum_{i=1}^{r} X_i M X_i^\dagger \tag{4.6}$$

for all $M \in L(\mathcal{H}_A)$.

5. Stinespring representation: *There exists* $\mathcal{H}_E$ *and* $V \in L(\mathcal{H}_A, \mathcal{H}_B \otimes \mathcal{H}_E)$ *such that*

$$\Phi[M] = \mathrm{Tr}_E[VMV^\dagger] \tag{4.7}$$

for all $M \in L(\mathcal{H}_A)$.

Moreover, r *in* 4 *and* $\dim \mathcal{H}_E$ *in* 5 *can be chosen as* $\mathrm{rank}(J^\Phi_{AB}) \leqslant \dim \mathcal{H}_A \dim \mathcal{H}_B$ *(or larger).*

Proof. The implications 1⇒2⇒3 are immediate. For the implication 4⇒5, simply define $\mathcal{H}_E = \mathbb{C}^r$ and $V := \sum_{i=1}^r X_i \otimes |i\rangle$ and verify that Eq. (4.7) reduces to Eq. (4.6). (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 5⇒4.) The implication 5⇒1 is also easy – both $M \mapsto VMV^\dagger$ and Tr_E are completely positive (see Practice Problem 3.3, 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 3⇒4 with $r = \mathrm{rank}\, J^\Phi_{AB}$. Since J^Φ_{AB} is PSD, we can use a spectral decomposition to write

$$J^\Phi_{AB} = \sum_{i=1}^{r} |v_i\rangle\langle v_i| \quad \text{for suitable } v_i \in \mathcal{H}_A \otimes \mathcal{H}_B. \tag{4.8}$$

(To get this form, restrict to the positive eigenvalues $\lambda_i > 0$ and absorb their square root $\sqrt{\lambda_i}$ into the normalization of the eigenvectors $|v_i\rangle$.) The v_i are vectors in $\mathcal{H}_A \otimes \mathcal{H}_B$, but we need to construct operators X_i in $L(\mathcal{H}_A, \mathcal{H}_B)$. This we can do similarly as in Practice Problem 2.4. Simply define

$$X_i := \sum_{a,b} \langle ab|v_i\rangle\, |b\rangle\langle a| \in L(\mathcal{H}_A, \mathcal{H}_B). \tag{4.9}$$

Then, using Eq. (4.5),

$$\begin{aligned}
\Phi[M] &= \mathrm{Tr}_A\left[(M_A^T \otimes I_B) J^\Phi_{AB}\right] = \sum_i \mathrm{Tr}_A\left[(M_A^T \otimes I_B)|v_i\rangle\langle v_i|\right] \\
&= \sum_i \sum_{a,b} \sum_{a',b'} \langle ab|v_i\rangle\langle v_i|a'b'\rangle\, \mathrm{Tr}_A\left[(M_A^T \otimes I_B)|ab\rangle\langle a'b'|\right] \\
&= \sum_i \sum_{a,b} \sum_{a',b'} \langle ab|v_i\rangle\langle v_i|a'b'\rangle\, \mathrm{Tr}_A\left[M_A^T|a\rangle\langle a'| \otimes |b\rangle\langle b'|\right] \\
&= \sum_i \sum_{a,b} \sum_{a',b'} \langle ab|v_i\rangle\langle v_i|a'b'\rangle \underbrace{\mathrm{Tr}[M_A^T|a\rangle\langle a'|]}_{=\langle a'|M_A^T|a\rangle = \langle a|M_A|a'\rangle} |b\rangle\langle b'| \\
&= \sum_i \sum_{a,b} \sum_{a',b'} \langle ab|v_i\rangle |b\rangle\langle a|M_A|a'\rangle\langle b'|\langle v_i|a'b'\rangle \\
&= \sum_i X_i M_A X_i^\dagger,
\end{aligned}$$

which concludes the proof. □

Theorem 4.2 is rather remarkable. Criterion 2 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 3 shows that we do not even have to check that $\Phi_{A\to B} \otimes \mathcal{I}_A$ sends every PSD operator to a PSD operator – it suffices to check this condition for an (unnormalized, if we wish) maximally entangled state. You can practice this technique in Homework Problem 4.2.

Criteria 4 and 5 are also very useful in practice, since many quantum channels are naturally given in this form. Indeed, your homework last week would have tremendously simplified with Theorem 4.2!

Remark 4.3 (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* $\Phi[M] = VMW^\dagger$ *(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.* (4.8)*. As these representations are much less useful we did not discuss this in class.*

4.2 Characterizing quantum channels

With Theorem 4.2 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 4.4 (When is a completely positive superoperator trace-preserving?). *For a completely positive superoperator* $\Phi_{A\to B}$*, the following statements are equivalent:*

1. $\Phi_{A\to B}$ *is trace-preserving (hence a quantum channel).*
2. Choi operator: $\operatorname{Tr}_B[J^\Phi_{AB}] = I_A$.
3. Kraus representation: $\sum_i X_i^\dagger X_i = I_A$ *for one/every Kraus representation.*
4. Stinespring representation: $V^\dagger V = I_A$ *for one/every Stinespring representation. That is,* V *is an isometry.*

In fact, the equivalence between 1 and 2 holds for arbitary superoperators (completely positive or not).

Proof. We will use the fact, which follows from Practice Problem 4.1, that for $X, Y \in L(\mathcal{H})$ it holds that

$$\operatorname{Tr}(XM) = \operatorname{Tr}(YM) \text{ for all } M \in L(\mathcal{H}) \;\Leftrightarrow X = Y. \tag{4.10}$$

By Eq. (4.5) for any $M_A \in L(\mathcal{H}_A)$

$$\operatorname{Tr}[\Phi_{A\to B}(M_A)] = \operatorname{Tr}[\operatorname{Tr}_A[(M_A^T \otimes I_B)J^\Phi_{AB}]] = \operatorname{Tr}[M_A^T \operatorname{Tr}_B[J^\Phi_{AB}]]$$

from which it clearly follows using Eq. (4.10) that 2 and 1 are equivalent since $\operatorname{Tr}(M_A^T) = \operatorname{Tr}(M_A)$. Next, consider a Kraus representation of the channel, and use the cyclicity of the trace to see that for all $M_A \in L(\mathcal{H}_A)$

$$\operatorname{Tr}[\Phi_{A\to B}(M_A)] = \operatorname{Tr}\Big[\sum_i X_i M_A X_i^\dagger\Big] = \operatorname{Tr}\Big[M_A\Big(\sum_i X_i X_i^\dagger\Big)\Big].$$

So, again using Eq. (4.10), we see that 1 and 3 are equivalent. Finally, for a Stinespring representation, again using the cyclicity of the trace we see that

$$\mathrm{Tr}[\Phi_{A\to B}(M_A)] = \mathrm{Tr}[\mathrm{Tr}_E[VM_AV^\dagger]] = \mathrm{Tr}[M_AV^\dagger V]$$

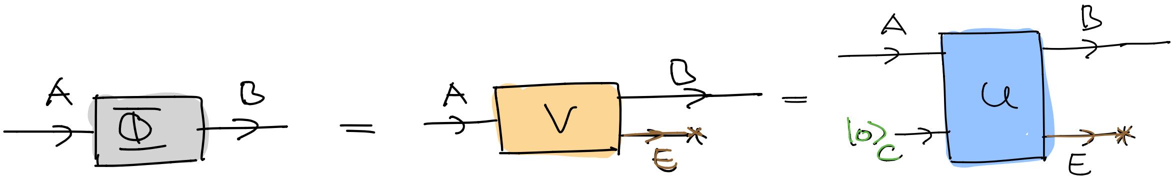
allowing us to conclude the equivalence of 1 and 4. □

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 but normalized to be a quantum state. This state is also known as the *Choi state* of $\Phi_{A\to B}$, and it is given by

$$\frac{1}{d_A} J^{\Phi}_{AB} = (\mathcal{I}_A \otimes \Phi_{A\to B})[|\Phi^+_{AA}\rangle\langle\Phi^+_{AA}|], \tag{4.11}$$

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 Practice Problem 4.5 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*:



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 4.5 (Channels). *Any quantum channel $\Phi_{A\to B}$ can be realized physically. That is, in principle, there exists a device that, given as input an arbitrary state ρ_A, outputs the state $\Phi_{A\to B}[\rho_B]$.*

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 Homework Problem 4.1 you will encounter an example of this proof strategy.

Remark 4.6 (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 $V_{A\to BE}, \tilde{V}_{A\to BE}$ of a channel $\Phi_{A\to B}$ are related by a unitary U_E on E, in the sense that*

$$\tilde{V}_{A\to BE} = (I_B \otimes U_E)V_{A\to BE}. \tag{4.12}$$

This follows from Lemma 2.11, because $|\Phi_{ABE}\rangle := (I_A \otimes V_{A\to BE})|\Phi^+_{AA}\rangle$ and $|\tilde{\Phi}_{ABE}\rangle := (I_A \otimes \tilde{V}_{A\to BE})|\Phi^+_{AA}\rangle$ are both purifications of the Choi state Eq. (4.11) 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. (4.12).

As a consequence, any two sets $\{X_i\}_{i=1}^r, \{Y_i\}_{i=1}^r$ of Kraus operators for a channel $\Phi_{A\to B}$ 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 $i = 1,\ldots,r$. This can be seen *by constructing the Stinespring isometries corresponding to these Kraus operators as in the proof of Theorem 4.2.*

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. (3.3) that the Hilbert-Schmidt inner product on $L(\mathcal{H})$ is given by $\langle M, N\rangle_{HS} = \operatorname{Tr}[M^\dagger N]$. This allows us to define the *adjoint* of a superoperator $\Phi \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$. Explicitly, this is the superoperator $\Phi^\dagger \in L(L(\mathcal{H}_B), L(\mathcal{H}_A))$ such that

$$\langle M_A, \Phi^\dagger[N_B]\rangle_{HS} = \langle \Phi[M_A], N_B\rangle_{HS} \qquad \forall M_A \in L(\mathcal{H}_A), N_B \in L(\mathcal{H}_B).$$

In Practice Problem 4.6 you will show the adjoint of a completely positive superoperator is again completely positive, and the ajoint of a trace-preserving superoperator is unital (meaning that $\Phi^\dagger[I_B] = I_A$).

Lecture 5

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. 9 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/.

5.1 Shannon entropy

Today we will work with classical probability distributions a lot, so let us denote by

$$P(\Sigma) := \left\{ p \colon \Sigma \to \mathbb{R}_{\geqslant 0} : \sum_{x \in \Sigma} p(x) = 1 \right\} \tag{5.1}$$

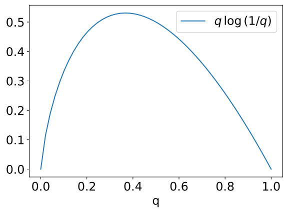
the set of all probability distributions on a finite set Σ. If X is a random variable then write $X \sim p$ to say that X is distributed according to p, i.e., $\Pr(X = x) = p(x)$ for all $x \in \Sigma$. As usual, we write $E[X] = \sum_{x \in \Sigma} p(x) x$ for the *expectation value* and $\operatorname{Var}(X) = E[X^2] - E[X]^2$ for the *variance* of a numerical random variable X. We now define the Shannon entropy.

Definition 5.1 (Shannon entropy). *The* Shannon entropy *of a probability distribution* $p \in P(\Sigma)$ *is defined by*

$$H(p) := \sum_{x \in \Sigma} p(x) \log \frac{1}{p(x)} = -\sum_{x \in \Sigma} p(x) \log p(x). \tag{5.2}$$

Here and throughout these lecture notes, log *always denotes the logarithm to* base 2 *(i.e.,* $\log 2 = 1$*).*

As stated, Eq. (5.2) 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 $q \to 0$, as illustrated in the following figure:



We can thus extend the definition of $H(p)$ by continuity, i.e., defining $p(x)\log\frac{1}{p(x)} = -p(x)\log p(x) = 0$ for $p(x) = 0$ in Eq. (5.2). Then $H(p)$ is a *continuous* function of $p \in P(\Sigma)$. This definition is also compatible with the interpretation that the Shannon entropy can be written as

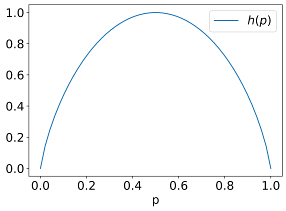
$$H(p) = E\left[\log\frac{1}{p(X)}\right] = -E[\log p(X)], \tag{5.3}$$

where $X \sim p$, since outcomes that appear with probability zero do not impact the expectation value.

Example 5.2 (Binary entropy function)**.** *The Shannon entropy of a probability distribution with two possible outcomes is given by the so-called* binary entropy function,

$$h(p) := H(\{p, 1-p\}) = p\log\frac{1}{p} + (1-p)\log\frac{1}{1-p},$$

where p *is the probability of one of the outcomes. This function looks as follows (cf. Practice Problem 7.5):*



Note that it is indeed continuous, but not Lipschitz continuous, as the derivative diverges for $p \to 0, 1$.

We now list some further properties of the Shannon entropy.

- *Nonnegativity:* $H(p) \geqslant 0$. 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).

  Proof. The lower bound holds since $f(q) = q\log\frac{1}{q} \geqslant 0$ 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. □

Before we state the next property, recall that a function $f\colon \mathbb{D} \to \mathbb{R}$ defined on a convex set $\mathbb{D}$ (e.g., an interval) is called *concave* if it holds that $qf(a) + (1-q)f(b) \leqslant f(qa + (1-q)b)$ 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 $f'' \leqslant 0$ then f is concave. If $f'' < 0$ then f is strictly concave.

Jensen's inequality states that, for any concave function f as above,

$$\sum_{x\in\Sigma} p(x)f(a(x)) \leqslant f\Big(\sum_{x\in\Sigma} p(x)a(x)\Big) \tag{5.4}$$

for any probability distribution $p \in P(\Sigma)$ and function $a\colon \Sigma \to \mathbb{D}$. (If $|\Sigma| = 2$ then this simply restates the definition of concavity.) Moreover, if f is strictly concave then equality in Eq. (5.4) holds if and only if a is constant on the set $\{x \in \Sigma : p(x) > 0\}$. We can also state Eq. (5.4) in probabilistic terms. If f is a concave function and A a random variable then

$$E[f(A)] \leqslant f(E[A]),$$

and for a strictly concave function we have equality iff A is constant.

- *Upper bound:*
  $$H(p) \leqslant \log \left|\{x : p(x) > 0\}\right| \leqslant \log |\Sigma|$$
  Moreover, $H(p) = \log|\Sigma|$ if and only if p is *uniform*, i.e., $p(x) = 1/|\Sigma|$ for all $x \in \Sigma$.

  Proof. This follows from Jensen's inequality, applied to the concave log function and $a(x) = 1/p(x)$. Indeed,
  $$H(p) = \sum_{x\in\Sigma,p(x)>0} p(x) \log \frac{1}{p(x)} \leqslant \log \sum_{x\in\Sigma,p(x)>0} p(x)\frac{1}{p(x)} = \log \left|\{x : p(x) > 0\}\right|,$$
  with equality if and only if all nonzero $p(x)$ are equal. Now the rest is clear. □

- *Concavity:* The Shannon entropy is a strictly concave function of $p \in P(\Sigma)$.

  Proof. 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 $q \in [0, \infty)$. Indeed, for $q > 0$,
  $$f'(q) = -\frac{1}{\ln 2} (\ln q + 1) \qquad \text{and so} \qquad f''(q) = -\frac{1}{\ln 2}\frac{1}{q} < 0.$$
  □

Definition 5.3 (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,*
$$\begin{aligned} p_{XY}(x, y) &= \Pr(X = x, Y = y), \\ p_X(x) &= \Pr(X = x) = \textstyle\sum_y p_{XY}(x, y), \\ p_Y(y) &= \Pr(Y = y) = \textstyle\sum_x p_{XY}(x, y). \end{aligned}$$
We already discussed and used this convention in Eq. (3.24). It will also be useful to write Σ_X for the domain of a random variable X, i.e., if $X \in P(\Sigma_X)$. This is completely analogous to our notation and conventions in the quantum case, see Definitions 2.5 and 2.8.

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.

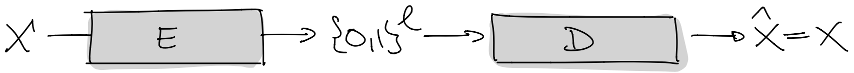
Today we will use this notation only to state the following two properties, which you will prove in Homework Problem 5.3.

- *Monotonicity:* $H(XY) \geqslant H(X)$ and $H(XY) \geqslant H(Y)$.
- *Subadditivity:* $H(X) + H(Y) \geqslant H(XY)$.

We now turn towards today's main goal, which is to give an interpretation of the Shannon entropy in the context of compression.

5.2 Lossy and lossless compression

Consider a data source modeled by a random variable $X \sim p \in P(\Sigma)$. We would like to compress 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 $\hat{X} := D(E(X))$ is equal to X (i.e., first compressing and then decompressing does recover the original input). This is illustrated in the following picture:



(5.5)

How small can we choose ℓ to be? The answer is given by the *raw bit content* of p, which is defined as follows:

$$H_0(p) := \log \left|\{x \in \Sigma : p(x) > 0\}\right|.$$

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 $\ell \geqslant H_0(p)$. 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:

1. *Lossy fixed-length compression:* We could allow a small probability of error, i.e., only demand that $\Pr(\hat{X} \neq X) \leqslant \delta$ for some $\delta > 0$.

2. *Lossless variable-length compression:* We could use bitstrings of different lengths $\ell = \ell(x)$ and try to minimize the average length.

Here is a concrete example:

Example 5.4. *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 (5.5).

However, 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)*
A	*0*
B	*1*
C	*1*

s	*D(s)*
0	*A*
1	*B*

Similarly, if we are willing to use bitstrings of varying length then the following encoder and decoder

x	*E(x)*
A	*0*
B	*10*
C	*11*

s	*D(s)*
0	*A*
10	*B*
11	*C*

achieves an average length of $0.98 \times 1 + 0.02 \times 2 = 1.01$.

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 Practice Problem 5.4.

5.3 Block codes, Shannon's source coding theorem, typical sets

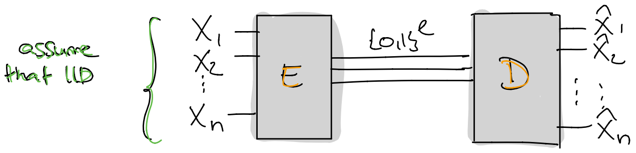
The basic assumption will be that our data source is *IID (or memoryless)*, which means that it emits symbols

$$X_1, X_2, \dots, X_n \overset{\text{IID}}{\sim} p$$

for some $p \in P(\Sigma)$. This notation means that the X_i are *independent and identically distributed (IID)* random variables such that each X_i has distribution p.

Remark 5.5. *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

$$\Pr(\hat{X}^n \neq X^n) \leqslant \delta.$$

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. (5.5). 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 $n \to \infty$).

Definition 5.6 (Code). *An* (n, R, δ)-code *for* $p \in P(\Sigma)$ *is a pair of functions*

$$E\colon \Sigma^n \to \{0,1\}^{\lfloor nR \rfloor} \qquad \textit{and} \qquad D\colon \{0,1\}^{\lfloor nR \rfloor} \to \Sigma^n$$

such that

$$\Pr\big(D(E(X^n)) \neq X^n\big) \leqslant \delta \tag{5.6}$$

for $X^n \overset{\text{IID}}{\sim} p$.

Note that the left-hand side of Eq. (5.6) can also be written as

$$\Pr\big(D(E(X^n)) \neq X^n\big) = \sum_{x^n \in \Sigma^n : D(E(x^n)) \neq x^n} p(x^n) = \sum_{x^n \in \Sigma^n : D(E(x^n)) \neq x^n} p(x_1) \cdots p(x_n),$$

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 5.7 (Shannon's source coding). *Let* $p \in P(\Sigma)$ *and* $\delta \in (0,1)$. *Then:*

1. *If* $R > H(P)$ *then there exists* n_0 *such that there exists an* (n, R, δ)*-code for all* $n \geqslant n_0$.
2. *If* $R < H(P)$ *then there exists* n_0 *such that no* (n, R, δ)*-codes exist for* $n \geqslant n_0$.

That is, the *optimal asymptotic compression rate* for an *IID source* described by a probability distribution is given by the Shannon entropy.

To prove Theorem 5.7, 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 5.8 (Typical set). *For* $p \in P(\Sigma)$, $n \in \mathbb{N}$, *and* $\varepsilon > 0$, *define the* typical set

$$\begin{aligned} T_{n,\varepsilon}(p) &:= \left\{ x^n \in \Sigma^n : \left| \frac{1}{n} \log \frac{1}{p(x^n)} - H(p) \right| \leqslant \varepsilon \right\} \\ &= \left\{ x^n \in \Sigma^n : \left| \frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(x_i)} - H(p) \right| \leqslant \varepsilon \right\} \end{aligned}$$

The following lemma summarizes the most important properties of the typical sets.

Lemma 5.9 (Asymptotic Equipartition Property, AEP). *The following properties hold:*

0. $2^{-n(H(p)+\varepsilon)} \leqslant p(x^n) \leqslant 2^{-n(H(p)-\varepsilon)}$ *for all* $x^n \in T_{n,\varepsilon}(p)$.
1. $|T_{n,\varepsilon}(p)| \leqslant 2^{n(H(p)+\varepsilon)}$.
2. *For* $X^n \overset{\text{IID}}{\sim} p$, *it holds that* $\Pr\big(X^n \notin T_{n,\varepsilon}(p)\big) \leqslant \frac{\sigma^2}{n\varepsilon^2}$. *Here,* $\sigma^2 = \operatorname{Var}(\log \frac{1}{p(X_i)})$ *is a constant that only depends on* p.

Proof. 0. This is just restating the definition.

1. This follows from

$$1 \geqslant \Pr\big(X^n \in T_{n,\varepsilon}(p)\big) \geqslant |T_{n,\varepsilon}(p)|\, 2^{-n(H(p)+\varepsilon)},$$

where the last step is the lower bound in part 0.

2. 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. (5.3)) and variance σ^2. Now,

$$\Pr\big(X^n \notin T_{n,\varepsilon}(p)\big) = \Pr\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} \log \frac{1}{p(X_i)} - H(p)\Big| > \varepsilon\Big) = \Pr\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} R_i - \mu\Big| > \varepsilon\Big)$$

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,

$$E[Y] = \mu \qquad \text{and} \qquad \operatorname{Var}(Y) = \frac{1}{n^2}\operatorname{Var}(R_1 + \dots + R_n) = \frac{1}{n}\operatorname{Var}(R_i) = \frac{\sigma^2}{n},$$

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

$$\Pr(|Y - E[Y]| > \varepsilon) \leqslant \frac{\operatorname{Var}(Y)}{\varepsilon^2}$$

to conclude the proof.

□

We are now in a good position to prove Shannon's source coding theorem.

Proof of Theorem 5.7. To prove part 1, let us choose $\varepsilon = \frac{R-H(p)}{2}$, noting that $\varepsilon > 0$. Then, using part 1 of Lemma 5.9,

$$|T_{n,\varepsilon}(p)| \leqslant 2^{n(H(p)+\varepsilon)} = 2^{n(R-\varepsilon)} \leqslant 2^{\lfloor nR \rfloor};$$

the final inequality holds provided we assume that $n \geqslant \frac{1}{\varepsilon}$. The above implies that there exists an *injective* map $E\colon T_{n,\varepsilon} \to \{0,1\}^{\lfloor nR \rfloor}$. Let us denote by $D\colon \{0,1\}^{\lfloor nR \rfloor} \to \Sigma^n$ 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,

$$\Pr\big(D(E(X^n)) \neq X^n\big) \leqslant \Pr(X^n \notin T_{n,\varepsilon}(p)) \leqslant \frac{\sigma^2}{n\varepsilon^2} \leqslant \delta,$$

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 2 of Lemma 5.9; the final inequality holds if we assume that $n \geqslant \frac{\sigma^2}{\varepsilon^2\delta}$. Thus we have proved that there exists an (n, R, δ)-code for any $n \geqslant n_0 := \max\{\frac{1}{\varepsilon}, \frac{\sigma^2}{\varepsilon^2\delta}\}$. We emphasize that n_0 only depends on p, δ, and R, as it should.

How about the proof of part 2? This is your Homework Problem 5.4! □

In Practice Problem 5.3 you can reflect on the practicalities of using typical sets for compression. In Practice Problem 5.4 you can discuss how to translate an (n, R, δ)-code into a corresponding lossless variable-length compression protocol.

Remark 5.10. *The typical sets constructed in the proof are in general not the smallest sets S_n with the property that $\Pr(X^n \in S_n) \geqslant 1 - \delta$. However, they are easy to handle mathematically as $n \to \infty$ and still small enough (this is the content of part 2 of Theorem 5.7).*

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!

Lecture 6

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 5.7 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.

6.1 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 6.1 (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 5.1). That is,*

$$H(\rho) := H(p) \tag{6.1}$$

where $p = (p(1), \ldots, p(d))$ *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:

$$H(\rho) = -\operatorname{Tr}[\rho \log \rho] \tag{6.2}$$

On the right-hand side, we take the logarithm of the operator ρ. 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 completely analogous to the definition of the square root $\sqrt{Q}$ in Lectures 2 and 3. Note that $\log Q$ is typically not PSD.

If ρ has some zero eigenvalues then $\log(\rho)$ is ill-defined. However, we can still define $\rho \log(\rho)$ by continuity for all $\rho \in D(\mathcal{H})$, in precisely the same way that we did for the Shannon entropy (see discussion below Definition 5.1). Then Eq. (6.2) is well-defined and holds for all $\rho \in D(\mathcal{H})$.

We now state some properties of the von Neumann entropy. The first two follow immediately from the corresponding properties of the Shannon entropy (see p. 52).

- *Nonnegativity:* $H(\rho) \geqslant 0$. Moreover, $H(\rho) = 0$ if and only if ρ is *pure* (i.e., $\rho = |\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle \in \mathcal{H}$).

- *Upper bound:*

$$H(\rho) \leqslant \log \operatorname{rank}(\rho) \leqslant \log \dim \mathcal{H}.$$

  Moreover, $H(\rho) = \log \dim \mathcal{H}$ if and only if ρ is *maximally mixed* (i.e., $\rho = \frac{I}{\dim \mathcal{H}}$).

- *Invariance under isometries:* $H(\rho) = H(V\rho V^\dagger)$ for any isometry V. This holds since the entropy only depends on the nonzero eigenvalues – but the latter are the same for ρ and $V\rho V^\dagger$.

- *Continuity:* The von Neumann entropy is continuous. This follows because the Shannon entropy is continuous and the sorted eigenvalues of a Hermitian operator depend continuously of an operator (but we will not prove this). In case you are curious about quantitative bounds: The *Fannes-Audenaert inequality* states that, for all $\rho, \sigma \in D(\mathcal{H})$,

$$|H(\rho) - H(\sigma)| \leqslant t \log(\dim \mathcal{H} - 1) + h(t),$$

  where $t = T(\rho, \sigma)$ is the trace distance between the two states and $h(t)$ denotes the binary Shannon entropy discussed in Example 5.2 and Practice Problem 7.5.

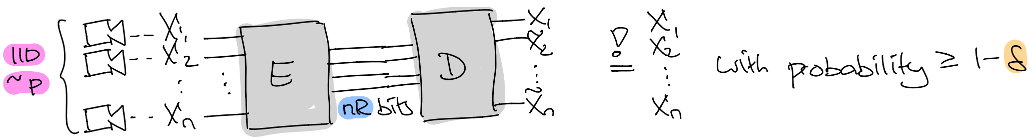
- *Concavity:* The von Neumann entropy is a strictly concave function of $\rho \in D(\mathcal{H})$. You will prove concavity in Homework Problem 6.4 (c) and strict concavity in Practice Problem 7.3. See p. 52 in the last lecture for the definition of concavity and strict concavity.

6.2 Motivation: Classical compression and correlations

Before we turn to compressing quantum data, let us briefly revisit the classical case. Recall from Definition 5.6 that an (n, R, δ)-code for a probability distribution $p \in P(\Sigma)$ consists of functions $E\colon \Sigma^n \to \{0,1\}^{\lfloor nR \rfloor}$ and $D\colon \{0,1\}^{\lfloor nR \rfloor} \to \Sigma^n$ such that

$$\Pr(\tilde{X}^n \neq X^n) \leqslant \delta \tag{6.3}$$

for $X_1, \ldots, X_n \overset{\text{IID}}{\sim} p$, where $\tilde{X}^n := D(E(X^n))$. Pictorially:



(6.4)

Shannon's source coding theorem asserts that $H(p)$ is the optimal rate for compression in this context (see Theorem 5.7 for the precise statement).

How about if X^n is *correlated* to another random variable Y? For example, suppose that $Y = X_1$, or $Y = X_1 \oplus \ldots \oplus X_n$ 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^nY} denote the joint distribution of (X^n, Y) and let $p_{\tilde{X}^nY}$ denote the joint distribution of $(\tilde{X}^n, Y)$. Then we would like to ask if it is true that $p_{X^nY} \approx p_{\tilde{X}^nY}$. This can be quantified by using the trace distance for probability distributions which is defined as follows:

Definition 6.2 (Trace distance)**.** *Given probability distributions* $p, q \in P(\Sigma)$*, 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 usual* ℓ_1*-norm of vectors.*

Note that this is nothing but the trace distance of the corresponding classical states. In Practice Problem 6.1, you will prove the following two properties:

1. If Z, $\tilde{Z}$ are random variables over Σ with distributions p, q, respectively, then

$$T(p, q) = \max_{S \subseteq \Sigma} \big(\Pr(Z \in S) - \Pr(\tilde{Z} \in S)\big). \tag{6.5}$$

2. If Z and $\tilde{Z}$ are as above and have a joint distribution then it holds that

$$T(p, q) \leqslant \Pr(Z \neq \tilde{Z}). \tag{6.6}$$

   Eq. (6.6) 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 6.3. *Let* $p \in P(\Sigma)$ *and* $E \colon \Sigma^n \to \{0, 1\}^{\lfloor Rn \rfloor}$, $D \colon \{0, 1\}^{\lfloor Rn \rfloor} \to \Sigma^n$ *be an arbitrary pair of functions. Then,* (E, D) *is an* (n, R, δ)*-code for* p *if and only if*

$$T(p_{X^n Y}, p_{\tilde{X}^n Y}) \leqslant \delta$$

for any joint distribution $p_{X^n Y}$ *of random variables* $X_1, \dots, X_n \overset{\text{IID}}{\sim} p$ *and* Y*, where* $p_{\tilde{X}^n Y}$ *denotes the joint distribution of* $\tilde{X}^n = D(E(X^n))$ *and* Y*.*

Proof. (⇒): Using the coupling inequality Eq. (6.6) for $Z = (X^n, Y)$ and $\tilde{Z} = (\tilde{X}^n, Y)$,

$$T(p_{X^n Y}, p_{\tilde{X}^n Y}) \leqslant \Pr(Z \neq \tilde{Z}) = \Pr(X^n \neq \tilde{X}^n) \leqslant \delta,$$

where the last inequality is Eq. (6.3), using that (E, D) is by assumption an (n, R, δ)-code.

(⇐): Choose $Y = X^n$. Then,

$$\Pr(\tilde{X}^n \neq X^n) = \Pr(\tilde{X}^n \neq Y) = \Pr(\tilde{X}^n \neq Y) - \underbrace{\Pr(X^n \neq Y)}_{=0} \leqslant T(p_{\tilde{X}^n Y}, p_{X^n Y}) \leqslant \delta,$$

where the first inequality is Eq. (6.5) for the event $S = \{(x^n, y) : x^n \neq y\}$. □

6.3 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 3.8 that $C(\mathcal{H}_A, \mathcal{H}_B)$ denotes the set of all quantum channels from $\mathcal{H}_A$ to $\mathcal{H}_B$.

Definition 6.4 (Quantum code). *An* (n, R, δ)-quantum code *for* $\rho \in D(\mathcal{H}_A)$ *is a pair of channels*

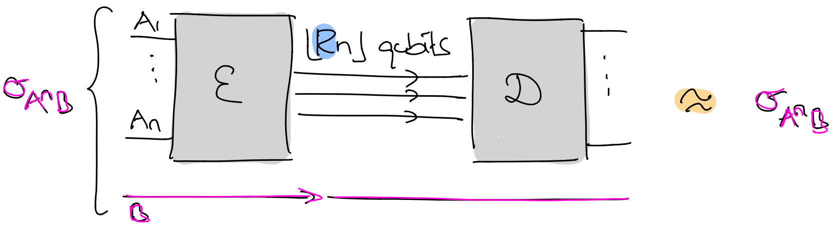
$$\mathcal{E} \in C\big(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}\big) \qquad \textit{and} \qquad \mathcal{D} \in C\big((\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}, \mathcal{H}_A^{\otimes n}\big)$$

such that

$$F\big(\sigma_{A^nB}, (\mathcal{D} \circ \mathcal{E} \otimes \mathcal{I}_B)[\sigma_{A^nB}]\big) \geqslant 1 - \delta \tag{6.7}$$

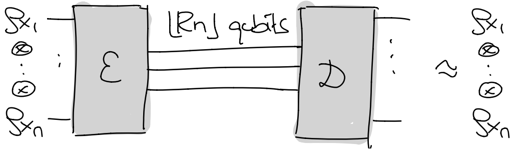
for all finite-dimensional $\mathcal{H}_B$ *and states* $\sigma_{A^nB} \in D(\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B)$ *such that* $\sigma_{A^n} = \rho_A^{\otimes n}$.

Here we use the fidelity rather than the trace distance – otherwise Definition 6.4 is completely analogous to the condition in Lemma 6.3. The following pictures illustrates the definition:



Definition 6.4 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. (6.7) so that it no longer refers to infinitely many options for σ_{A^nB}? (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 6.4 to compression of a source. In analogy to last lecture, 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 $\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}$ according to the IID distribution $p(x^n) = p(x_1) \cdots p(x_n)$. 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

$$\sum_{x^n \in \Sigma^n} p(x_1) \cdots p(x_n) F\big(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}, \mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]]\big) \geqslant 1 - \delta. \tag{6.8}$$

This looks similar to Eqs. (6.3) and (6.4), 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. (6.8) 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. (6.7), 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,

$$\sum_{x^n \in \Sigma^n} p(x^n)\, F\big(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}, \mathcal{D}[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]]\big) = F\big(\sigma_{A^n Y}, (\mathcal{D} \circ \mathcal{E} \otimes \mathcal{I}_{X^n})[\sigma_{A^n X^n}]\big) \geqslant 1 - \delta,$$

where the equality holds thanks to last week's Homework Problem 5.1 and the inequality is simply by Eq. (6.7) in the definition of a quantum code, applied to the state $\sigma_{A^n X^n}$.

Thus we have proved that Eq. (6.7) implies Eq. (6.8), meaning that a quantum code for ρ can be used for compressing any quantum source with average output state ρ. In Homework Problem 6.1 you will show that in general the converse is *not* true. This makes sense, since Eq. (6.8) refers to a single source, while we just proved that Eq. (6.7) 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 $\rho = \sum_x p(x) \rho_x$.

6.4 Channel fidelity

We now turn to the second question raised above – how can we check the condition in Eq. (6.7) without having to consider all possible states $\sigma_{A^n B}$? We start with a definition that abstracts the situation.

Definition 6.5 (Channel fidelity). *Given a channel $\mathcal{T}_A \in C(\mathcal{H}_A, \mathcal{H}_A)$ and a state ρ_A, define the* channel fidelity *as*

$$F(\mathcal{T}_A, \rho_A) := \inf \big\{ F\big(\sigma_{AB}, (\mathcal{T}_A \otimes \mathcal{I}_B)[\sigma_{AB}]\big) \;:\; \mathcal{H}_B,\ \sigma_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B) \text{ such that } \sigma_A = \rho_A \big\}.$$

Given this definition, we can rephrase Eq. (6.7) in the definition of a quantum code as

$$F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geqslant 1 - \delta. \tag{6.9}$$

Why is this progress? It turns out that we can always compute the channel fidelity by considering an arbitrary purification.

Lemma 6.6. *Let* $\sigma_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ *be an arbitrary purification of* ρ_A*. Then,*

$$F(\mathcal{T}_A, \rho_A) = F\big(\sigma_{AB}, (\mathcal{T}_A \otimes \mathcal{I}_B)[\sigma_{AB}]\big).$$

Proof. This follows readily from the fidelity's monotonicity and invariance under isometries. □

As a consequence we find a simple expression in terms of a Kraus representation.

Corollary 6.7. *Let* $\mathcal{T}_A[M_A] = \sum_i X_i M_A X_i^\dagger$ *be a Kraus representation. Then,*

$$F(\mathcal{T}_A, \rho_A) = \sqrt{\sum_i \left|\mathrm{Tr}[X_i \rho_A]\right|^2}.$$

Proof. Let $\sigma_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ be an arbitrary purification of ρ_A. Then,

$$\begin{aligned} F(\mathcal{T}_A, \rho_A)^2 &= F\big(\sigma_{AB}, (\mathcal{T}_A \otimes \mathcal{I}_B)[\sigma_{AB}]\big)^2 \\ &= \big\langle\Psi_{AB}\big|(\mathcal{T}_A \otimes \mathcal{I}_B)\big[|\Psi_{AB}\rangle\langle\Psi_{AB}|\big]\big|\Psi_{AB}\big\rangle \\ &= \sum_i \big\langle\Psi_{AB}\big|(X_i \otimes I_B)|\Psi_{AB}\rangle\langle\Psi_{AB}|(X_i^\dagger \otimes I_B)\big|\Psi_{AB}\big\rangle \\ &= \sum_i |\langle\Psi_{AB}|X_i \otimes I_B|\Psi_{AB}\rangle|^2 = \sum_i |\mathrm{Tr}[X_i \rho_A]|^2, \end{aligned}$$

where we first used Lemma 6.6, then Eq. (3.13) to evaluate the fidelity, and finally the Kraus representation of $\mathcal{T}_A$. □

6.5 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 6.8 (Schumacher compression). *Let* $\rho \in D(\mathcal{H}_A)$ *and* $\delta \in (0,1)$*. Then:*

1. *If* $R > H(\rho)$ *then there exists* n_0 *such that there exists an* (n, R, δ)*-quantum code for all* $n \geqslant n_0$*.*
2. *If* $R < H(\rho)$ *then there exists* n_0 *such that no* (n, R, δ)*-quantum codes exist for* $n \geqslant n_0$*.*

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 6.9 (Typical subspace and projector). *For* $\rho \in D(\mathcal{H}_A)$*,* $n \in \mathbb{N}$*, and* $\varepsilon > 0$*, define the* typical subspace

$$S_{n,\varepsilon}(\rho) = \mathrm{span}\,\{|e_{y_1}\rangle \otimes \cdots \otimes |e_{y_n}\rangle \;:\; y^n \in T_{n,\varepsilon}(q)\},$$

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

$$\begin{aligned}\rho^{\otimes n} &= \sum_{y^n} q(y_1)\cdots q(y_n)\, (|e_{y_1}\rangle \otimes \cdots \otimes |e_{y_n}\rangle)\,(\langle e_{y_1}| \otimes \cdots \otimes \langle e_{y_n}|) \\ &= \sum_{y^n} q(y_1)\cdots q(y_n)\, |e_{y_1}\rangle\langle e_{y_1}| \otimes \cdots \otimes |e_{y_n}\rangle\langle e_{y_n}|,\end{aligned} \tag{6.10}$$

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

$$\Pi_{n,\varepsilon} = \sum_{y^n \in T_{n,\varepsilon}(q)} |e_{y_1}\rangle\langle e_{y_1}| \otimes \cdots \otimes |e_{y_n}\rangle\langle e_{y_n}|. \tag{6.11}$$

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 6.10 (Quantum Asymptotic Equipartition Property, AEP)**.** *With notation as above, the following properties hold:*

0. *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* $2^{-n(H(\rho)\pm\varepsilon)}$*,*
1. $\operatorname{rank}\Pi_{n,\varepsilon} = \dim S_{n,\varepsilon}(\rho) = |T_{n,\varepsilon}(q)| \leqslant 2^{n(H(\rho)+\varepsilon)}$*,*
2. $\operatorname{Tr}[\Pi_{n,\varepsilon}\rho^{\otimes n}] \geqslant 1 - \frac{\sigma^2}{n\varepsilon}$*, where* σ^2 *is a constant that only depends on the eigenvalues of* ρ*.*

Proof. These properties follow from the corresponding properties in Lemma 5.9. For property 1, this is immediate. To prove the other properties, note that Eqs. (6.10) and (6.11) 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 0 from the corresponding property in Lemma 5.9. And since the preceding implies that

$$\operatorname{Tr}[\Pi_{n,\varepsilon}\rho^{\otimes n}] = \sum_{y^n \in T_{n,\varepsilon}(q)} q(y^n) = \Pr(Y^n \in T_{n,\varepsilon}(q)),$$

where $Y_1, \dots, Y_n \overset{\text{IID}}{\sim} p$, property 2 likewise follows from Lemma 5.9. □

We now prove Schumacher's theorem.

Proof of Theorem 6.8. To prove part 1, 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 1 of Lemma 6.10,

$$\operatorname{rank}\Pi_{n,\varepsilon} \leqslant 2^{n(H(\rho)+\varepsilon)} = 2^{n(R-\varepsilon)} \leqslant 2^{\lfloor nR \rfloor} = \dim\big((\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}\big); \tag{6.12}$$

the final inequality holds for large enough n (e.g., if $n \geqslant \frac{1}{\varepsilon}$). Eq. (6.12) implies that there exists a linear map $V\colon \mathcal{H}_A^{\otimes n} \to (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}$ 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 $D = \operatorname{rank} \Pi_{n,\varepsilon}$, $\{|\varphi_i\rangle\}_{i=1}^{D}$ 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

$$\begin{aligned}
\mathcal{E}[M] &:= VMV^\dagger + \operatorname{Tr}[\sqrt{I - V^\dagger V} M \sqrt{I - V^\dagger V}]\, \alpha, \\
\mathcal{D}[M] &:= V^\dagger M V + \operatorname{Tr}[\sqrt{I - VV^\dagger} M \sqrt{I - VV^\dagger}]\, \beta,
\end{aligned}$$

where α and β are arbitrary states. Note that $I - V^\dagger V$ and $I - VV^\dagger$ are PSD (since $V^\dagger V = \Pi_{n,\varepsilon}$ and $VV^\dagger$ 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 $\mathcal{E} \in C(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor})$ and $\mathcal{D} \in C((\mathbb{C}^2)^{\otimes \lfloor nR \rfloor}, \mathcal{H}_A^{\otimes n})$. It remains to verify Eq. (6.7) or, equivalently, Eq. (6.9). 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 Practice Problem 4.4, this means that $\mathcal{D} \circ \mathcal{E}$ has a Kraus representation starting with $V^\dagger V = \Pi_{n,\varepsilon}$. Hence, Corollary 6.7 implies that

$$F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geqslant \left|\operatorname{Tr}[V^\dagger V \rho^{\otimes n}]\right| = \operatorname{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}]$$

But now property 2 in Lemma 6.10 shows that the right-hand side is $\geqslant 1 - \delta$ if we choose n sufficiently large. This concludes the proof of part 1.

How about part 2? You already gave this a try in Practice Problem 6.4 and we will follow the argument sketched therein. Fix $\delta \in (0,1)$ and $R < H(\rho)$. First, note that if P is an arbitrary orthogonal projection of rank $\leqslant 2^{nR}$ then

$$\begin{aligned}
\operatorname{Tr}[P\rho^{\otimes n}] &= \operatorname{Tr}[P\Pi_{n,\varepsilon}\rho^{\otimes n}] + \operatorname{Tr}[P(I - \Pi_{n,\varepsilon})\rho^{\otimes n}] \\
&\leqslant \underbrace{\|P\|_1}_{\leqslant 2^{nR}} \underbrace{\|\Pi_{n,\varepsilon}\rho^{\otimes n}\|_\infty}_{\leqslant 2^{-n(H(\rho)-\varepsilon)}} + \underbrace{\operatorname{Tr}[(I - \Pi_{n,\varepsilon})\rho^{\otimes n}]}_{1 - \operatorname{Tr}[\Pi_{n,\varepsilon}\rho^{\otimes n}]} \\
&\leqslant 2^{-n\varepsilon} + \left(1 - \operatorname{Tr}[\Pi_{n,\varepsilon}\rho^{\otimes n}]\right)
\end{aligned} \tag{6.13}$$

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. (3.7) and the operator norm using property 0 in Lemma 6.10. For the right-hand side term, we simply used that $P \leqslant I$ and rewrote the result. In view of property 2 in Lemma 6.10, the expression Eq. (6.13) converges to 0 as $n \to \infty$.

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 $X_i \in L(\mathcal{H}_A^{\otimes n}, (\mathbb{C}^2)^{\otimes \lfloor nR \rfloor})$, it has necessarily rank $\leqslant 2^{nR}$. 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 $\leqslant 2^{nR}$. We now evaluate the channel fidelity using Corollary 6.7 and obtain

$$\begin{aligned}
F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n})^2 &= \sum_k \left|\operatorname{Tr}[Z_k \rho^{\otimes n}]\right|^2 = \sum_k \left|\operatorname{Tr}[P_k Z_k \rho^{\otimes n}]\right|^2 \\
&= \sum_k \left|\operatorname{Tr}[Z_k \sqrt{\rho^{\otimes n}} \sqrt{\rho^{\otimes n}} P_k]\right|^2 \leqslant \sum_k \operatorname{Tr}[Z_k^\dagger Z_k \rho^{\otimes n}] \operatorname{Tr}[P_k \rho^{\otimes n}],
\end{aligned}$$

where the inequality is by the Cauchy-Schwarz inequality for operators [Eq. (3.6)]. 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 4.4. This implies that

[1] For example, we can use $V = \sum_{y^n \in T_{n,\varepsilon}(q)} |E(y^n)\rangle \left(\langle e_{y_1}| \otimes \cdots \otimes \langle e_{y_n}|\right)$, where $E\colon T_{n,\varepsilon} \to \{0,1\}^{\lfloor nR \rfloor}$ 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 $E(y^n) \in \{0,1\}^{\otimes \lfloor nR \rfloor}$.

$r(k) := \mathrm{Tr}[Z_k^\dagger Z_k \rho^{\otimes n}]$ is a probability distribution. But then,

$$F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n})^2 \leqslant \sum_k r(k)\,\mathrm{Tr}[P_k \rho^{\otimes n}] \leqslant 2^{-n\varepsilon} + \big(1 - \mathrm{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}]\big)$$

by Eq. (6.13). By property 2 in Lemma 6.10, the right-hand side converges to 0 as $n \to \infty$. As a consequence, $F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geqslant 1 - \delta$ can only hold for finitely many n. In other words, (n, R, δ)-codes can only exist for finitely many values of n. □

Lecture 7

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, 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 Homework Problem 6.3, 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.

7.1 Entropies of subsystems

To study the entropies of subsystems, it is useful to first introduce some notation.

Definition 7.1 (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 Homework Problem 6.2, as well as for the Shannon entropy (Definition 5.3).

How are these entropies related? Let us first consider two very extreme cases:

- If ρ_{AB} is pure then $H(AB) = 0$ and

  $$H(A) = H(B). \tag{7.1}$$

  The latter is often called the *entanglement entropy* of ρ_{AB}.

  Proof. The former holds because the eigenvalues of a pure state are $1, 0, \ldots, 0$. The latter follows from the Schmidt decomposition, which implies that ρ_A and ρ_B have the same nonzero eigenvalues (Corollary 2.13). □

- 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,

$$\begin{aligned} H(AB) = \sum_{i,j} p_i q_j \log \frac{1}{p_i q_j} &= \sum_{i,j} p_i q_j \log \frac{1}{p_i} + \sum_{i,j} p_i q_j \log \frac{1}{q_j} \\ &= \sum_i p_i \log \frac{1}{p_i} + \sum_j q_j \log \frac{1}{q_j} = H(A) + H(B). \end{aligned}$$

□

Next, we list some general properties. We first discuss the extent to which the subadditivity and monotonicity properties of the Shannon entropy (see p. 53 and Homework Problem 5.3) generalize to the quantum case.

- *Subadditivity:*

  $$H(A) + H(B) \geqslant H(AB) \tag{7.2}$$

  Moreover, equality holds if *and only if* $\rho_{AB} = \rho_A \otimes \rho_B$.

  You proved this inequality on Homework Problem 6.3 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, $H(AB) \ngeqslant H(A)$ and $H(AB) \ngeqslant H(B)$. You discussed this in Homework Problem 6.2.

- However, for classical-quantum states ρ_{XB} we do have the monotonicity inequalities

  $$H(XB) \geqslant H(X) \quad \text{and} \quad H(XB) \geqslant H(B). \tag{7.3}$$

  You proved the first inequality in Homework Problem 6.4; the second will be on Practice Problem 8.4.

- *Araki-Lieb (or triangle) inequality:*

  $$H(AB) \geqslant \big|H(A) - H(B)\big|. \tag{7.4}$$

  We can think of Eq. (7.4) as a weaker form of monotonicity (not to be confused with Eq. (7.6) below). Indeed, if $H(AB) \geqslant H(A)$ and $H(AB) \geqslant H(B)$ were true then these would imply Eq. (7.4).

  Proof. Choose any purification ρ_{ABC} of ρ_{AB}. Then:

  $$H(AB) = H(C) \geqslant H(BC) - H(B) = H(A) - H(B),$$

  where the first and last step hold since ρ_{ABC} is pure [Eq. (7.1)] and the inequality is subadditivity [Eq. (7.2)]. Likewise,

  $$H(AB) = H(C) \geqslant H(AC) - H(A) = H(B) - H(A),$$

  which proves the other half of Eq. (7.4). □

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
$$H(AC) + H(BC) \geqslant H(ABC) + H(C). \tag{7.5}$$
Clearly, this inequality reduces to Eq. (7.2) if there is no C system, which justifies the terminology. Eq. (7.5) is much harder to prove than Eq. (7.2) and we will not have time to do this in the lecture (cf. the closely related monotonicity property of the quantum relative entropy that we will discuss in Lecture 8).

- *Weak monotonicity:* For all ρ_{ABC}, it holds that
$$H(AC) + H(BC) \geqslant H(A) + H(B). \tag{7.6}$$
This inequality follows from Eq. (7.5) by using a purification – in the same way that Eq. (7.4) follows from Eq. (7.2) – as you get to prove in Practice Problem 7.2. The name is justified since if $H(AC) \geqslant H(A)$ and $H(BC) \geqslant H(B)$ were true then these would imply Eq. (7.6).

7.2 Mutual information

In this section we will discuss the mutual information, which is a useful way to quantify correlations in quantum states.

Definition 7.2 (Mutual information)**.** *The* mutual information *of a quantum state* ρ_{AB} *is defined as*
$$I(A:B)_\rho := H(A)_\rho + H(B)_\rho - H(AB)_\rho. \tag{7.7}$$
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 7.3. *For a maximally entangled state*
$$\rho_{AB} = |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|, \quad |\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big)$$
the mutual information is $I(A:B) = 1 + 1 - 0 = 2$*. In contrast, for a classical maximally correlated state*
$$\rho_{AB} = \frac{1}{2}\big(|00\rangle\langle 00| + |11\rangle\langle 11|\big)$$
we have $I(A:B) = 1 + 1 - 1 = 1$*, since the overall state is not pure but mixed.*

We now list some useful properties, several of which follow directly from the results of Section 7.1:

- *Nonnegativity:* $I(A:B) \geqslant 0$. 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. (7.2)], including the condition for equality. □

- *Invariance under isometries:* For any state ρ_{AB} and isometries $V_{A\to A'}$, $W_{B\to B'}$, we have

  $$I(A:B)_\rho = I(A':B')_\sigma,$$

  where $\sigma_{A'B'} := (V_{A\to A'} \otimes W_{B\to B'})\rho_{AB}(V^\dagger_{A\to A'} \otimes W^\dagger_{B\to B'})$.

  Proof. This follows from the invariance of the von Neumann entropy under isometries (see p. 60) once we recognize that $\sigma_{A'} = V\rho_A V^\dagger$ and $\sigma_{B'} = W\rho_B W^\dagger$. □

- *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. □

- *Upper bound:* Let $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$. Then,

  $$I(A:B) \leqslant 2\min\{H(A), H(B)\} \leqslant 2\log\min\{d_A, d_B\}. \tag{7.8}$$

  For classical-quantum states ρ_{XB}, we have the stronger upper bound

  $$I(X:B) \leqslant \min\{H(X), H(B)\} \leqslant \log\min\{d_X, d_B\}. \tag{7.9}$$

  In particular, Eq. (7.9) holds for classical states and joint probability distributions. In Homework Problems 7.1 and 7.2 you will investigate under which conditions the upper bounds in Eqs. (7.8) and (7.9) hold with equality.

  Proof. The first inequality in Eq. (7.8) follows from the Araki-Lieb inequality [Eq. (7.4)]. Indeed, $H(A)+H(B)-H(AB) = I(A:B) \leqslant 2H(A)$ is equivalent to $H(AB) \geqslant H(B) - H(A)$, and similarly for the other bound. Likewise, the first bound in Eq. (7.9) is equivalent to the monotonicity inequalities in Eq. (7.3). □

- *Monotonicity:* For all ρ_{ACE},

  $$I(A:CE) \geqslant I(A:C). \tag{7.10}$$

  (We label the subsystems ACE rather than ABC to avoid confusion in the below.)

  Proof. This is simply a rewriting of strong subadditivity [Eq. (7.5)]. □

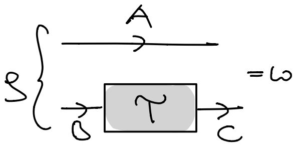
The latter property is equivalent to the following general result:

Lemma 7.4 (Data processing inequality). *Let $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a state, $\mathcal{T}_{B\to C} \in C(\mathcal{H}_B, \mathcal{H}_C)$ a channel, and $\omega_{AC} = (\mathcal{I}_A \otimes \mathcal{T}_{B\to C})[\rho_{AB}]$. Then,*

$$I(A:B)_\rho \geqslant I(A:C)_\omega,$$

By symmetry, the same 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 7.4 reduces to the monotonicity property of the mutual information (simply choose $B = CE$ and $\mathcal{T} = \mathrm{Tr}_E$).

Proof of Lemma 7.4. Any channel has a Stinespring representation $\mathcal{T}_{B\to C}[M_B] = \mathrm{Tr}_E\left[VM_BV^\dagger\right]$, where $V = V_{B\to CE} \in L(\mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_E)$ is an isometry [Lemma 4.4]. Note that

$$\omega_{ACE} = \left(I_A \otimes V_{B\to CE}\right)\rho_{AB}\left(I_A \otimes V_{B\to CE}\right)^\dagger$$

is an extension of ω_{AC} (i.e., $\mathrm{Tr}_E[\omega_{ACE}] = \omega_{AC}$). As a consequence,

$$I(A:B)_\rho = I(A:CE)_\omega \geqslant I(A:C)_\omega,$$

using that the mutual information is invariant under isometries and monotonic. □

Next week we will discuss a nice application of the data processing inequality known as Holevo's Theorem. This theorem 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.

Lecture 8

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 $\rho_{AB} = \mathrm{Tr}_C[\rho_{ABC}]$) and the inequalities among them, such as the strong subadditivity:

$$H(AB) + H(BC) \geqslant H(ABC) + H(B).$$

This is equivalent to the monotonicity of the mutual information:

$$I(A : B) \leqslant I(A : BC)$$

where $I(A : B) = H(A) + H(B) - H(AB)$. We also saw the data processing inequality for the mutual information:

$$I(A : B)_\rho \geqslant I(A : C)_\omega \tag{8.1}$$

where $\omega_{AC} = (\mathcal{I}_A \otimes \mathcal{T}_{B\to C})[\rho_{AB}]$ is obtained by applying a channel $\mathcal{T}_{B\to C}$ on the B system of ρ_{AB} (intuitively, local processing of individual subsystems can only decrease the mutual information between them).

8.1 Holevo bound

Assume we draw an element $x \in \Sigma$ with probability p_x, record its value in a classical register X with Hilbert space $\mathcal{H}_X = \mathbb{C}^\Sigma$, and create an arbitrary state $\rho_x \in D(\mathcal{H}_B)$ associated to x in a separate register B. Then the resulting cq-state

$$\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \in D(\mathcal{H}_X \otimes \mathcal{H}_B) \tag{8.2}$$

represents the ensemble $\{p_x, \rho_x\}$. To such an ensemble, or the corresponding cq-state, we associate the so-called Holevo χ-quantity.

Definition 8.1 (Holevo χ-quantity). *The* Holevo χ-quantity *of an ensemble* $\{p_x, \rho_x\}$ *is*

$$\chi(\{p_x, \rho_x\}) := I(X : B) = H\Big(\sum_x p_x \rho_x\Big) - \sum_x p_x H(\rho_x),$$

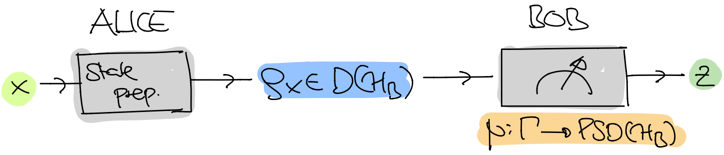
where the mutual information is computed in the cq-state $\rho_{XB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x$.

To verify the second equality, use that $H(XB) = H(p) + \sum_x p_x H(\rho_x)$, as you proved in Homework Problem 6.4 (a). Using part (b) of the same homework problem (or the nonnegativity of the mutual information) we find that χ is nonnegative:

$$0 \leqslant \chi(\{p_x, \rho_x\}) \leqslant H\Big(\sum_x p_x \rho_x\Big) \leqslant \log \dim \mathcal{H}_B. \tag{8.3}$$

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:


ALICE
x
State prep.
$\rho_x \in D(\mathcal{H}_B)$
BOB
z
$\mu: \Gamma \to PSD(\mathcal{H}_B)$

Here, Alice has a classical message $x \in \Sigma$ with distribution $p \in P(\Sigma)$ which she would like to communicate to Bob. For this, she sends Bob a quantum state $\rho_x \in D(\mathcal{H}_B)$, and Bob applies a measurement $\mu\colon \Gamma \to PSD(\mathcal{H}_B)$. 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

$$p(x, z) = p(x) \operatorname{Tr}[\rho_x \mu(z)].$$

In Homework Problem 2.1, 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 $\Pr(X = Z) \leqslant 2^{n-m}$, i.e., we need to send $n \geqslant m$ 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 8.2 (Holevo). $I(X:Z) \leqslant \chi(\{p_x, \rho_x\})$ *for any ensemble* $\{p_x, \rho_x\}$ *and measurement* μ.

Holevo's theorem is a simple consequence of the data processing inequality (Lemma 7.4), which in turn relies on the very nontrivial strong subadditivity inequality.

Proof. Let ρ_{XB} be the cq-state from Eq. (8.2) that represents the ensemble $\{p_x, \rho_x\}$, let $\mu\colon \Gamma \to PSD(\mathcal{H}_B)$ be an arbitrary measurement on system B, and let $\Phi_{B\to Z}[\sigma] := \sum_{z\in\Gamma} \operatorname{Tr}[\sigma\mu(z)]\, |z\rangle\langle z|$, with output space $\mathcal{H}_Z$, be the quantum channel corresponding to μ. Then by the data processing inequality for the mutual information, Eq. (8.1),

$$\chi(\{p_x, \rho_x\}) = I(X:B)_\rho \geqslant I(X:Z)_\omega$$

where

$$\omega_{XZ} = (\mathcal{I}_X \otimes \Phi_{B\to Z})[\rho_{XB}] = \sum_{x\in\Sigma} p_x |x\rangle\langle x| \otimes \Phi[\rho_x] = \sum_{x\in\Sigma, z\in\Gamma} p(x,z) |x\rangle\langle x| \otimes |z\rangle\langle z|$$

is the resulting output state after the measurement. That was easy! □

Recall from Eq. (8.3) that $\chi(\{p_x, \rho_x\}) \leqslant \log \dim \mathcal{H}_B$ where $\rho_x \in D(\mathcal{H}_B)$. Together with Holevo's bound this implies that $I(X:Z) \leqslant \log \dim \mathcal{H}_B$ where $\dim \mathcal{H}_B$ is the dimension of the quantum states in the ensemble. What this means intuitively is that Alice cannot transmit more than $\log \dim \mathcal{H}_B$ classical bits to Bob by sending a quantum state of dimension $\dim \mathcal{H}_B$. In other words, by sending n qubits she cannot reliably transmit more than n classical bits.

Remark 8.3. *The above considerations are closely related to one of the most fundamental problems in quantum information theory: Given access to a quantum channel $\mathcal{N}_{A\to B}$ (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 $\sigma_{B,x} := \mathcal{N}_{A\to B}[\rho_{A,x}]$. Let $\chi(\mathcal{N}_{A\to B})$ denote the supremum of $\chi(\{p_x, \sigma_{B,x}\})$ over all ensembles obtained in this way. Then the classical capacity of $\mathcal{N}_{A\to B}$ is given by $\lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}_{A\to B}^{\otimes n})$. Proving this result is out of scope for this introductory lecture, but you can consult the books by Watrous or Wilde for details.*

8.2 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 8.4 (Relative entropy)**.** *Let $p, q \in P(\Sigma)$ be probability distributions. The* relative entropy *of p with respect to q is*

$$D(p\|q) = \begin{cases} \sum_{x\in\Sigma} p(x)\log\frac{p(x)}{q(x)} & \text{if } \{x\colon q(x)=0\} \subseteq \{x\colon p(x)=0\}, \\ \infty & \text{otherwise.} \end{cases} \tag{8.4}$$

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. 51 that $\lim_{a\to 0} a\log a = 0$. 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.

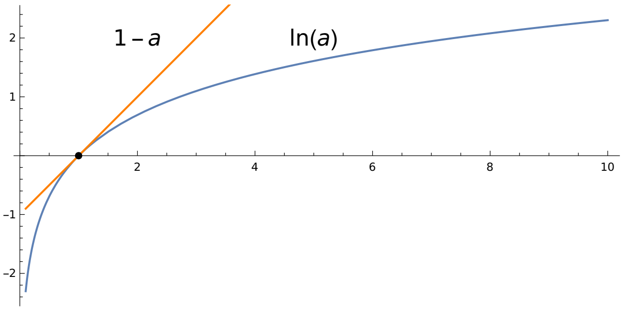
$p(x)$	$q(x)$	$p(x)\log\frac{p(x)}{q(x)}$
$= 0$	$= 0$	0
$= 0$	> 0	0
> 0	$= 0$	∞
> 0	> 0	finite

The condition for when $D(p\|q)$ in Eq. (8.4) 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:* $D(p\|q) \geqslant 0$, with equality iff $p = q$.

  Proof. Without loss of generality, we assume that $p(x) \geqslant 0$ for all x. Note that $\ln a \leqslant a - 1$, with equality iff $a = 1$.



Since $\log a = \frac{\ln a}{\ln 2}$,

$$\begin{aligned}
D(p\|q) &= \sum_x p(x)\left(-\log\frac{q(x)}{p(x)}\right) \\
&\geqslant \frac{1}{\ln 2}\sum_x p(x)\left(1-\frac{q(x)}{p(x)}\right) \\
&= \frac{1}{\ln 2}\left(\sum_x p(x) - \sum_x q(x)\right) = 0,
\end{aligned}$$

completing the proof. □

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 $p, q \in P(\Sigma)$ 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 $H(p) \leqslant \log|\Sigma|$, with equality if and only if p is uniform. We proved this already on p. 53 in Lecture 5.

As another application, let $p_{XY} \in P(\Sigma \times \Gamma)$ 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 $I(X:Y)_{p_{XY}} \geqslant 0$, 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).

8.3 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 8.5 (Quantum relative entropy). *Let* $\rho, \sigma \in D(\mathcal{H})$ *be quantum states. The* quantum relative entropy *of* ρ *with respect to* σ *is*

$$D(\rho\|\sigma) = \begin{cases} \mathrm{Tr}[\rho \log \rho] - \mathrm{Tr}[\rho \log \sigma] & \textit{if } \ker \sigma \subseteq \ker \rho, \\ \infty & \textit{otherwise}. \end{cases}$$

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 $\mathrm{im}\, \rho \subseteq \mathrm{im}\, \sigma$. For example, $D(|0\rangle\langle 0| \,\|\, |+\rangle\langle +|) = \infty$ since $\mathrm{span}\{|0\rangle\} \nsubseteq \mathrm{span}\{|+\rangle\}$.

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 $D(\rho\|\sigma) = D(p\|q)$.

- *Monotonicity:* For any $\rho, \sigma \in D(\mathcal{H})$ and any $\Phi \in C(\mathcal{H}, \mathcal{H}')$:

  $$D(\rho\|\sigma) \geqslant D(\Phi[\rho]\|\Phi[\sigma]). \tag{8.5}$$

  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):* $D(\rho\|\sigma) \geqslant 0$, with equality iff $\rho = \sigma$.

  Proof. Let $\mu \colon \Omega \to \mathrm{PSD}(\mathcal{H})$ be a quantum measurement and denote by $\Phi \in C(\mathcal{H}, \mathcal{X})$ where $\mathcal{X} = \mathbb{C}^\Omega$ the quantum channel

  $$\Phi[\omega] := \sum_{x \in \Omega} \mathrm{Tr}[\mu(x)\omega]\,|x\rangle\langle x|$$

  that implements the measurement μ. Denote by p and q the probability distributions resulting from measuring ρ and σ, respectively:

  $$p(x) := \mathrm{Tr}[\mu(x)\rho], \qquad q(x) := \mathrm{Tr}[\mu(x)\sigma].$$

  Note that

  $$\Phi[\rho] = \sum_{x \in \Omega} p(x)|x\rangle\langle x|, \qquad \Phi[\sigma] = \sum_{x \in \Omega} q(x)|x\rangle\langle x|$$

  are diagonal. By monotonicity,

  $$D(\rho\|\sigma) \geqslant D(\Phi[\rho]\|\Phi[\sigma]) = D(p\|q) \geqslant 0,$$

  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, you will show in Homework Problem 8.1 that, for any $\rho, \sigma \in D(\mathcal{H})$, 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, $D(\rho\|\sigma) \geqslant D(p\|q) > 0$ by a similar argument as above. Hence the quantum relative entropy is also a divergence. □

- *Joint convexity:* Let Σ be a finite set, $p \in P(\Sigma)$ a probability distribution, and $(\rho_x)_{x\in\Sigma}$ and $(\sigma_x)_{x\in\Sigma}$ families of states in $D(\mathcal{H})$.

  $$D(\sum_{x\in\Sigma} p_x\rho_x \| \sum_{x\in\Sigma} p_x\sigma_x) \leqslant \sum_{x\in\Sigma} p_x D(\rho_x\|\sigma_x).$$

  You will show this in Homework Problem 8.3.

- 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 $\rho, \sigma \in D(\mathbb{C}^d)$ where $\sigma = I/d$ is the maximally mixed state. You will show in Practice Problem 8.3 that

$$D(\rho\|\sigma) = -H(\rho) + \log d,$$

implying that $H(\rho) \leqslant \log d$, with equality iff $\rho = I/d$ is maximally mixed. We know this already from p. 60 in Lecture 6.

Let ρ_{AB} be a bipartite state with marginals $\rho_A = \operatorname{Tr}_B \rho_{AB}$ and $\rho_B = \operatorname{Tr}_A \rho_{AB}$. You will show in Practice Problem 8.3 that

$$D(\rho_{AB}\|\rho_A \otimes \rho_B) = I(A:B)_{\rho_{AB}},$$

implying $I(A:B)_{\rho_{AB}} \geqslant 0$, 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. (7.2) in Lecture 7.

Finally, we can also derive the monotonicity of the mutual information [Eq. (7.10)] from the monotonicity of the relative entropy [Eq. (8.5)]. Namely, by choosing $\Phi = \operatorname{Tr}_E$, we obtain

$$I(A:BE)_{\rho_{ABE}} = D(\rho_{ABE}\|\rho_A \otimes \rho_{BE}) \geqslant D(\rho_{AB}\|\rho_A \otimes \rho_B) = I(A:B)_{\rho_{AB}}.$$

As discussed last week, this inequality is in turn equivalent to strong subadditivity [Eq. (7.5)].

Lecture 9

Entanglement

Last week we discussed a scenario where Alice wants to transmit a classical message to Bob by sending a quantum state, and we proved Holevo's bound which establishes an upper bound on how much classical information can be extracted from an ensemble of quantum states. An important consequence of Holevo's bound is that one cannot reliably transmit more than n classical bits by sending n qubits.

Last time we also introduced the classical and quantum relative entropy, and saw how it can be used to quickly rederive entropic inequalities. The main idea was to evaluate the relative entropy on a particular pair of states, and then use nonegativity (Klein's inequality) or monotonicity of the relative entropy.

This week we will look at the opposite problem from last week, namely the problem of transmitting quantum information by sending classical bits. Wait, does this even make sense? This is clearly impossible for several reasons:

- Classical information is a special case of quantum information. We would not need quantum information (or quantum mechanics for that matter) if we could do the same things classically.

- There are more quantum states than bit strings. Indeed, the set of quantum states of a given dimension is continuous and infinite while the set of n-bit strings is discrete and finite.

If Alice wants to classically send an n-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 a list of all matrix entries of ρ. This would be an exponentially long list since ρ is of size $2^n \times 2^n$. Moreover, it would describe ρ only approximately since the matrix entries can be given only to a finite precision. And even if they went through all this trouble, the state reconstructed by Bob would not preserve the correlations Alice's state might have had with an external system. For example, if $\rho_A = \operatorname{Tr}_R[|\Psi_{AR}\rangle\langle\Psi_{AR}|]$ where R is some reference system that is not accessible to Bob, the state he reconstructs would not be correlated with R.

Nevertheless, one can still wonder if this can be achieved in some weaker sense. That is, can we somehow transmit a quantum state without ever actually sending any qubits (at least, not any qubits that depend on the state that we want to transmit)? In other words, would it help if we were to exchange some qubits *before* we get the actual state but cannot exchange any further qubits afterwards? Surprisingly, in such scenario it is possible to perfectly transmit a quantum state by sending only classical information, a procedure known as *quantum teleportation*.

9.1 Pauli matrices and Bell states

Mathematically, the quantum teleportation protocol has to do with a nice interplay between the Pauli matrices, Bell states, and the swap operation.

Recall from Eq. (1.10) of Lecture 1 that the *Pauli matrices* are as follows:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that $X^2 = Y^2 = Z^2 = I$. You can check that $ZX = -XZ = iY$, so the four Pauli matrices (up the annoying i in Y) can also be expressed as follows:

$$Z^0X^0 = I, \qquad Z^0X^1 = X, \qquad Z^1X^0 = Z, \qquad Z^1X^1 = iY.$$

They are related in a nice way to the two-qubit *swap operation*, which is defined as

$$W|a, b\rangle = |b, a\rangle,$$

for all $a, b \in \{0, 1\}$. In Practice Problem 9.4 you will show that

$$\begin{aligned} W &= \frac{1}{2}(I \otimes I + X \otimes X + Y \otimes Y + Z \otimes Z) \\ &= \frac{1}{2} \sum_{z,x \in \{0,1\}} Z^zX^x \otimes X^xZ^z. \end{aligned}$$

We denote the canonical two-qubit maximally entangled state between systems A and B by

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

This state is one of the four *Bell states*:

$$\begin{aligned} |\Phi^{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |\Phi^{01}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ |\Phi^{10}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), & |\Phi^{11}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \end{aligned} \tag{9.1}$$

These two-qubit states form an orthonormal basis of $\mathbb{C}^{\Sigma \times \Sigma}$ where $\Sigma = \{0, 1\}$.

You will show in Practice Problem 9.1 that the four Pauli matrices (again, up to the annoying i in Y) are related to the four Bell states as follows:

$$\begin{aligned} |\Phi^{zx}\rangle &= (Z^zX^x \otimes I)|\Phi^+\rangle \\ &= (I \otimes X^xZ^z)|\Phi^+\rangle, \end{aligned}$$

for all $z, x \in \{0, 1\}$. This *local conversion* property of Bell states is very surprising – if Alice and Bob each posses one qubit of a Bell state, any of them can apply a Pauli matrix on their respective qubit and convert their joint state to any of the other four Bell states.

We still need one more property of Bell states, namely that they can be prepared / unprepared from the corresponding standard basis state $|z, x\rangle$ as follows:

$$|\Phi^{zx}\rangle = \mathrm{CNOT}\,(H \otimes I)\,|z, x\rangle, \qquad |z, x\rangle = (H \otimes I)\,\mathrm{CNOT}\,|\Phi^{zx}\rangle, \tag{9.2}$$

where CNOT $:= |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$ is the controlled-NOT operation and H is the Hadamard operation:

$$\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Note that both operations are their own inverses: $\mathrm{CNOT}^\dagger = \mathrm{CNOT}$ and $H^\dagger = H$, hence it is enough to verify only one of the identities in Eq. (9.2) and the other follows automatically.

9.2 Teleportation

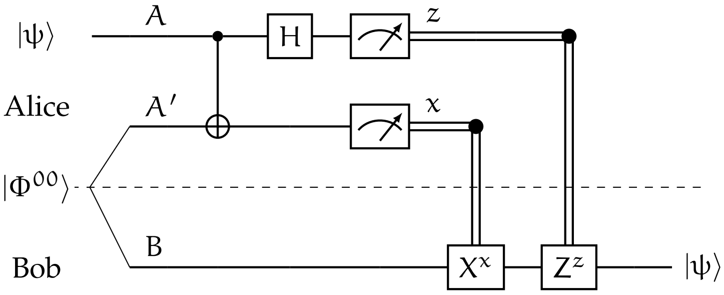
Let $|\psi\rangle \in \mathbb{C}^2$ be an arbitrary qubit state. The teleportation protocol can be derived from the following *teleportation identity*:

$$|\psi_A\rangle \otimes |\Phi^{00}_{A'B}\rangle = \frac{1}{2} \sum_{z,x \in \{0,1\}} |\Phi^{zx}_{AA'}\rangle \otimes X^x Z^z |\psi_B\rangle, \tag{9.3}$$

which you will prove in Homework Problem 9.1. The *teleportation protocol* is then as follows:

1. Start with state $|\psi_A\rangle \otimes |\Phi^{00}_{A'B}\rangle$ where registers AA' belong to Alice and B belongs to Bob.
2. Alice measures her registers AA' in the Bell basis $|\Phi^{zx}_{AA'}\rangle$ and sends the two measurement outcomes $z, x \in \{0, 1\}$ to Bob. Note from Eq. (9.2) that measuring in the Bell basis is equivalent to applying $(H \otimes I)\,\mathrm{CNOT}$ and then measuring in the standard basis.
3. Bob applies the *Pauli correction* $Z^z X^x$ on his qubit B to recover Alice's state $|\psi\rangle$.

Here is a graphical[1] depiction of the teleportation protocol:



Before we verify that this protocol indeed works, we need a way to compute the post-measurement state on Bob's system. Recall Axiom 2.6 that describes the Born's rule for measuring a subsystem: if you measure the A system of a state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ with measurement $\mu_A : \Omega \to \mathrm{PSD}(\mathcal{H}_A)$, you get the outcome $\omega \in \Omega$ with probability

$$p_\omega = \mathrm{Tr}\big[\rho_{AB}(\mu_A(\omega) \otimes I_B)\big].$$

The measurement channel corresponding to μ_A has classical output space $\mathcal{H}_X = \mathbb{C}^\Omega$ and acts as

$$\Phi_{A\to X}[\sigma_A] = \sum_{\omega \in \Omega} \mathrm{Tr}\big[\sigma_A \mu_A(\omega)\big]\, |\omega\rangle\langle\omega|_X,$$

[1] Confusingly, in quantum circuits time goes from left to right, so the order of gates is reversed compared to symbolic expressions.

for all $\sigma_A \in D(\mathcal{H}_A)$. You derived in Practice Problem 5.2 that the joint post-measurement state on XB is classical-quantum:

$$\rho_{XB} = (\Phi_{A\to X} \otimes \mathcal{I}_B)[\rho_{AB}] = \sum_{\omega\in\Omega} |\omega\rangle\langle\omega|_X \otimes \operatorname{Tr}_A\left[\rho_{AB}(\mu_A(\omega) \otimes I_B)\right].$$

By appropriately normalizing each term, we can think of this as an ensemble $\{p_\omega, \rho_{B,\omega}\}$. Let us formalize this observation as an axiom.

Axiom 9.1 (Post-measurement state)**.** *If a quantum system* AB *is in state* ρ_{AB} *and we perform a measurement* $\mu_A \colon \Omega \to \mathrm{PSD}(\mathcal{H}_A)$ *on subsystem* A*, then we obtain outcome* $\omega \in \Omega$ *with probability given by Born's rule [Eq.* (2.6)*], i.e.,*

$$p_\omega = \operatorname{Tr}\left[\rho_{AB}(\mu_A(\omega) \otimes I_B)\right].$$

In case the outcome is ω *then the state of system B after the measurement is given by*

$$\rho_{B,\omega} = \frac{1}{p_\omega} \operatorname{Tr}_A\left[\rho_{AB}(\mu_A(\omega) \otimes I_B)\right].$$

This state is called the post-measurement state *corresponding to outcome* ω*. The ensemble* $\{p_\omega, \rho_{B,\omega}\}$ *of post-measurement states is described by the classical-quantum state* $\rho_{XB} = \sum_{\omega\in\Omega} p_\omega |\omega\rangle\langle\omega|_X \otimes \rho_{B,\omega}$.

To verify the correctness of the teleportation protocol, note from Eq. (9.3) that the original input state $|\psi_A\rangle \otimes |\Phi^{00}_{A'B}\rangle$ is equivalent to a linear combination of four terms. In each term, Alice's qubits are in one of the four Bell states $|\Phi^{zx}_{AA'}\rangle$ while Bob's qubit in the corresponding term is in the state $X^x Z^z |\psi_B\rangle$. If Alice measures her two qubits in the Bell basis, she gets each of the four possible outcomes with probability $1/4$. If she sends the measurement outcomes z and x to Bob, he can recover the original state $|\psi\rangle$ by applying $Z^z X^x$ to cancel out the undesirable Pauli factor in $X^x Z^z |\psi_B\rangle$ (recall that $X^2 = Z^2 = I$).

You will show in Homework Problem 9.1 that the teleportation protocol not only transmits Alice's state to Bob but also preserves any correlations it might have had with some external reference system. If you were to trace out this reference system, the remaining state on Alice's system A' would be mixed. Hence you can think of this as an argument that proves the correctness of the teleportation protocol also for mixed states.

There is another interesting protocol known as *superdense coding* which is dual to teleportation since it achieves the opposite conversion: it lets you transmit two classical bits by sending only one qubit (you will derive this protocol in Homework Problem 9.2). At first this might seem to violate Holevo's bound from the previous lecture, since each qubit can transmit at most one classical bit. However, the catch is that the protocol also consumes one shared copy of the two-qubit state $|\Phi^{00}\rangle$. This state is often called an *EPR pair*, for Einstein, Podolsky, and Rosen who wrote a famous paper about it, or an *ebit*, for "entangled bit". Teleportation and superdense coding can thus be summarized as the following two *resource inequalities*:

$$\begin{aligned} \text{teleportation:} \qquad & \text{ebit} + 2[c \to c] \geqslant [q \to q], \\ \text{superdense coding:} \qquad & \text{ebit} + [q \to q] \geqslant 2[c \to c], \end{aligned}$$

where $[c \to c]$ denotes one bit of classical communication and $[q \to q]$ 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".

Remark 9.2. *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 shared uniformly random bit.*

9.3 Entangled vs separable states

Teleportation and superdense coding raise many questions. For example:

- *What is so special about $|\Phi^+_{AB}\rangle$, the Bell states, or any other maximally entangled state?* Recall that a general maximally entangled state is of the form $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\alpha_i\rangle \otimes |\beta_i\rangle$ where $\{|\alpha_i\rangle\}$ and $\{|\beta_i\rangle\}$ are some orthonormal bases.

- *What is entanglement anyway?* We will see a formal definition in this class! One can also study more complicated forms of entanglement, such as between more than two systems, however in this course we focus only on the bipartite case which already is complicated enough, especially for mixed states.

- *How can we detect if a given state is entangled?* We will see some results today. In particular, for mixed states it is generally very hard, but for pure states it is quite easy.

- *How much entanglement is there in a given state $|\Psi_{AB}\rangle$?* You will see an example today in class, and another one in Practice Problem 9.3. One can also give a more operational answer to this question by asking how many qubits can be teleported by using $|\Psi_{AB}\rangle$ as a resource, or how many shared ebits are needed to construct $|\Psi_{AB}\rangle$ without using further quantum communication? We will discuss this more in subsequent lectures.

- *How can we manipulate entanglement?* Teleportation and superdense coding shows that we should treat entanglement as a resource, so the set of allowed operations for manipulating it should be such that they cannot create more entanglement out of thin air (in particular, quantum communication is not allowed). We will see in the next lecture that *Local Operations and Classical Communication* (LOCC) is the right set of allowed operations.

Instead of defining what entanglement is, let us define what it is *not* – the technical term for "not entangled" is *separable*.

Remark 9.3. *For the notion of entanglement to make sense in the first place, you need a system consisting of at least two (complementary) subsystems, say A and B. For example, it does not make sense to talk about the entanglement of a single-qubit state.*

Definition 9.4 (Separability). *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces. Then*

- *an operator $M_{AB} \in \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is* separable *if*

$$M_{AB} = \sum_i P_{A,i} \otimes Q_{B,i} \tag{9.4}$$

for some $P_{A,i} \in \mathrm{PSD}(\mathcal{H}_A)$ and $Q_{B,i} \in \mathrm{PSD}(\mathcal{H}_B)$;

- *a state $\rho_{AB} \in \mathrm{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is* separable *if*

$$\rho_{AB} = \sum_i p_i \rho_{A,i} \otimes \rho_{B,i} \tag{9.5}$$

for some probability distribution p and states $\rho_{A,i} \in \mathrm{D}(\mathcal{H}_A)$ and $\rho_{B,i} \in \mathrm{D}(\mathcal{H}_B)$;

- *a pure state* $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}| \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is* separable *if and only if it is a product state, that is, if and only if*

$$|\Psi_{AB}\rangle\langle\Psi_{AB}| = |\alpha_A\rangle\langle\alpha_A| \otimes |\beta_B\rangle\langle\beta_B| \tag{9.6}$$

or, equivalently,

$$|\Psi_{AB}\rangle = |\alpha_A\rangle \otimes |\beta_B\rangle \tag{9.7}$$

for unit vectors $|\alpha_A\rangle \in \mathcal{H}_A$ *and* $|\beta_B\rangle \in \mathcal{H}_B$.

We denote the set of separable operators on $\mathcal{H}_A \otimes \mathcal{H}_B$ *by* $\mathrm{Sep}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *and the set of separable states by* $\mathrm{SepD}(\mathcal{H}_A \otimes \mathcal{H}_B)$. *A state is called* entangled *if it is not separable.*

You will show in Practice Problem 9.1 that subsequent parts of the above definition are obtained by simply restricting the general notion of a separable operator first to mixed and then to pure states. In particular, if a pure state can be written as in Eq. (9.4) or Eq. (9.5), then it is actually a product, as in Eq. (9.6)!

Note that there is no sum in Eq. (9.7) – we simply demand that $|\Psi_{AB}\rangle = |\alpha_A\rangle \otimes |\beta_B\rangle$. Indeed, if we had a sum in this condition, it would not be restrictive at all since any pure state can be expressed as $|\Psi_{AB}\rangle = \sum_i c_i |\alpha_{A,i}\rangle \otimes |\beta_{B,i}\rangle$ for some $c_i \in \mathbb{C}$ (or even $c_i \geqslant 0$) and pure states $|\alpha_{A,i}\rangle \in \mathcal{H}_A$ and $|\beta_{B,i}\rangle \in \mathcal{H}_B$ thanks to the Schmidt decomposition.

As an example, note that any classical state ρ_{XY} is separable since

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

This justifies the idea that entanglement captures the non-classical part of correlations.

Remark 9.5. *In the classical case, the distinction between a product and a correlated distribution is similar to the distinction between a product and an entangled pure state. A probability distribution* $p_{XY} \in P(\Sigma \times \Gamma)$ *is* product *if* $p_{XY}(x,y) = p_X(x)p_Y(y)$ *for all* $x \in \Sigma$ *and* $y \in \Gamma$, *and* correlated *otherwise. Entanglement captures an even stronger notion of correlations since classical states are not entangled even if they are classically correlated.*

Remark 9.6. *Whether a given state is entangled or not is not affected by how we choose the local basis within each subsystem. For example,* $(U_A \otimes U_B)|\Phi^+_{AB}\rangle$ *is (maximally) entangled for any choice of the local unitaries* U_A *and* U_B. *However, a global basis change can map entangled states to separable states and vice versa. For example, recall from Eq. (9.2) that we can map any Bell state* $|\Phi^{zx}\rangle$ *to the corresponding standard basis state* $|z,x\rangle$ *by the operation* $U_{AB} = (H \otimes I)\,\mathrm{CNOT}$. *This is a global operation thanks to the* CNOT *gate that acts on both qubits. While the Bell states are maximally entangled, the standard basis states* $|z\rangle \otimes |x\rangle$ *are product.*

Now that we have introduced the notion of entanglement, one may ask how to determine whether a given state is entangled or not, and how to measure the amount of entanglement? Recall from Eq. (7.1) that $H(A) = H(B)$ whenever ρ_{AB} is a pure state. This quantity is a useful measure of entanglement.

Definition 9.7 (Entanglement entropy)**.** *If* $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *then* $H(A) = H(B)$. *This quantity is known as* entanglement entropy *of* $|\psi_{AB}\rangle$.

In Practice Problem 9.1, you will show that a pure state $|\psi_{AB}\rangle$ is separable if and only if its entanglement entropy is zero. This is also equivalent to $|\psi_{AB}\rangle$ having Schmidt rank one (recall from Lemma 2.12 that Schmidt rank is the number of non-zero coefficients in the Schmidt decomposition).

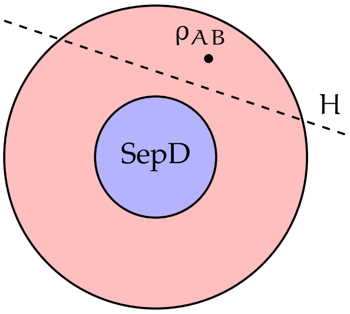
The set of separable states is clearly *convex* – it is defined as the convex hull of product states. In fact, it is also the convex hull of *pure* product states, since without loss of generality we can take the states $\rho_{A,i}$ and $\rho_{B,i}$ in Eq. (9.5) to be pure. Indeed, if $\rho_{A,i} = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ for some pure states $|\psi_j\rangle$, we can simply substitute this in Eq. (9.5) and expand it further by appropriately increasing the range of the summation.

The set of separable states is also *compact* (i.e., closed and bounded). This can be shown by noting that the set of pure states (i.e., the unit sphere) is closed and bounded, and that compactness is preserved under tensor products and convex hulls.

Since the set of separable states is convex and compact, we can use the hyperplane separation theorem. That is, for any entangled state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$, we can find a Hermitian operator $H \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that

1. $\langle H, \rho_{AB} \rangle_{HS} < 0$ and
2. $\langle H, \sigma_{AB} \rangle_{HS} \geqslant 0$, for any separable state $\sigma_{AB} \in \mathrm{SepD}(\mathcal{H}_A \otimes \mathcal{H}_B)$.

where $\langle M, N \rangle_{HS} := \mathrm{Tr}[M^\dagger N]$ is the Hilbert-Schmidt inner product, see Eq. (3.3). Such H is called an *entanglement witness* for ρ_{AB}. Here is a graphical depiction of the situation:



It is not immediately clear how to find an entanglement witness H for a given entangled state ρ_{AB}. Specifically, how to check that $\langle H, \sigma_{AB} \rangle_{HS} \geqslant 0$ for all $\sigma_{AB} \in \mathrm{SepD}(\mathcal{H}_A \otimes \mathcal{H}_B)$? Indeed, if one could easily check this by somehow iterating through all separable states σ_{AB}, there would be no need to look for an entanglement witness in the first place – one could instead just iterate through all $\sigma_{AB} \in \mathrm{SepD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and check whether $\sigma_{AB} = \rho_{AB}$. The following theorem offers an alternative way of checking separability. While it suffers from the same issue, we will later specialize it to a weaker (one-way) test that can be executed more easily.

Theorem 9.8 (Horodecki)**.** *A state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ is separable iff $(\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}] \geqslant 0$ for all unital positive superoperators $\Psi_{B\to A} : L(\mathcal{H}_B) \to L(\mathcal{H}_A)$.*

Recall that a superoperator $\Psi_{B\to A}$ is *unital* if $\Psi[I_B] = I_A$ and *positive* if $\Psi[P] \geqslant 0$ for all $P \geqslant 0$. Note that the roles of systems A and B in Theorem 9.8 can be exchanged due to symmetry.

Proof. The forward implication is straightforward: if ρ_{AB} is separable then

$$\begin{aligned}(\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}] &= (\mathcal{I}_A \otimes \Psi_{B\to A})\left[\sum_i p_i \rho_{A,i} \otimes \rho_{B,i}\right] \\ &= \sum_i p_i \rho_{A,i} \otimes \Psi_{B\to A}[\rho_{B,i}] \geqslant 0\end{aligned}$$

since Ψ is positive.

We will prove the converse by showing the contrapositive. That is, for any entangled state ρ_{AB} there exists a unital positive $\Psi_{B\to A}$ such that $(\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}] \not\geqslant 0$ (i.e., it has a negative eigenvalue). We will do this in two steps. First, let $\Phi^+_{AA} = |\Phi^+_{AA}\rangle\langle\Phi^+_{AA}|$ be the canonical maximally entangled state on $\mathcal{H}_A \otimes \mathcal{H}_A$ and note that

$$\begin{aligned}\langle\Phi^+_{AA}|\,(\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}]\,|\Phi^+_{AA}\rangle &= \langle\Phi^+_{AA}, (\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}]\rangle_{\mathrm{HS}} \\ &= \langle(\mathcal{I}_A \otimes \Psi^\dagger_{A\to B})[\Phi^+_{AA}], \rho_{AB}\rangle_{\mathrm{HS}} \\ &= \dim(\mathcal{H}_A)\,\langle J^{\Psi^\dagger}_{AB}, \rho_{AB}\rangle_{\mathrm{HS}}\end{aligned} \tag{9.8}$$

where $J^{\Psi^\dagger}_{AB}$ is the Choi operator of $\Psi^\dagger$, see Eq. (4.2). Next, note that for any positive operators $P \in L(\mathcal{H}_A)$ and $Q \in L(\mathcal{H}_B)$,

$$\begin{aligned}\langle P, \Psi(Q)\rangle_{\mathrm{HS}} &= \operatorname{Tr}\big[P^\dagger\Psi(Q)\big] \\ &= \dim(\mathcal{H}_A)\,\langle\Phi^+|\big(\overline{P}\otimes\Psi(Q)\big)|\Phi^+\rangle \\ &= \dim(\mathcal{H}_A)\,\langle\Phi^+, (\mathcal{I}\otimes\Psi)[\overline{P}\otimes Q]\rangle_{\mathrm{HS}} \\ &= \dim(\mathcal{H}_A)\,\langle(\mathcal{I}\otimes\Psi^\dagger)[\Phi^+], \overline{P}\otimes Q\rangle_{\mathrm{HS}} \\ &= \langle J^{\Psi^\dagger}_{AB}, \overline{P}\otimes Q\rangle_{\mathrm{HS}}.\end{aligned} \tag{9.9}$$

Since we assumed ρ_{AB} to be entangled, let H be its entanglement witness, i.e., H is a Hermitian matrix such that (i) $\langle H, \rho_{AB}\rangle_{\mathrm{HS}} < 0$ and (ii) $\langle H, \sigma_{AB}\rangle_{\mathrm{HS}} \geqslant 0$ for all $\sigma_{AB} \in \mathrm{SepD}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We can use this H to define the superoperator Ψ by setting $J^{\Psi^\dagger} = H$. Note that $(\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}]$ is not positive semidefinite since $\langle\Phi^+_{AA}|\,(\mathcal{I}_A \otimes \Psi_{B\to A})[\rho_{AB}]\,|\Phi^+_{AA}\rangle = \dim(\mathcal{H}_A)\,\langle J^{\Psi^\dagger}_{AB}, \rho_{AB}\rangle_{\mathrm{HS}} < 0$ thanks to Eq. (9.8) and property (i). Moreover, Ψ is positive since $\langle P, \Psi(Q)\rangle = \langle J^{\Psi^\dagger}, \overline{P}\otimes Q\rangle_{\mathrm{HS}} \geqslant 0$ thanks to Eq. (9.9) and property (ii).

What remains to show is that Ψ can be made unital without ruining the other two properties. The idea is to slightly perturb Ψ so that $\Psi(I)$ is full rank. Then we can define a new superoperator $\tilde{\Psi} : M \mapsto \Psi(I)^{-1/2}\Psi(M)\Psi(I)^{-1/2}$ which is clearly unital. Moreover $\tilde{\Psi}$ is also positive and, when applied to the B system of ρ_{AB}, produces an operator that is not positive semidefinite. □

Although the separability test provided by the above theorem is "if and only if", i.e., it can conclusively certify both entanglement as well as separability, it is hard to apply in practice. For separable states one has to iterate through all Ψ, which is not feasible, while for entangled states it is not clear what Ψ to choose. Luckily, in the second case one particular map – the transpose map – often does the job and the resulting test, known as the *partial transpose test*, can certify the entanglement of many states. However, this test is one-sided – it can only certify entanglement but not separability. In other words, we cannot claim a state to be separable if it fails the partial transpose test. It could simply be that this particular choice of Ψ was not suitable for detecting the entanglement in this particular state and that we should try other maps.

Corollary 9.9 (Partial transpose test)**.** *Let $\mathcal{T}[X] = X^{\mathsf{T}}$ denote the transpose map. If $(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_{AB}]$ has a negative eigenvalue then the state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ is entangled.*

We refer to $\mathcal{T}_A \otimes \mathcal{I}_B$ as the *partial transposition* operation and call $(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_{AB}]$ the *partial transpose* of ρ_{AB}. Note that due to symmetry we could equally well apply the transpose operation on the other system in the partial transpose test. This would not affect the conclusion since

$$(\mathcal{I}_A \otimes \mathcal{T}_B)[\rho_{AB}] = \big((\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_{AB}]\big)^{\mathsf{T}},$$

so in both cases the resulting operator has the same spectrum.

As an example, let us use the partial transpose test to show that our favorite two-qubit state $|\Phi^+_{AB}\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ is indeed entangled. Recall from Remark 2.9 that

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \quad \Phi^+_{AB} = |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}| = \frac{1}{2}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}\begin{pmatrix}1 & 0 & 0 & 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 0 & 0 & 1\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\1 & 0 & 0 & 1\end{pmatrix}.$$

It may not be immediately obvious how to transpose, say, the system A. However, we already did this in the example on Page 42. First, note that

$$\begin{aligned}\Phi^+_{AB} &= \frac{1}{2}\big(|00\rangle + |11\rangle\big)\big(\langle 00| + \langle 11|\big)\\ &= \frac{1}{2}\big(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\big)\\ &= \frac{1}{2}\big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|\big)\end{aligned}$$

By linearity, we can now apply $\mathcal{T}$ on the first register of each term and note that $(|0\rangle\langle 1|)^{\mathsf{T}} = |1\rangle\langle 0|$:

$$\begin{aligned}(\mathcal{T}_A \otimes \mathcal{I}_B)[\Phi^+_{AB}] &= \frac{1}{2}\big(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 0| \otimes |0\rangle\langle 1| + |0\rangle\langle 1| \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|\big)\\ &= \frac{1}{2}\begin{pmatrix}1 & 0 & 0 & 0\\0 & 0 & 1 & 0\\0 & 1 & 0 & 0\\0 & 0 & 0 & 1\end{pmatrix}.\end{aligned}$$

This matrix has a negative eigenvalue because of the central 2 × 2 block that looks like the Pauli X matrix. More concretely, you can verify that

$$\frac{1}{2}\begin{pmatrix}1 & 0 & 0 & 0\\0 & 0 & 1 & 0\\0 & 1 & 0 & 0\\0 & 0 & 0 & 1\end{pmatrix}\begin{pmatrix}0\\1\\-1\\0\end{pmatrix} = -\frac{1}{2}\begin{pmatrix}0\\1\\-1\\0\end{pmatrix},$$

so the partial transpose of Φ^+_{AB} has a negative eigenvalue: $-1/2$.

You will have an opportunity to apply the partial transpose test in Homework Problem 9.3. It is useful to know (particularly in Homework Problem 9.4) that, for small systems, the partial transpose test can be shown to work both ways, i.e., it can also conclusively detect separability.

Remark 9.10 (Converse for 2 × 2 and 2 × 3 systems). *If* $\dim(\mathcal{H}_A) = 2$ *and* $\dim(\mathcal{H}_B) \leqslant 3$ *then* $(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_{AB}] \geqslant 0$ *implies that* $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is separable.*

Lecture 10

Separable maps and LOCC

Last week we started discussing entanglement – a notion that applies to quantum systems with at least two subsystems, A and B. Recall from Definition 9.4 that a quantum state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ is entangled if it is *not* separable, and we call ρ_{AB} separable if

$$\rho_{AB} = \sum_i p_i \rho_{A,i} \otimes \rho_{B,i}$$

for some probability distribution p and states $\rho_{A,i} \in D(\mathcal{H}_A)$ and $\rho_{B,i} \in D(\mathcal{H}_B)$.

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:

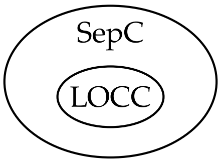
- *local operations*, such as unitary operations, isometries, measurement or, more generally, a quantum channel applied to one of the subsystems (e.g., $\Phi_{A \to A'}$ or $\Psi_{B \to B'}$);
- *classical communication* (exchanging classical messages between the two subsystems) can increase classical 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. We can then alternatively think of entanglement as the resource that cannot be 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. For example, 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.

10.1 Separable superoperators

While LOCC plays a central role in quantum information theory, unfortunately it is very hard to deal with mathematically. Therefore we often relax it to a slightly larger class of operations known as separable operations or SepC:

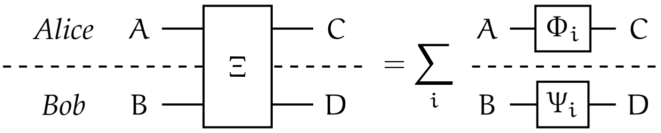


Let us define this set more formally. Recall from Definition 3.8 that we denote the set of all completely positive maps from $\mathcal{H}_A$ to $\mathcal{H}_B$ by $CP(\mathcal{H}_A, \mathcal{H}_B) \subset L(L(\mathcal{H}_A), L(\mathcal{H}_B))$.

Definition 10.1 (Separable channel). *A completely positive map* $\Xi \in \mathrm{CP}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$ *is* separable *if*

$$\Xi = \sum_i \Phi_i \otimes \Psi_i,$$

for some $\Phi_i \in \mathrm{CP}(\mathcal{H}_A, \mathcal{H}_C)$ *and* $\Psi_i \in \mathrm{CP}(\mathcal{H}_B, \mathcal{H}_D)$ *where* Φ_i *acts on Alice's side and* Ψ_i *on Bob's:*



A trace-preserving separable map is called a separable quantum channel. *We denote the sets of separable maps and separable quantum channels by* $\mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$ *and* $\mathrm{SepC}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$*, respectively.*

By definition, the sets of separable maps and separable quantum channels are related as follows:

$$\mathrm{SepC}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D) = \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D) \cap \mathrm{C}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$$

where C denotes the set of quantum channels. Just like for separable states, the colon ":" in the notation signifies how the systems are split between the two parties.

You will show in Practice Problem 10.2 that the composition of separable maps is also separable, and that a completely positive map Ξ is separable if and only if its Kraus representation is of the form

$$\Xi(X) = \sum_{x \in \Sigma} (A_x \otimes B_x) X (A_x \otimes B_x)^\dagger, \tag{10.1}$$

for some $A_x \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_C)$ and $B_x \in \mathrm{L}(\mathcal{H}_B, \mathcal{H}_D)$.

Thanks to the Choi-Jamiołkowski isomorphism (see Lemma 4.1) we can relate the separability of superoperators to the separability of operators, a notion we are already familiar with from the previous lecture (see Definition 9.4).

Before we do this, it is useful to formally introduce the following correspondence between matrices and bipartite pure states, the reverse of which we already encountered in Eq. (4.9). Intuitively, this is equivalent to cutting a matrix M into column vectors and then stacking them on top of each other to obtain one long column vector.

Definition 10.2 (Vectorization). *Let* $M \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$ *where* $\mathcal{H}_A = \mathbb{C}^\Sigma$ *and* $\mathcal{H}_B = \mathbb{C}^\Gamma$*. The* vectorization *of* M *is given by*

$$|M_{AB}\rangle := \sum_{\substack{a \in \Sigma \\ b \in \Gamma}} \langle b|M|a\rangle \, |a, b\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

In particular, if $M = |b\rangle\langle a|$*, for some* $a \in \Sigma$ *and* $b \in \Gamma$*, then* $|M_{AB}\rangle = |a, b\rangle$*.*

In Practice Problem 10.1, you will prove the following extremely useful *vectorization identity*:

$$(A \otimes B)|M\rangle = |BMA^\mathsf{T}\rangle, \tag{10.2}$$

for all $A \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_C)$, $B \in \mathrm{L}(\mathcal{H}_B, \mathcal{H}_D)$, $M \in \mathrm{L}(\mathcal{H}_A, \mathcal{H}_B)$.

Lemma 10.3. *Let* $\Xi \in \mathrm{CP}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$. *Then*

$$\Xi \in \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D) \qquad \Longleftrightarrow \qquad VJ^{\Xi}_{AB,CD}V^\dagger \in \mathrm{Sep}(\mathcal{H}_A \otimes \mathcal{H}_C : \mathcal{H}_B \otimes \mathcal{H}_D)$$

where $V|a,b,c,d\rangle = |a,c,b,d\rangle$, *for all* $|a\rangle \in \mathcal{H}_A$, $|b\rangle \in \mathcal{H}_B$, $|c\rangle \in \mathcal{H}_C$, $|d\rangle \in \mathcal{H}_D$.

Proof. Recall from Eq. (10.1) that

$$\Xi(X) = \sum_{x \in \Sigma} (A_x \otimes B_x) X (A_x \otimes B_x)^\dagger.$$

For simplicity, let us assume that the sum contains only one term and $A_x = |c\rangle\langle a|$ and $B_x = |d\rangle\langle b|$, for some standard basis states $|a\rangle$, $|b\rangle$, $|c\rangle$, $|d\rangle$. Then according to Eq. (4.1) or Eq. (4.2),

$$J^{\Xi}_{AB,CD} = |a,b,c,d\rangle\langle a,b,c,d|$$

and hence

$$VJ^{\Xi}_{AB,CD}V^\dagger = |a,c,b,d\rangle\langle a,c,b,d| = |a,c\rangle\langle a,c| \otimes |b,d\rangle\langle b,d|.$$

This is clearly a product operator and hence separable. More generally, you can show that

$$VJ^{\Xi}_{AB,CD}V^\dagger = \sum_x |A_x\rangle\langle A_x| \otimes |B_x\rangle\langle B_x| \tag{10.3}$$

where $|A_x\rangle = \sum_{a,c} \langle c|A_x|a\rangle\, |a,c\rangle$ and $|B_x\rangle = \sum_{b,d} \langle d|B_x|b\rangle\, |b,d\rangle$ are the vectorizations of the Kraus operators A_x and B_x. The operator in Eq. (10.3) is clearly separable across $AB : CD$. The reverse implication follows by running the same argument backwards. □

10.2 Entanglement rank

In the previous lecture we encountered two types of states – those that are entangled and those that are not (i.e., separable states). 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 Lemma 2.12 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 product states have Schmidt rank 1 while a maximally entangled state of dimension d has Schmidt rank d. We can extend this notion to a general PSD operator by decomposing it in terms of pure states with as small Schmidt rank as possible.

Definition 10.4 (Entanglement rank). *We write* $P_{AB} \in \mathrm{Ent}_r(\mathcal{H}_A : \mathcal{H}_B) \subseteq \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *if*

$$P_{AB} = \sum_x |\Psi_{AB,x}\rangle\langle\Psi_{AB,x}|,$$

where each $|\Psi_{AB,x}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *has Schmidt rank at most* r. *The* entanglement rank *of* $P_{AB} \in \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is the smallest* r *such that* $P_{AB} \in \mathrm{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$.

For pure states, the entanglement rank coincides with the Schmidt rank since the decomposition consists only of a single term. Sometimes it is useful to write the vectors $|\Psi_{AB,x}\rangle$ as $|M_x\rangle$, for some $M_x \in L(\mathcal{H}_A, \mathcal{H}_B)$. In this case it is useful to note that the Schmidt rank of $|M_x\rangle$ is equal to $\operatorname{rank}(M_x)$.

Note that larger entanglement rank corresponds to more entanglement since

$$\mathrm{Sep} = \mathrm{Ent}_1 \subset \cdots \subset \mathrm{Ent}_r \subset \mathrm{Ent}_{r+1} \subset \cdots \subset \mathrm{Ent}_n = \mathrm{PSD}$$

where $n = \min\{\dim \mathcal{H}_A, \dim \mathcal{H}_B\}$ and all inclusions are strict. While entanglement rank is only a rough measure of entanglement as it only takes on integer values, $r \in \{1, \dots, n\}$, it is still meaningful. Indeed, the next theorem shows that separable quantum channels cannot increase the entanglement rank (in particular, they cannot create entangled states out of separable ones).

Theorem 10.5 (Separable maps cannot increase entanglement rank). *If* $\Xi \in \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$ *and* $P \in \mathrm{Ent}_r(\mathcal{H}_A : \mathcal{H}_B)$ *then* $\Xi(P) \in \mathrm{Ent}_r(\mathcal{H}_C : \mathcal{H}_D)$.

Proof. We can write $P = \sum_y |M_y\rangle\langle M_y|$, for some $M_y \in L(\mathcal{H}_A, \mathcal{H}_B)$ such that $\operatorname{rank}(M_y) \leqslant r$. Recall from Eq. (10.1) that there exist Kraus operators $A_x \in L(\mathcal{H}_A, \mathcal{H}_C)$ and $B_x \in L(\mathcal{H}_B, \mathcal{H}_D)$ such that

$$\begin{aligned}\Xi(P) &= \sum_x \sum_y (A_x \otimes B_x)|M_y\rangle\langle M_y|(A_x \otimes B_x)^\dagger \\ &= \sum_x \sum_y |B_x M_y A_x^T\rangle\langle B_x M_y A_x^T|,\end{aligned}$$

where we used the vectorization identity from Eq. (10.2). Since $\operatorname{rank}(B_x M_y A_x^T) \leqslant \operatorname{rank}(M_y) \leqslant r$, we conclude that $\Xi(P) \in \mathrm{Ent}_r(\mathcal{H}_C : \mathcal{H}_D)$. □

As a special case of this theorem, we conclude that the set of separable operators is closed under separable maps.

Corollary 10.6 (Separable maps preserve separability). *If* $\Xi \in \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$ *and* $P \in \mathrm{Sep}(\mathcal{H}_A : \mathcal{H}_B)$ *then* $\Xi(P) \in \mathrm{Sep}(\mathcal{H}_C : \mathcal{H}_D)$.

We will shortly define LOCC channels and you will show that they are separable. Because of this, the above two results specialize also to LOCC. In particular, LOCC maps cannot increase the entanglement rank.

10.3 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 10.7 (Instrument). *An* instrument *is a collection of completely positive maps* $\{\Phi_\omega : \omega \in \Omega\} \subset \mathrm{CP}(\mathcal{H}_A, \mathcal{H}_B)$ *such that* $\sum_{\omega \in \Omega} \Phi_\omega \in \mathrm{C}(\mathcal{H}_A, \mathcal{H}_B)$. *When applied to a state* $\rho \in \mathrm{D}(\mathcal{H}_A)$, *it produces an outcome* $\omega \in \Omega$ *with probability* $\operatorname{Tr}[\Phi_\omega[\rho]]$ *and changes* ρ *to*

$$\rho_\omega = \frac{\Phi_\omega[\rho]}{\operatorname{Tr}[\Phi_\omega[\rho]]} \in \mathrm{D}(\mathcal{H}_B).$$

You will show in Practice Problem 10.4 that any instrument can be implemented by a quantum channel, followed by an orthonormal measurement.

We can now formally define the set of LOCC quantum channels that can be implemented only by local operations and classical communication.

Definition 10.8 (LOCC channel). *Let* $\Xi \in \mathrm{C}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$. *Then*

- Ξ *is a* one-way right LOCC channel *if*

$$\Xi = \sum_{\omega \in \Omega} \Phi_\omega \otimes \Psi_\omega$$

where $\{\Phi_\omega : \omega \in \Omega\} \subset \mathrm{CP}(\mathcal{H}_A, \mathcal{H}_C)$ *is an instrument and each* $\Psi_\omega \in \mathrm{C}(\mathcal{H}_B, \mathcal{H}_D)$ *is a quantum channel;*

- Ξ *is a* one-way left LOCC channel *if it is of the same form but each* $\Phi_\omega \in \mathrm{C}(\mathcal{H}_A, \mathcal{H}_C)$ *is a quantum channel and* $\{\Psi_\omega : \omega \in \Omega\} \subset \mathrm{CP}(\mathcal{H}_B, \mathcal{H}_D)$ *is an instrument;*

- Ξ *is an* LOCC channel *if it is a finite composition of the above.*

We denote the set of all LOCC channels with input $A : B$ *and output* $C : D$ *by* $\mathrm{LOCC}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$.

Intuitively, a one-way right LOCC protocol consists of Alice performing a local channel, followed by a measurement. She then sends the measurement outcome $\omega \in \Omega$ to Bob. Depending on the value of ω, bob applies a channel Ψ_ω. One-way right LOCC protocols are similar, except the communication is from Bob to Alice and Alice's channel Φ_ω depends on Bob's measurement outcome $\omega \in \Omega$.

You will show in Practice Problem 10.3 than any LOCC channel is separable, i.e.,

$$\mathrm{LOCC}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D) \subseteq \mathrm{SepC}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D). \tag{10.4}$$

10.4 Separable and LOCC measurements

It is often useful to consider measurements that are either separable or LOCC. For example, in the context of state discrimination. Imagine that Alice and Bob share a state selected from some ensemble, and they want to determine which state it is, however they cannot exchange any quantum information and can communication only classically. This corresponds to performing an LOCC measurement. In this context, it is interesting to compare how well do LOCC measurements perform compared to the slightly more general separable measurements. Before we can ask these questions, we first need to define these two types of measurements.

Definition 10.9 (Separable and LOCC measurements). *Let* $\mu : \Omega \to \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *be a measurement on systems* A *and* B. *Let* $\mathcal{H}_X = \mathcal{H}_Y = \mathbb{C}^\Omega$ *and* $\Phi_\mu \in \mathrm{C}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_X \otimes \mathcal{H}_Y)$ *be the quantum-to-classical channel corresponding to* μ*:*

$$\Phi_\mu(X) = \sum_{\omega \in \Omega} \mathrm{Tr}\big[\mu[\omega] X\big] \, |\omega\rangle\langle\omega|_X \otimes |\omega\rangle\langle\omega|_Y.$$

The measurement μ *is* separable / LOCC *if the channel* Φ_μ *is separable / LOCC.*

You can check that a measurement is separable iff each measurement operator is separable.

Lemma 10.10. *A measurement* $\mu : \Omega \to \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is separable across* $A : B$ *iff* $\mu(\omega) \in \mathrm{Sep}(\mathcal{H}_A : \mathcal{H}_B)$*, for each* $\omega \in \Omega$.

We can also define one-way right and left LOCC measurements. This is a joint measurement where the first party performs a local measurement ν, sends the outcome $\gamma \in \Gamma$ to the second party who then adaptively performs a measurement π_γ that depends on the received value γ.

Definition 10.11 (One-way right / left LOCC measurement)**.** *A measurement* $\mu : \Omega \to \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is* one-way right LOCC *if there exists a measurement* $\nu : \Gamma \to \mathrm{PSD}(\mathcal{H}_A)$ *on Alice's side and, for each* $\gamma \in \Gamma$*, a measurement* $\pi_\gamma : \Omega \to \mathrm{PSD}(\mathcal{H}_B)$ *on Bob's side such that*

$$\mu(\omega)_{AB} = \sum_{\gamma \in \Gamma} \nu(\gamma)_A \otimes \pi_\gamma(\omega)_B,$$

for every $\omega \in \Omega$*. Similarly, a the measurement* μ *is* one-way left LOCC *if*

$$\mu(\omega)_{AB} = \sum_{\gamma \in \Gamma} \pi_\gamma(\omega)_A \otimes \nu(\gamma)_B,$$

where the roles of Alice and Bob have been exchanged.

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 10.12 (Perfect one-way LOCC measurement for discriminating orthogonal pure states)**.** *If* $|\Psi_0\rangle, |\Psi_1\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *are orthogonal states, then there exists a one-way LOCC measurement* $\mu : \{0, 1\} \to \mathrm{PSD}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *such that* $\langle\Psi_0|\mu(0)|\Psi_0\rangle = \langle\Psi_1|\mu(1)|\Psi_1\rangle = 1$.

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 orthogonal pure states.

Lecture 11

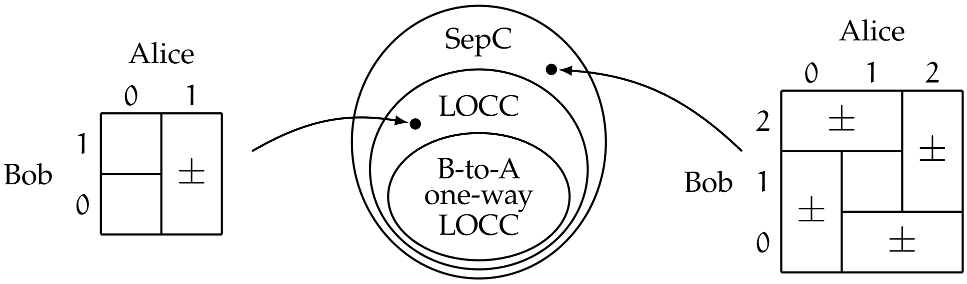
Majorization and Nielsen's theorem

Last week we looked at separable and LOCC maps and 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 Homework Problem 10.2 you saw the following set of orthogonal product states:

$$|\Psi_1\rangle = |0\rangle \otimes |0\rangle, \qquad |\Psi_2\rangle = |0\rangle \otimes |1\rangle, \qquad |\Psi_3\rangle = |1\rangle \otimes |+\rangle, \qquad |\Psi_4\rangle = |1\rangle \otimes |-\rangle$$

where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. They correspond to the tiles shown on the left:

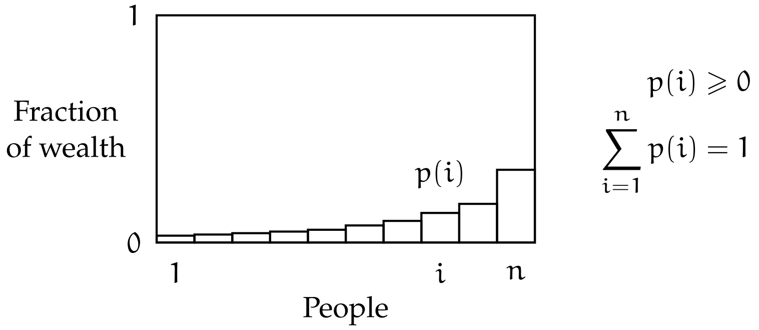


These states can be perfectly discriminated by a separable measurement or a one-way LOCC measurement from Alice to Bob, but not by a one-way LOCC measurement from Bob to Alice. Hence, as illustrated in the above diagram, the corresponding measurement is in LOCC but not in one-way LOCC from Bob to Alice. 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 corresponding measurement is clearly separable since these states form an orthonormal product basis.

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. Since the answer to this problem is closely tied with the notion of majorization, 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:

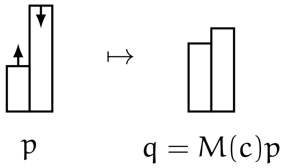


Given two probability distributions p and q, how can we tell which one corresponds to a "more equal" distribution of wealth? Clearly, $p = (1, 0, \dots, 0)$ is the least equal distribution of wealth and $q = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$ is the most equal. How about the rest and how can we compare two distributions?

One obvious way to increase equality is to take from the rich and give to the poor. Let's 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:

$$M(c) = cI + (1-c)X = \begin{pmatrix} c & 1-c \\ 1-c & c \end{pmatrix},$$

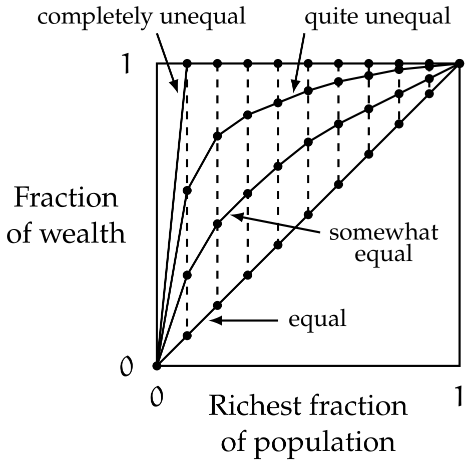
for some $c \in [0, 1]$. If $c \in (0, 1)$, applying this to the wealth of two individuals always has the effect of decreasing the gap between their wealth because the new values are convex combinations of the old ones[1]:



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.

Another 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 $p(1) \geqslant p(2) \geqslant \cdots \geqslant p(n)$, 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:

[1] When $c > 1/2$, the roles of the two individuals in terms of their richness are swapped.



If f_q lies completely below f_p then the distribution q is more equal than p. Each Robin Hood move has the effect of pushing the corresponding cumulative curve downwards. Note that not all pairs of curves can be compared since it is possible for two curves to intersect.

Let us now turn these two intuitive ways of comparing wealth inequality into precise mathematical statements and show that the two approaches discussed above are actually equivalent. Before we do this, let us introduce two central concepts – doubly stochastic and permutation matrices.

Definition 11.1 (Stochastic and doubly stochastic matrices)**.** *Let* $A \in L(\mathbb{R}^\Sigma)$.

- *We call* A stochastic *if*

  1. $A_{ij} \geqslant 0$, *for all* $i, j \in \Sigma$,
  2. $\sum_{i\in\Sigma} A_{ij} = 1$, *for all* $j \in \Sigma$.

  This is equivalent to saying that each column of A *is a probability distribution.*

- *We call* A doubly stochastic *if it is stochastic and*

  3. $\sum_{j\in\Sigma} A_{ij} = 1$, *for all* $i \in \Sigma$.

  Equivalently, each row and each column of A *is a probability distribution.*

- *We say that* A *is a* permutation matrix *if it is doubly stochastic and*

  4. $A_{ij} \in \{0, 1\}$, *for all* $i, j \in \Sigma$.

  Equivalently, each row and each column of A *contains exactly one entry* 1 *and the rest are zeroes.*

It is a simple but important observation that stochastic matrices are precisely those that preserve the ℓ^1-norm of vectors. Note that stochastic and even doubly stochastic matrices are generally not invertible, e.g., the matrix $\frac{1}{2}\left(\begin{smallmatrix}1&1\\1&1\end{smallmatrix}\right)$. Permutation matrices stand out as precisely those stochastic matrices that are invertible and whose inverse is also stochastic. In fact, inverting a permutation matrix A is particularly simple since A^T is the inverse of A. Note also that any convex combination of permutations is doubly stochastic, since each row and column is a convex combination of the standard basis vectors and hence a probability distribution. Surprisingly, the converse claim is also true.

Theorem 11.2 (Birkhoff–von Neumann). *$A \in L(\mathbb{R}^\Sigma)$ is doubly stochastic iff there exists a probability distribution p over S_Σ (the set of all permutations acting on Σ) such that*

$$A = \sum_{\pi \in S_\Sigma} p(\pi) V_\pi,$$

where $V_\pi \in L(\mathbb{R}^\Sigma)$ is the permutation matrix corresponding to π, i.e., $(V_\pi)_{ij} = \delta_{i,\pi(j)}$.

Let us mention another fact without proof, namely that a matrix is doubly stochastic iff it can be decomposed as a sequence of Robin Hood moves. Since each Robin Hood move makes a wealth distribution more equal, so does a doubly stochastic matrix. This motivates the following definition.

Definition 11.3 (Majorization). *Let $u, v \in \mathbb{R}^\Sigma$. Then u* majorizes *v if $v = Au$, for some doubly stochastic $A \in L(\mathbb{R}^\Sigma)$. We write this as $u \succ v$ or $v \prec u$.*

To remember this notation, think of "$\succ$" as pointing in the direction in which the mapping is applied, i.e., $u \succ v$ means that u is converted to v by applying some mapping A. If u and v are probability distributions, then $u \succ v$ means that v is more equal than u. It is not immediately clear how to check this condition since it seems to require going through all doubly stochastic matrices. Luckily, there is an equivalent condition that is simpler to check.

Let $r(u) \in \mathbb{R}^\Sigma$ denote the *reverse sorting* of vector $u \in \mathbb{R}^\Sigma$, i.e.,

$$r_1(u) \geqslant \cdots \geqslant r_n(u) \qquad \text{and} \qquad \{r_1(u), \ldots, r_n(u)\} = \{u_1, \ldots, u_n\}.$$

Then the following theorem provides a simple way to check the majorization condition (you will derive other equivalent conditions in Practice Problem 11.2).

Theorem 11.4. *Let $u, v \in \mathbb{R}^n$. Then $v \prec u$ iff $\sum_{i=1}^m r_i(v) \leqslant \sum_{i=1}^m r_i(u)$, for all $m \in \{1, \ldots, n-1\}$, and $\sum_{i=1}^n r_i(v) = \sum_{i=1}^n r_i(u)$.*

When restricted to probability distributions, we do not need to check the last condition since the entries automatically sum to one.

Example 11.5 (Extreme distributions). *Let $p = (1, 0, \ldots, 0)$ denote a deterministic distribution and $q = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ denote the uniform distribution on n elements. Intuitively, they correspond to a maximally unequal and maximally equal way to distribute wealth, respectively. Indeed, one can easily check using Theorem 11.4 that, for any probability distribution s on n elements,*

$$q \prec s \prec p.$$

This means that one can always perform the conversions

$$q \leftarrow\!\shortmid s \leftarrow\!\shortmid p$$

either by a sequence of Robin Hood moves or by applying doubly stochastic matrices.

11.2 Majorization for Hermitian operators

Recall that a diagonal density matrix can be identified with the probability distribution it contains on the diagonal and hence considered classical. Thus one might wonder whether the theory of majorization has a "quantum" extension that deals with Hermitian operators instead of vectors? Indeed, such a generalization is possible. To obtain it, we need to ask what operations will play the role of permutations?

Recall from our earlier discussion that stochastic matrices are precisely those that map probability distributions to probability distributions, i.e., they preserve the ℓ^1-norm of vectors. Quantum channels play the same role for quantum states – they map density matrices to density matrices (even when applied to a subsystem). Among all stochastic matrices, permutations stand out as precisely those that are invertible and whose inverse is also stochastic. In the quantum case, unitary channels are the invertible ones. This motivates the following definition for a quantum analogue of a doubly stochastic matrix.

Definition 11.6 (Mixed-unitary channel). *$\Phi \in C(\mathcal{H})$ is a* mixed-unitary channel *if*

$$\Phi(M) = \sum_{i \in \Sigma} p(i) U_i M U_i^\dagger,$$

for some set Σ, a probability distribution $p \in P(\Sigma)$, and unitaries $U_i \in U(\mathcal{H})$.

The following is then a quantum generalization of Definition 11.3.

Definition 11.7 (Majorization for Hermitian operators). *Let A and B be Hermitian operators on $\mathcal{H}$. Then A majorizes B if $B = \Phi(A)$, for some mixed-unitary $\Phi \in C(\mathcal{H})$. We write this as $A \succ B$ or $B \prec A$.*

It is even less obvious how one is supposed to check this condition compared to the classical one in Definition 11.3. Luckily, the following theorem[2] expresses this condition entirely in terms of the much simpler classical condition, which we already know how to check thanks to Theorem 11.4.

Theorem 11.8 (Uhlmann). *Let A and B be Hermitian operators on $\mathcal{H}$. Then $B \prec A$ iff $\lambda(B) \prec \lambda(A)$, where $\lambda(A) \subset \mathbb{R}$ denotes the spectrum (i.e., the set of eigenvalues) of A.*

Note that for diagonal matrices A, the spectrum $\lambda(A)$ is just the set of diagonal entries of A. Hence, majorization for diagonal operators reduces to majorization for vectors, thus recovering the classical notion.

11.3 Nielsen's theorem

Other than generalizing the notion of majorization to operators, this lecture so far has been almost entirely classical (indeed, we have not seen a single bra or ket yet). Also, it is not clear how all this relates to our main topic of the last two weeks – entanglement and LOCC. This connection is established by Nielsen's theorem, which establishes a condition under which one bipartite pure state can be converted into another under (one-way) LOCC. This condition is expressed succinctly in terms of majorization of the reduced states.

Theorem 11.9 (Nielsen). *Let $|u_{AB}\rangle, |v_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be pure states. The following are equivalent:*

1. $\mathrm{Tr}_A\big[|u\rangle\langle u|\big] \prec \mathrm{Tr}_A\big[|v\rangle\langle v|\big]$.
2. $\Xi\big[|u\rangle\langle u|\big] = |v\rangle\langle v|$, *for some one-way LOCC protocol $\Xi \in \mathrm{LOCC}(A : B)$ from Alice to Bob.*

[2] Not to be confused with the more famous Uhlmann's Theorem 3.5.

3. *Same, but Ξ is one-way LOCC from Bob to Alice.*

4. *Same, but $\Xi \in \mathrm{SepC}(A : B)$.*

Remark 11.10. *Note that here, unlike in Definitions 11.3 and 11.7, the direction of majorization is* opposite *to the direction in which the processing occurs: we transform $|u\rangle$ to $|v\rangle$, while the majorization sign is pointing towards $|u\rangle$. See Example 11.11 below for an illustration of why it is so.*

Proof. (1 ⇒ 2) The main idea of the proof is to take the mixed-unitary channel from Definition 11.6, which we get thanks to the majorization condition, and turn it into a one-way LOCC protocol. Doing this is not straightforward at all and takes a number of steps.

First, you will show in Practice Problem 11.3 that, for any $L, R \in L(\mathcal{H}_A, \mathcal{H}_B)$,

$$\mathrm{Tr}_A \left[|L\rangle\langle R|\right] = LR^\dagger. \tag{11.1}$$

In particular, if $u, v \in L(\mathcal{H}_A, \mathcal{H}_B)$ are operators whose vectorizations coincide with $|u\rangle$ and $|v\rangle$, respectively, then the majorization identity between the reduced states is equivalent to

$$uu^\dagger \prec vv^\dagger.$$

By Definition 11.7, there exists a mixed-unitary channel $\Phi \in C(\mathcal{H}_B)$ such that $uu^\dagger = \Phi(vv^\dagger)$. Recall from Definition 11.6 that $\Phi(M) = \sum_{i\in\Sigma} p(i) W_i M W_i^\dagger$, for some set Σ, a probability distribution p over Σ, and unitaries $W_i \in U(\mathcal{H}_B)$. Hence,

$$uu^\dagger = \sum_{i\in\Sigma} p(i) W_i v v^\dagger W_i^\dagger.$$

Based on this relation, we want to construct a one-way LOCC protocol from Alice to Bob.

Let us first write the relation in a more symmetric way by introducing another sum:

$$uu^\dagger = \left(\sum_{i\in\Sigma} \sqrt{p(i)}(W_i v) \otimes \langle i|\right)\left(\sum_{i'\in\Sigma} \sqrt{p(i')}(W_{i'} v) \otimes |i'\rangle\right) = ww^\dagger$$

where $w \in L(\mathcal{H}_C, \mathcal{H}_B)$ with $\mathcal{H}_C := \mathcal{H}_A \otimes \mathbb{C}^\Sigma$ is defined as

$$w := \sum_{i\in\Sigma} \sqrt{p(i)}(W_i v) \otimes \langle i|.$$

Given that $uu^\dagger = ww^\dagger$, how are the operators u and w related?

To find a relationship between u and w, let

$$u = \sum_{j=1}^{r} s_j |b_j\rangle\langle a_j| \tag{11.2}$$

be the singular value decomposition of u (see Lemma 2.16). Here $r := \mathrm{rank}(u)$ is the rank of u or the Schmidt rank of $|u\rangle$ (see Lemma 2.12), $s_j > 0$ are the singular values of u or the Schmidt coefficients of $|u\rangle$, and $\{|a_j\rangle : j = 1, \ldots, r\}$ and $\{|b_j\rangle : j = 1, \ldots, r\}$ are some orthonormal sets of vectors in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Note from Eq. (11.2) that

$$uu^\dagger = \sum_{j,k=1}^{r} s_j s_k |b_j\rangle\langle a_j|a_k\rangle\langle b_k| = \sum_{j=1}^{r} s_j^2 |b_j\rangle\langle b_j| = ww^\dagger, \tag{11.3}$$

so $uu^\dagger$ and $ww^\dagger$ both have eigenvalues s_j^2 with corresponding eigenvectors $|b_j\rangle$. Hence w has singular value decomposition

$$w = \sum_{j=1}^{r} s_j |b_j\rangle\langle c_j|,$$

for some orthonormal set of vectors $\{|c_j\rangle : j = 1, \dots, r\}$ in $\mathcal{H}_C$. Note from Eq. (11.2) that u and w have the same singular values s_j and left singular vectors $|b_j\rangle$.

Since $\{|a_j\rangle : j = 1, \dots, r\}$ and $\{|c_j\rangle : j = 1, \dots, r\}$ are two orthonormal bases of the same dimension, we can find an isometry $V \in U(\mathcal{H}_A, \mathcal{H}_C)$ such that $V|a_j\rangle = |c_j\rangle$, for all $j \in \Sigma$. Equivalently,

$$uV^\dagger = w = \sum_{i\in\Sigma} \sqrt{p(i)}(W_i v) \otimes \langle i|. \tag{11.4}$$

Based on this observation, let us devise a one-way LOCC protocol from Alice to Bob. Recall from Definition 10.8 that such protocol is described by a channel $\Xi \in C(\mathcal{H}_A \otimes \mathcal{H}_B)$ of the form

$$\Xi = \sum_{i\in\Sigma} \Phi_i \otimes \Psi_i,$$

where $\{\Phi_i : i \in \Sigma\} \subset CP(\mathcal{H}_A)$ is an instrument (see Definition 10.7) on Alice's side and each $\Psi_i \in C(\mathcal{H}_B)$ is a channel on Bob's side. Let us choose the superoperators Φ_i and Ψ_i as follows:

$$\Phi_i(M_A) := A_i M_A A_i^\dagger, \qquad \Psi_i(M_B) := U_i M_B U_i^\dagger, \tag{11.5}$$

for all $M_A \in L(\mathcal{H}_A)$ and $M_B \in L(\mathcal{H}_B)$, where

$$\{A_i : i \in \Sigma\} \subset L(\mathcal{H}_A), \qquad \{U_i : i \in \Sigma\} \subset U(\mathcal{H}_B)$$

are Kraus operators of Alice's measurement and the corresponding basis change operators for Bob. The overall one-way LOCC protocol then acts as

$$\Xi(M_{AB}) = \sum_{i\in\Sigma} (A_i \otimes U_i) M_{AB} (A_i \otimes U_i)^\dagger,$$

for all $M_{AB} \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$. Note that here Σ plays the role of possible messages Alice may transmit to Bob: if Alice gets measurement outcome $i \in \Sigma$, she sends the value of i to Bob and he applies the corresponding basis change U_i.

Note from Eq. (11.5) that each Ψ_i is indeed a quantum channel, as it is just a unitary change of basis. Moreover, each Φ_i is indeed completely positive. Hence, we only need to make sure that $\{\Phi_i : i \in \Sigma\}$ is indeed an instrument by checking that $\sum_{i\in\Sigma} \Phi_i$ is trace-preserving. According to Lemma 4.4, this amounts to checking that

$$\sum_{i\in\Sigma} A_i^\dagger A_i = I_A. \tag{11.6}$$

Let us now proceed to actually construct the protocol by appropriately choosing the operators A_i and U_i in Eq. (11.5). Recall from Eq. (10.2) that

$$(A_i \otimes U_i)|u\rangle = |U_i u A_i^T\rangle.$$

We would like this to be equal to $\sqrt{p(i)}|v\rangle$ since then the output of the channel would be

$$\Xi(|u\rangle\langle u|) = \sum_{i\in\Sigma} (A_i \otimes U_i)|u\rangle\langle u|(A_i \otimes U_i)^\dagger = \sum_{i\in\Sigma} |U_i u A_i^\mathsf{T}\rangle\langle U_i u A_i^\mathsf{T}| = \sum_{i\in\Sigma} p(i)|v\rangle\langle v| = |v\rangle\langle v|,$$

as desired. The unvectorized version of the desired identity $|U_i u A_i^\mathsf{T}\rangle = \sqrt{p(i)}|v\rangle$ is

$$U_i u A_i^\mathsf{T} = \sqrt{p(i)}v. \tag{11.7}$$

Recall from Eq. (11.4) that

$$uV^\dagger = \sum_{i\in\Sigma} \sqrt{p(i)}(W_i v) \otimes \langle i|.$$

We can make this look more like the desired identity in Eq. (11.7) by selecting only one value of $i \in \Sigma$ and canceling out the undesired unitary W_i:

$$W_i^\dagger(uV^\dagger)(I_A \otimes |i\rangle) = \sqrt{p(i)}v.$$

To recover Eq. (11.7), we choose

$$U_i := W_i^\dagger, \qquad\qquad A_i^\mathsf{T} := V^\dagger(I_A \otimes |i\rangle).$$

In other words, $A_i^\dagger = V^\mathsf{T}(I_A \otimes |i\rangle)$ and $A_i = (I_A \otimes \langle i|)\overline{V}$.

It remains to verify that this corresponds to a well-defined one-way LOCC protocol from Alice to Bob. Bob's superoperators Ψ_i in Eq. (11.5) are clearly quantum channels since all U_i are unitary. To show that Alice's superoperators Φ_i form a proper instrument, we need to verify Eq. (11.6) to make sure that A_i are valid Kraus operators:

$$\begin{aligned}
\sum_{i\in\Sigma} A_i^\dagger A_i &= \sum_{i\in\Sigma} V^\mathsf{T}(I_A \otimes |i\rangle)(I_A \otimes \langle i|)\overline{V} \\
&= \sum_{i\in\Sigma} V^\mathsf{T}(I_A \otimes |i\rangle\langle i|)\overline{V} \\
&= V^\mathsf{T}(I_A \otimes I_{|\Sigma|})\overline{V} \\
&= V^\mathsf{T}\overline{V} \\
&= \overline{V^\dagger V} \\
&= I_A,
\end{aligned}$$

where we used the fact that $V \in U(\mathcal{H}_A, \mathcal{H}_C)$, where $\mathcal{H}_C = \mathcal{H}_A \otimes \mathbb{C}^\Sigma$, is an isometry.

($1 \Rightarrow 3$) Same, but with the roles of Alice and Bob exchanged.

($2 \Rightarrow 4$) and ($3 \Rightarrow 4$) Every LOCC channel is separable, see Eq. (10.4).

($4 \Rightarrow 1$) The main idea of the proof is to restate the desired majorization condition in terms of the simpler notion for probability distributions (Theorem 11.8) and then restate that again in terms of partial sums (Theorem 11.4). This simplified condition is similar to a variational characterization of eigenvalues. We can prove it by truncating the singular value decomposition of u and combining it with Kraus operators of Ξ.

Let $\Xi \in \text{SepC}(A : B)$ be a separable channel (see Definition 10.1) with Kraus operators $\{A_i \otimes B_i : i \in \Sigma\} \subset L(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that

$$\Xi(|u\rangle\langle u|) = \sum_{i\in\Sigma} (A_i \otimes B_i)|u\rangle\langle u|(A_i \otimes B_i)^\dagger = |v\rangle\langle v|. \tag{11.8}$$

Since $|v\rangle\langle v|$ is a rank-1 matrix, each term in the above sum must be of the form $p(i)|v\rangle\langle v|$, for some probability distribution $p \in P(\Sigma)$. By using our most beloved vectorization identity from Eq. (10.2), we conclude that

$$|B_i u A_i^{\mathsf{T}}\rangle\langle B_i u A_i^{\mathsf{T}}| = p(i)|v\rangle\langle v|,$$

for every $i \in \Sigma$. Taking partial trace over A and using Eq. (11.1) (which is now our second most beloved vectorization identity),

$$B_i u A_i^{\mathsf{T}} \overline{A}_i u^\dagger B_i^\dagger = p(i) v v^\dagger. \tag{11.9}$$

Recall that our goal is to show that $uu^\dagger \prec vv^\dagger$. By Theorem 11.8 (which is now our second most favorite Uhlmann's Theorem), this is equivalent to

$$\lambda(uu^\dagger) \prec \lambda(vv^\dagger), \tag{11.10}$$

where $\lambda(M)$ denotes the spectrum of M and "$\prec$" is the much simpler notion of majorization for vectors (see Definition 11.3). We can restate our goal further by invoking Theorem 11.4, which expresses the majorization condition for vectors in terms of partial sums.

Let $n = \dim(\mathcal{H}_B)$ and let $\lambda_j(M)$ denote the j-th largest eigenvalue of $M \in \mathrm{PSD}(\mathcal{H}_B)$, i.e.,

$$\lambda_1(M) \geqslant \cdots \geqslant \lambda_j(M) \geqslant \cdots \geqslant \lambda_n(M).$$

Since $uu^\dagger = \mathrm{Tr}_A\big[|u\rangle\langle u|\big]$ and $|u\rangle$ is a unit vector,

$$\sum_{j=1}^{n} \lambda_j(uu^\dagger) = \mathrm{Tr}[uu^\dagger] = 1 = \mathrm{Tr}[vv^\dagger] = \sum_{j=1}^{n} \lambda_j(vv^\dagger),$$

which is the last condition in Theorem 11.4. It remains to show that, for all $m \in \{1, \dots, n\}$,

$$\sum_{j=m}^{n} \lambda_j(vv^\dagger) \leqslant \sum_{j=m}^{n} \lambda_j(uu^\dagger), \tag{11.11}$$

which (thanks to Practice Problem 11.2) is equivalent[3] to the remaining conditions in Theorem 11.4.

Since $\lambda_j(cM) = c\lambda_j(M)$, for any $j \in \{1, \dots, n\}$, $c > 0$, and $M \in \mathrm{PSD}(\mathcal{H}_B)$,

$$\sum_{j=m}^{n} \lambda_j(vv^\dagger) = \sum_{j=m}^{n} \sum_{i\in\Sigma} \lambda_j(p(i)vv^\dagger) = \sum_{i\in\Sigma} \sum_{j=m}^{n} \lambda_j\big(B_i u A_i^{\mathsf{T}} \overline{A}_i u^\dagger B_i^\dagger\big), \tag{11.12}$$

where we used Eq. (11.9) and the distribution p defined earlier. Let us denote the monstrous argument of λ_j by $P_i \in \mathrm{PSD}(\mathcal{H}_B)$. Since the sum is over the $n - m + 1$ smallest eigenvalues of P_i,

$$\sum_{j=m}^{n} \lambda_j(P_i) \leqslant \mathrm{Tr}\big[\Pi_{i,m} P_i\big], \tag{11.13}$$

for any projector $\Pi_{i,m} \in \mathrm{PSD}(\mathcal{H}_B)$ of $\mathrm{rank}(\Pi_{i,m}) \geqslant n - m + 1$, thanks to the variational characterization of the eigenvalues of P_i.

[3] Note that the roles of u and v in $\lambda(uu^\dagger) \prec \lambda(vv^\dagger)$ in Eq. (11.10) are reversed compared to Definition 11.3 and Theorem 11.4, which are stated for $v \prec u$. However, the order of summation in Eq. (11.11) is also reversed, which gives us back the correct inequality thanks to the equivalence of the first two conditions in Practice Problem 11.2.

For each $i \in \Sigma$, let us choose the projector $\Pi_{i,m}$ so that $\Pi_{i,m}B_i|b_j\rangle = 0$, for all $j \in \{1, \dots, m-1\}$, where B_i comes from Eq. (11.8) and $|b_j\rangle$ is the j-th left singular vector of u (see Eq. (11.2)) or 0 if $j > r = \operatorname{rank}(u)$. Since $\operatorname{span}\{B_i|b_j\rangle : j = 1, \dots, m-1\}$ has dimension at most $m-1$, we do not violate the restriction that $\operatorname{rank}(\Pi_{i,m}) \geqslant n - m + 1$.

Next, let us truncate the singular value decomposition of u in Eq. (11.2) and define

$$u_m := \sum_{j=m}^{r} s_j|b_j\rangle\langle a_j|. \tag{11.14}$$

By our choice of $\Pi_{i,m}$,

$$\Pi_{i,m}B_i u = \sum_{j=1}^{r} s_j \Pi_{i,m}B_i|b_j\rangle\langle a_j| = \sum_{j=m}^{r} s_j \Pi_{i,m}B_i|b_j\rangle\langle a_j| = \Pi_{i,m}B_i u_m.$$

By taking the conjugate transpose of both sides, $u^\dagger B_i^\dagger \Pi_{i,m} = u_m^\dagger B_i^\dagger \Pi_{i,m}$. As a consequence,

$$\operatorname{Tr}\left[\Pi_{i,m}P_i\right] = \operatorname{Tr}\left[\Pi_{i,m}P_{i,m}\right] \tag{11.15}$$

where $P_{i,m} := B_i u_m A_i^T \overline{A}_i u_m^\dagger B_i^\dagger$ is the truncated cousin of $P_i = B_i u A_i^T \overline{A}_i u^\dagger B_i^\dagger$.

To summarize, we know from Eq. (11.12) that

$$\sum_{j=m}^{n} \lambda_j(vv^\dagger) = \sum_{i\in\Sigma}\sum_{j=m}^{n} \lambda_j(P_i).$$

Moreover, by combining Eqs. (11.13) and (11.15), we also know that

$$\sum_{j=m}^{n} \lambda_j(P_i) \leqslant \operatorname{Tr}\left[\Pi_{i,m}P_i\right] = \operatorname{Tr}\left[\Pi_{i,m}P_{i,m}\right] \leqslant \operatorname{Tr}\left[P_{i,m}\right].$$

Putting these two observations together,

$$\sum_{j=m}^{n} \lambda_j(vv^\dagger) \leqslant \sum_{i\in\Sigma} \operatorname{Tr}\left[P_{i,m}\right] = \operatorname{Tr}\left[\sum_{i\in\Sigma} P_{i,m}\right]. \tag{11.16}$$

Our goal is to relate the right-hand side to $\sum_{j=m}^{n} \lambda_j(uu^\dagger)$, so that we can finally prove Eq. (11.11).

Let us investigate the operators $P_{i,m}$ in more detail. By using both vectorization identities (Eqs. (4.9) and (11.1)) backwards,

$$\begin{aligned} P_{i,m} &= B_i u_m A_i^T \overline{A}_i u_m^\dagger B_i^\dagger \\ &= \operatorname{Tr}_A\left[|B_i u_m A_i^T\rangle\langle B_i u_m A_i^T|\right] \\ &= \operatorname{Tr}_A\left[(A_i \otimes B_i)|u_m\rangle\langle u_m|(A_i \otimes B_i)^\dagger\right]. \end{aligned}$$

Note that

$$\sum_{i\in\Sigma} P_{i,m} = \operatorname{Tr}_A\left[\Xi\left[|u_m\rangle\langle u_m|\right]\right],$$

where Ξ is the separable channel from Eq. (11.8) that we started with. Since Ξ is trace-preserving and $\operatorname{Tr}_A\left[|u_m\rangle\langle u_m|\right] = u_m u_m^\dagger$,

$$\operatorname{Tr}\left[\sum_{i\in\Sigma} P_{i,m}\right] = \operatorname{Tr}\left[\Xi\left[|u_m\rangle\langle u_m|\right]\right] = \operatorname{Tr}\left[|u_m\rangle\langle u_m|\right] = \operatorname{Tr}\left[\operatorname{Tr}_A\left[|u_m\rangle\langle u_m|\right]\right] = \operatorname{Tr}\left[u_m u_m^\dagger\right].$$

Substituting the singular value decomposition of u_m from Eq. (11.14),

$$\operatorname{Tr}\left[u_m u_m^\dagger\right] = \operatorname{Tr}\left[\sum_{j=m}^{r} s_j |b_j\rangle\langle a_j| \sum_{k=m}^{r} s_k |a_k\rangle\langle b_k|\right] = \sum_{j=m}^{r} s_j^2$$

since $\langle a_j|a_k\rangle = \langle b_j|b_k\rangle = \delta_{j,k}$. Combining this with Eq. (11.16),

$$\sum_{j=m}^{n} \lambda_j(vv^\dagger) \leqslant \operatorname{Tr}\left[\sum_{i\in\Sigma} P_{i,m}\right] = \operatorname{Tr}\left[u_m u_m^\dagger\right] = \sum_{j=m}^{r} s_j^2.$$

We will be done with proving Eq. (11.11) if we manage to show that

$$\sum_{j=m}^{r} s_j^2 = \sum_{j=m}^{n} s_j^2 = \sum_{j=m}^{n} \lambda_j(uu^\dagger).$$

The first equality follows since $\operatorname{rank}(u) = r$, so $s_j = 0$ for $j > r$. For the second equality, note that $s_j^2 = \lambda_j(uu^\dagger)$, for all $j \in \{1, \dots, n\}$, which follows from an earlier calculation in Eq. (11.3) showing that

$$uu^\dagger = \sum_{j=1}^{r} s_j^2 |b_j\rangle\langle b_j|.$$

This concludes the proof of the final implication and hence the theorem. □

To illustrate the use of Nielsen's theorem, consider the following example, which is a quantum version of Example 11.5.

Example 11.11 (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 $|\Psi_{AB}\rangle = |\alpha_A\rangle \otimes |\beta_B\rangle$, for some $|\alpha_A\rangle \in \mathcal{H}_A$ and $|\beta_B\rangle \in \mathcal{H}_B$, denote any product state and let $|\Phi^+_{AB}\rangle := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |i_A\rangle \otimes |i_B\rangle$ denote the maximally entangled state on this system. We see from Example 11.5 that*

$$\operatorname{Tr}_A\left[|\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|\right] \prec \operatorname{Tr}_A\left[|\Omega_{AB}\rangle\langle\Omega_{AB}|\right] \prec \operatorname{Tr}_A\left[|\Psi_{AB}\rangle\langle\Psi_{AB}|\right].$$

It follows from Nielsen's theorem that, for any pure state $|\Omega_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, one can perform the following sequence of conversions by one-way LOCC:

$$|\Phi^+_{AB}\rangle \mapsto |\Omega_{AB}\rangle \mapsto |\Psi_{AB}\rangle.$$

In other words, the maximally entangled state can be converted to any pure state, and any 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).

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. This Intuitively corresponds to comparing how unequal are the wealth distributions corresponding to these probability distributions – a distribution q is more equal than p, written $q \prec p$, if q can be obtained from p by taking away from the rich and giving to the poor.

This notion has a natural quantum counterpart where probability distributions are replaced by density matrices. Using this more general notion of majorization for Hermitian operators, Nielsen's theorem provides an elegant answer to the following problem: can $|u_{AB}\rangle$ be converted to $|v_{AB}\rangle$ by LOCC between systems A and B? This is possible if and only if $\mathrm{Tr}_A\left[|u\rangle\langle u|\right] \prec \mathrm{Tr}_A\left[|v\rangle\langle v|\right]$. This condition is easy to check by computing the eigenvalues of the reduced states and checking whether they obey the desired majorization relation.

Intuitively, if $|u_{AB}\rangle$ be converted to $|v_{AB}\rangle$ by LOCC then $|u_{AB}\rangle$ is more entangled than $|v_{AB}\rangle$ because local operations and classical communicaiton 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 (the same is also true for probability distributions). Even when two states are comparable, Nielsen's theorem doesn't 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 which tells us how much entanglement the state has. This number should be easy to compute and also have some intuitive operational interpretation. In this class we will see how this can be done for bipartite pure states.

12.1 Entanglement transformations

A convenient way to measure the amount of entanglement for bipartite states would be to choose a "golden standard" state and ask how many of these states can be obtained from the given state by LOCC. The canonical maximally entangled two-qbuit state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}$$

is a natural choice of such "golden standard". We will denote its density matrix by

$$\phi := |\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

This approach is analogous to the idea of distillation – a process by which a large amount of an impure substance is refined to a smaller amount of a more concentrated and pure substance.

One can also ask the opposite question – how many copies of $|\Phi^+_{AB}\rangle$ are required to produce one copy of some desired state? Recall from Example 11.11 that, according to Nielsen's theorem, any pure two-qubit state can be obtained from $|\Phi^+_{AB}\rangle$ by LOCC. However, if the desired target state is not too entangled, more copies might be obtainable from a single copy of $|\Phi^+_{AB}\rangle$. Unfortunately, Nielsen's theorem does not directly address this question.

Another shortcoming of Nielsen's theorem is exactness – the theorem gives an iff condition for when one bipartite state can be converted to another *exactly*. However, from practical perspective, getting sufficiently close to the desired target state might already be good enough.

To address these shortcomings, we would like to introduce a robust way to quantify the amount of entanglement in any bipartite state. Since we are interested in conversion *rates* – namely, the number of copies consumed versus the number of copies produced – we need to refer to bipartite spaces that consist of several copies of the same system. There is a slight ambiguity of how the registers are ordered. If we have n copies of $\mathcal{H}_A \otimes \mathcal{H}_B$, the resulting Hilbert space is

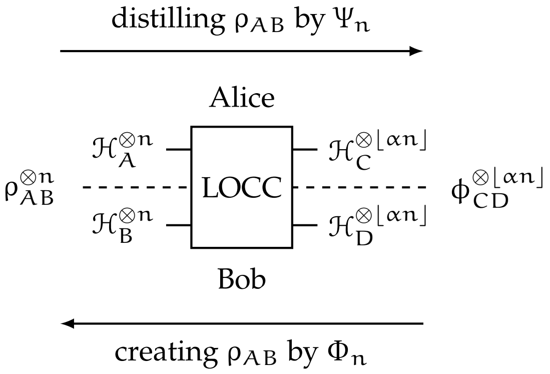
$$(\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes n} = (\mathcal{H}_A \otimes \mathcal{H}_B) \otimes \cdots \otimes (\mathcal{H}_A \otimes \mathcal{H}_B).$$

However, since the LOCC protocol is performed with respect to the Alice versus Bob separation, we need to group together all A systems and all B systems:

$$\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_B^{\otimes n} = \underbrace{(\mathcal{H}_A \otimes \cdots \otimes \mathcal{H}_A)}_{\text{Alice}} \otimes \underbrace{(\mathcal{H}_B \otimes \cdots \otimes \mathcal{H}_B)}_{\text{Bob}}$$

Permuting the systems around amounts to a non-trivial operation. However, we will not explicitly write it since the order of registers should be clear from the context.

Since the input and output systems can generally have different dimensions, we will denote them by $A : B$ and $C : D$, where A and C belgon to Alice and B and D belong to Bob. Then the opposite processes of entanglement distillation and creation can be illustrated as follows:


distilling ρ_{AB} by Ψ_n
Alice
$\mathcal{H}_A^{\otimes n}$
$\mathcal{H}_C^{\otimes\lfloor\alpha n\rfloor}$
$\rho_{AB}^{\otimes n}$
LOCC
$\phi_{CD}^{\otimes\lfloor\alpha n\rfloor}$
$\mathcal{H}_B^{\otimes n}$
$\mathcal{H}_D^{\otimes\lfloor\alpha n\rfloor}$
Bob
creating ρ_{AB} by Φ_n

Using these conventions, let us formally define the rates of entanglement distillation and creation for a given state ρ_{AB}.

Definition 12.1 (Distillable entanglement)**.** *The* distillable entanglement $E_D(\rho_{AB})$ *of state* $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is the supremum over all* $\alpha \geqslant 0$ *for which there exists a sequence of LOCC channels*

$$\Psi_n \in \mathrm{LOCC}(\mathcal{H}_A^{\otimes n}, \mathcal{H}_C^{\otimes \lfloor \alpha n \rfloor} : \mathcal{H}_B^{\otimes n}, \mathcal{H}_D^{\otimes \lfloor \alpha n \rfloor})$$

such that

$$\lim_{n \to \infty} F\big(\Psi_n(\rho_{AB}^{\otimes n}), \phi_{CD}^{\otimes \lfloor \alpha n \rfloor}\big) = 1.$$

Definition 12.2 (Entanglement cost)**.** *The* entanglement cost $E_C(\rho_{AB})$ *of state* $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ *is the infimum over all* $\alpha \geqslant 0$ *for which there exists a sequence of LOCC channels*

$$\Phi_n \in \mathrm{LOCC}(\mathcal{H}_C^{\otimes \lfloor \alpha n \rfloor}, \mathcal{H}_A^{\otimes n} : \mathcal{H}_D^{\otimes \lfloor \alpha n \rfloor}, \mathcal{H}_B^{\otimes n})$$

such that

$$\lim_{n \to \infty} F\big(\Phi_n(\phi_{CD}^{\otimes \lfloor \alpha n \rfloor}), \rho_{AB}^{\otimes n}\big) = 1.$$

How does distillable entanglement and entanglement cost compare? 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! The following lemma captures our intuition that it is more difficult to create than destroy.

Lemma 12.3 (No free lunch)**.** *For any state* $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$*,* $E_C(\rho_{AB}) \geqslant E_D(\rho_{AB})$*.*

Proof. Let's try to approximately implement by LOCC the map $\Psi_n \circ \Phi_n$ such that

$$\phi^{\otimes m} \xmapsto{\Phi_n} \rho^{\otimes n} \xmapsto{\Psi_n} \phi^{\otimes k},$$

for some integers $m, n, k \geqslant 0$. Note that $\phi^{\otimes m}$ is equivalent to a maximally entangled state of dimension 2^m, hence its Schmidt rank is 2^m. For pure states, Schmidt rank coincides with the entanglement rank (see Definition 10.4). Since $\Psi_n \circ \Phi_n$ is also LOCC (and thus separable), the output state $(\Psi_n \circ \Phi_n)(\phi^{\otimes m})$ has entanglement rank at most 2^m by Theorem 10.5. You will show in Homework Problem 12.3 that $F(\sigma, \phi^{\otimes k})^2 \leqslant r/2^k$, for any state σ of entanglement rank r. Hence,

$$F\big((\Psi_n \circ \Phi_n)(\phi^{\otimes m}), \phi^{\otimes k}\big)^2 \leqslant 2^m/2^k = 2^{m-k}. \tag{12.1}$$

By Definitions 12.1 and 12.2 of entanglement cost and distillable entanglement, for all $\varepsilon > 0$ there exists n such that

$$F\big(\Phi_n(\phi^{\otimes m}), \rho^{\otimes n}\big) > 1 - \varepsilon,$$
$$F\big(\Psi_n(\rho^{\otimes n}), \phi^{\otimes k}\big) > 1 - \varepsilon.$$

By Practice Problem 12.2,

$$F\big((\Psi_n \circ \Phi_n)(\phi^{\otimes m}), \phi^{\otimes k}\big) > 1 - 4\varepsilon.$$

Taking $\varepsilon < 1/16$,

$$F\big((\Psi_n \circ \Phi_n)(\phi^{\otimes m}), \phi^{\otimes k}\big)^2 > \left(1 - \frac{1}{4}\right)^2 = \frac{9}{16} > \frac{1}{2}.$$

Comparing with Eq. (12.1) we get $2^{m-k} > 2^{-1}$, concluding that $m > k - 1$ or $m \geqslant k$. In other words, we get at most as many copies of ϕ as we started with. Since $m = \lfloor \alpha n \rfloor$ and $k = \lfloor \beta n \rfloor$, $\alpha \geqslant \beta$ and thus $E_C(\rho_{AB}) \geqslant E_D(\rho_{AB})$. □

12.2 For pure states, distillation and creation cost the same

Since Definitions 12.1 and 12.2 and Lemma 12.3 apply to general mixed states ρ_{AB}, it is tempting to ask if $E_C(\rho_{AB})$ and $E_D(\rho_{AB})$ are actually equal for any mixed state ρ_{AB}? Surprisingly, the answer is "No!" – there are mixed states for which $E_C(\rho_{AB}) > E_D(\rho_{AB})$. 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. Instead, we will fully resolve the case of pure states. Here the situation is much more simple since the distillable entanglement and entanglement cost are the same, and equal to entanglement entropy (see Definition 9.7).

Theorem 12.4. *For any* pure *state* $\rho_{AB} = |u\rangle\langle u|_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$,

$$E_D(\rho_{AB}) = H(\rho_A) = H(\rho_B) = E_C(\rho_{AB}).$$

Proof. We already know that $E_D(\rho_{AB}) \leqslant E_C(\rho_{AB})$, and we know from Schmidt decomposition that $H(\rho_A) = H(\rho_B) = H(p)$, where $\sqrt{p}$ are the Schmidt coefficients of $|u_{AB}\rangle$:

$$|u_{AB}\rangle = \sum_{x \in \Sigma} \sqrt{p(x)}\, |a_x\rangle_A \otimes |b_x\rangle_B.$$

Our strategy will be to show that $E_C(\rho_{AB}) \leqslant H(p) \leqslant E_D(\rho_{AB})$. We divide the rest of the proof into two parts that show these two inequalities.

Let us first show that $E_C(\rho_{AB}) \leqslant H(p)$. Recall from Definition 5.8 and Lemma 5.9 that, for any $n \geqslant 1$ and $\varepsilon > 0$, the set $T_{n,\varepsilon}(p)$ of ε-typical strings with respect to p consists of those $x_1 \cdots x_n \in \Sigma^n$ for which

$$2^{-n(H(p)+\varepsilon)} < p(x_1) \cdots p(x_n) < 2^{-n(H(p)-\varepsilon)}.$$

Define a vector

$$|v_{n,\varepsilon}\rangle := \sum_{x_1 \cdots x_n \in T_{n,\varepsilon}(p)} \sqrt{p(x_1) \cdots p(x_n)}\, \big(|a_1\rangle \otimes \cdots \otimes |a_n\rangle\big) \otimes \big(|b_1\rangle \otimes \cdots \otimes |b_n\rangle\big)$$

and note that

$$p_{n,\varepsilon} := \big\||v_{n,\varepsilon}\rangle\big\|^2 = \sum_{x_1 \cdots x_n \in T_{n,\varepsilon}(p)} p(x_1) \cdots p(x_n) = \Pr\big(X^n \in T_{n,\varepsilon}(p)\big),$$

where X is a random variable on Σ distributed according to p. Recall from the AEP Lemma 5.9 that this probability is close to 1:

$$p_{n,\varepsilon} \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}, \tag{12.2}$$

where σ is a constant that depends only on p. Just like in the proof of Shannon's source coding Theorem 5.7, this probability goes to 1 as $n \to \infty$, so the state $|v_{n,\varepsilon}\rangle$ is asymptotically normalized. However, for finite n, $p_{n,\varepsilon} \leqslant 1$, so let

$$|w_{n,\varepsilon}\rangle := \frac{|v_{n,\varepsilon}\rangle}{\sqrt{p_{n,\varepsilon}}}$$

denote the normalized version of $|v_{n,\varepsilon}\rangle$.

The reduced state of $|w_{n,\varepsilon}\rangle$ on systems $A_1 \cdots A_n$ has eigenvalues of the form $p(x)/p_{n,\varepsilon}$, for some $x \in T_{n,\varepsilon}(p)$, so by the AEP Lemma 5.9 they satisfy

$$\frac{2^{-n(H(p)+\varepsilon)}}{p_{n,\varepsilon}} < \lambda_j\left(\mathrm{Tr}_{B_1\cdots B_n}\left[|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\right]\right) < \frac{2^{-n(H(p)-\varepsilon)}}{p_{n,\varepsilon}}, \tag{12.3}$$

for all $j = 1, \dots, |T_{n,\varepsilon}(p)|$, while the remaining eigenvalues are zero.

Let us now bound the entanglement cost $E_C(\rho_{AB})$ of $|u_{AB}\rangle$. Recall that $E_C(\rho_{AB})$ is the infimum over all $\alpha \geqslant 0$ such that $\phi^{\otimes \lfloor \alpha n \rfloor}$ can be approximately converted to $\rho^{\otimes n}$ by LOCC. Take any $\alpha > H(p)$ and let $\varepsilon > 0$ be sufficiently small so that $\alpha > H(p) + 2\varepsilon$. Let $n > 1/\varepsilon$ so that $n\varepsilon > 1$. Then

$$m := \lfloor \alpha n \rfloor \geqslant \lfloor n(H(p)+\varepsilon) + n\varepsilon \rfloor > n(H(p)+\varepsilon). \tag{12.4}$$

We want to create as many copies of $|u_{AB}\rangle$ as possible from m copies of $\phi_{CD} = |\Phi^+_{CD}\rangle\langle\Phi^+_{CD}|$. Since the reduced state of $\phi^{\otimes m}$ is maximally mixed,

$$\lambda_j\left(\mathrm{Tr}_{D_1\cdots D_m}[\phi^{\otimes m}]\right) = 2^{-m}, \tag{12.5}$$

for $j = 1, \dots, 2^m$. Note from Eq. (12.4) that

$$2^{-m} \leqslant 2^{-n(H(p)+\varepsilon)} \leqslant \frac{2^{-n(H(p)+\varepsilon)}}{p_{n,\varepsilon}}$$

since $p_{n,\varepsilon} \leqslant 1$. Combining this with Eqs. (12.3) and (12.5),

$$\lambda_j\left(\mathrm{Tr}_{D_1\cdots D_m}[\phi^{\otimes m}]\right) < \lambda_j\left(\mathrm{Tr}_{B_1\cdots B_n}\left[|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\right]\right),$$

for all $j = 1, \dots, |T_{n,\varepsilon}(p)|$. Hence, we get the following majorization relation:

$$\sum_{j=1}^{k} \lambda_j\left(\mathrm{Tr}_{D_1\cdots D_m}[\phi^{\otimes m}]\right) \leqslant \sum_{j=1}^{k} \lambda_j\left(\mathrm{Tr}_{B_1\cdots B_n}\left[|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\right]\right),$$

for all $k = 1, \dots, 2^m$, where the right-hand side is equal to 1 for any $k \geqslant |T_{n,\varepsilon}(p)|$. By Nielsen's Theorem 11.9, there exists an LOCC channel Φ_n such that

$$\Phi_n(\phi^{\otimes m}) = |w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|,$$

with the conversion being exact.

While the resulting state $|w_{n,\varepsilon}\rangle$ is not exactly the same as the desired target state $|u\rangle^{\otimes n}$, they are sufficiently close. Indeed, the fidelity between the two states is

$$\begin{aligned} F\big(|u\rangle\langle u|^{\otimes n}, |w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\big)^2 = \big|\langle u|^{\otimes n}|w_{n,\varepsilon}\rangle\big|^2 &= \frac{1}{p_{n,\varepsilon}}\big|\langle u|^{\otimes n}|v_{n,\varepsilon}\rangle\big|^2 \\ &= \frac{1}{p_{n,\varepsilon}}\left(\sum_{x_1\cdots x_n\in T_{n,\varepsilon}(p)} p(x_1)\cdots p(x_n)\right)^2 \\ &= \frac{p_{n,\varepsilon}^2}{p_{n,\varepsilon}} = p_{n,\varepsilon}. \end{aligned} \tag{12.6}$$

Recall from Eq. (12.2) that we can make this arbitrarily close to 1 by choosing n large enough. In particular, we can achieve squared fidelity larger than $1-\delta$ by choosing $n > \max\{\frac{1}{\varepsilon}, \frac{\sigma^2}{\varepsilon^2\delta}\}$, just like in the proof of Theorem 5.7.

The proof of the second inequality $H(p) \leqslant E_D(\rho_{AB})$ is very similar. Let $\alpha < H(p)$ and $\varepsilon \in (0,1)$ be small enough so that $\alpha < H(p) - 2\varepsilon$. Let $n \geqslant -\frac{1}{\varepsilon}\log(1-\varepsilon)$ so that $-n\varepsilon \leqslant \log(1-\varepsilon)$. Then

$$m := \lfloor \alpha n \rfloor \leqslant n(H(p)-\varepsilon) - n\varepsilon \leqslant n(H(p)-\varepsilon) + \log(1-\varepsilon), \tag{12.7}$$

so

$$2^{-m} \geqslant 2^{-n(H(p)-\varepsilon)-\log(1-\varepsilon)} = \frac{2^{-n(H(p)-\varepsilon)}}{1-\varepsilon}.$$

Since $p_{n,\varepsilon} \to 1$ as $n \to \infty$,

$$2^{-m} \geqslant \frac{2^{-n(H(p)-\varepsilon)}}{p_{n,\varepsilon}},$$

for all sufficiently large n.

Recall from Eqs. (12.3) and (12.5) that the eigenvalues of the reduced states satisfy

$$\lambda_j\left(\mathrm{Tr}_{B_1\cdots B_n}\left[|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\right]\right) < \frac{2^{-n(H(p)-\varepsilon)}}{p_{n,\varepsilon}} \leqslant 2^{-m} = \lambda_j\left(\mathrm{Tr}_{D_1\cdots D_m}[\phi^{\otimes m}]\right),$$

for all $j = 1, \ldots, 2^m$ (for distillation we are in the regime $|T_{n,\varepsilon}(p)| \geqslant 2^m$). This implies the majorization relation

$$\sum_{j=1}^{k} \lambda_j\left(\mathrm{Tr}_{B_1\cdots B_n}\left[|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\right]\right) \leqslant \sum_{j=1}^{k} \lambda_j\left(\mathrm{Tr}_{D_1\cdots D_m}[\phi^{\otimes m}]\right)$$

for all k, where the right-hand side hits 1 at $k = 2^m$. By Nielsen's Theorem 11.9, there exists an LOCC channel Ψ_n such that

$$\Psi_n(|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|) = \phi^{\otimes m}.$$

By monotonicity of fidelity under Ψ_n (see Homework Problem 4.1),

$$\begin{aligned} F\big(\Psi_n(|u\rangle\langle u|^{\otimes n}), \phi^{\otimes m}\big)^2 &= F\Big(\Psi_n(|u\rangle\langle u|^{\otimes n}), \Psi_n(|w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|)\Big)^2 \\ &\geqslant F\big(|u\rangle\langle u|^{\otimes n}, |w_{n,\varepsilon}\rangle\langle w_{n,\varepsilon}|\big)^2 \\ &= p_{n,\varepsilon}, \end{aligned}$$

where the last equality follows from Eq. (12.6). Since this goes to 1 as $n \to \infty$, we have proved that $E_D(\rho_{AB}) \geqslant H(p)$. □

Lecture 13

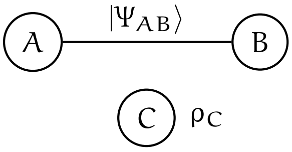
Monogamy of entanglement

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 the given state by LOCC. In Theorem 12.4, 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. Instead of the usual bipartite setting, we will extend it to include 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 would violate monogamy.

13.1 Sharing classical vs quantum correlations

You showed in Homework Problem 2.3 (d) 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*.

$\rho_{ABC} =$ 

In contrast, a classical state that is maximally correlation can be shared with an arbitrary number of parties. To see this, consider the following classical tripartite state

$$\rho_{ABC} = \frac{1}{2}(|000\rangle\langle 000| + |111\rangle\langle 111|), \tag{13.1}$$

which corresponds to flipping an unbiased coin and telling the outcome to all three parties. Note that any two parties are maximally correlated since

$$\rho_{AB} = \rho_{BC} = \rho_{AC} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|). \tag{13.2}$$

Moreover, one can easily distribute this correlation even further by attaching a fresh qubit in state $|0\rangle$ and performing a CNOT operation with this qubit as a target.

Why does this not work in the quantum case? Let's see what happens if we try to use the same construction. 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 should be $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Written as a density matrix,

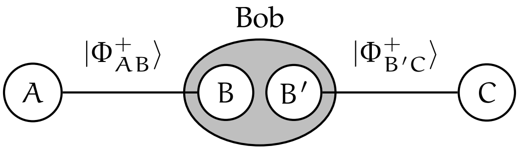
$$|\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|).$$

However, the actual reduced states are exactly the same as in Eq. (13.2):

$$\rho_{AB} = \rho_{BC} = \rho_{AC} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|),$$

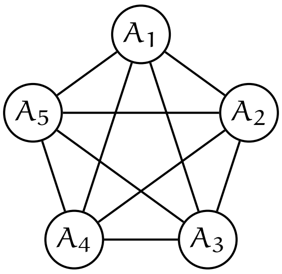
meaning that each pair of parties shares a maximal *classical* correlation, not a maximally entangled state. This is a very stark manifestation of monogamy of entanglement – we were hoping for all pairs of parties to be a maximally entangled, while in reality each pair shares a separable state that has no entanglement whatsoever!

The only way to actually share a maximally entangled state with two parties is by increasing the dimension. For example, if Bob has two qubits B and B′, he can share a maximally entangled with A and C as follows:



However, the system B is completely uncorrelated with C, and B′ is completely uncorrelated with A. Even if the dimension of system A is increased, Bob cannot share more entanglement with it while still maintaining a maximally entangled state with C.

Let us now consider a more complicated situation with n parties denoted by $A_1, \dots, A_n$. Assume their joint state $\rho_{A_1 \cdots A_n}$ 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}.



Intuitively, ρ_{AB} should not be too entangled because each party 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 symmetric n-party extension. Results of this form are known as *de Finetti* theorems.

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 $n \geqslant 1$, 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:*
>
> $$R_\pi\big(|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle\big) := |\psi_{\pi^{-1}(1)}\rangle \otimes \cdots \otimes |\psi_{\pi^{-1}(n)}\rangle, \tag{13.3}$$
>
> *for all $|\psi_1\rangle, \dots, |\psi_n\rangle \in \mathcal{H}$. The* symmetric subspace *of $\mathcal{H}^{\otimes n}$ is then defined as*
>
> $$\mathrm{Sym}^n(\mathcal{H}) := \big\{|\Phi\rangle \in \mathcal{H}^{\otimes n} : R_\pi|\Phi\rangle = |\Phi\rangle, \forall \pi \in S_n\big\}.$$

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 any $\pi, \tau \in S_n$. You will show this and $R_\pi^\dagger = R_{\pi^{-1}}$, for any $\pi \in S_n$, in Practice Problem 13.1 (c), which makes the map $\pi \mapsto R_\pi$ a unitary representation of the symmetric group S_n.*

Remark 13.3. *Make sure to not confuse R_π with the permutation matrix V_π from Theorem 11.2! The distinction is that R_π permutes the systems while V_π permutes the standard basis states according to π. In particular, R_π is of size $d^n \times d^n$, where $\dim \mathcal{H} = d$, while V_π is of size $n \times n$.*

Here are some basic observations about $\mathrm{Sym}^n(\mathcal{H})$:

- For any $|\psi\rangle \in \mathcal{H}$, $|\psi\rangle^{\otimes n} \in \mathrm{Sym}^n(\mathcal{H})$ since $R_\pi|\psi\rangle^{\otimes n} = |\psi\rangle^{\otimes n}$.
- For any $\pi \in S_n$, $|\Phi\rangle \in \mathrm{Sym}^n(\mathcal{H})$ iff $R_\pi|\Phi\rangle \in \mathrm{Sym}^n(\mathcal{H})$.
- If $|\Phi_1\rangle, |\Phi_2\rangle \in \mathrm{Sym}^n(\mathcal{H})$ then $|\Phi_1\rangle + |\Phi_2\rangle \in \mathrm{Sym}^n(\mathcal{H})$.

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). *The case of two qubits corresponds to $n = 2$ and $d = 2$. In this case,*
>
> $$\mathrm{Sym}^2(\mathbb{C}^2) = \mathrm{span}\left\{|0,0\rangle, |1,1\rangle, \frac{|0,1\rangle + |1,0\rangle}{\sqrt{2}}\right\}.$$
>
> *The remaining vector $(|0,1\rangle - |1,0\rangle)/\sqrt{2}$ (also known as the* singlet state*) is anti-symmetric.*

Since $\mathrm{Sym}^n(\mathcal{H})$ is a subspace of $\mathcal{H}^{\otimes n}$, we can write down a projector onto this subspace. You will show in Practice Problem 13.1 (d) that the following is an orthogonal projection onto $\mathrm{Sym}^n(\mathcal{H})$:

$$\Pi_n := \frac{1}{n!} \sum_{\pi \in S_n} R_\pi. \tag{13.4}$$

In particular, $\Pi_n^\dagger = \Pi_n$ and $\Pi_n^2 = \Pi_n$, i.e., Π_n is Hermitian and a projector. 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 Practice Problem 13.1 (d) you will show that $\Pi_n|\Phi\rangle \in \mathrm{Sym}^n(\mathcal{H})$, for any $|\Phi\rangle \in \mathcal{H}^{\otimes n}$. Moreover, if $|\Phi\rangle \in \mathrm{Sym}^n(\mathcal{H})$ then $\Pi_n|\Phi\rangle = |\Phi\rangle$. In fact, for any $k \geqslant 0$ and $|\Phi\rangle \in \mathrm{Sym}^{k+n}$, the following more general identity holds:

$$(I_k \otimes \Pi_n)|\Phi\rangle = |\Phi\rangle, \tag{13.5}$$

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*

$$\Pi_2 = \frac{1}{2}(I + F),$$

where $F \in U(\mathcal{H}^{\otimes 2})$ 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 $\mathrm{Sym}^n(\mathbb{C}^d)$, for any $n \geqslant 1$ and $d \geqslant 1$? We can symmetrize the standard basis states, thus projecting them to the symmetric subspace.

Let $\Lambda_{n,d}$ denote the set of all integers $t_1, \dots, t_d \geqslant 0$ such that $\sum_{i=1}^d t_i = n$:

$$\Lambda_{n,d} := \left\{(t_1, \dots, t_d) \in \mathbb{Z}^d : t_1, \dots, t_d \geqslant 0, \sum_{i=1}^{d} t_i = n\right\}.$$

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:

$$|T_{t_1,\dots,t_d}\rangle := \underbrace{\overbrace{|1\rangle \otimes \cdots \otimes |1\rangle}^{t_1} \otimes \overbrace{|2\rangle \otimes \cdots \otimes |2\rangle}^{t_2} \otimes \cdots \otimes \overbrace{|d\rangle \otimes \cdots \otimes |d\rangle}^{t_d}}_{n}, \tag{13.6}$$

where t_i denotes the number of terms $|i\rangle$ occurring in the tensor product.

Example 13.6 (Two qubits). *The set $\Lambda_{d,n}$ and the corresponding basis states for two qubits are*

$$\Lambda_{2,2} = \{(2,0), (1,1), (0,2)\}, \qquad |T_{2,0}\rangle = |0,0\rangle, \qquad |T_{1,1}\rangle = |0,1\rangle, \qquad |T_{0,2}\rangle = |1,1\rangle.$$

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 $\mathrm{Sym}^n(\mathbb{C}^d)$:*

$$\mathrm{Sym}^n(\mathbb{C}^d) = \mathrm{span}\{\Pi_n|T_{t_1,\dots,t_d}\rangle : (t_1, \dots, t_d) \in \Lambda_{n,d}\}.$$

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 of $(\mathbb{C}^d)^{\otimes n}$. Since $\Pi_n R_\pi|\Phi\rangle = \Pi_n|\Phi\rangle$, for all $\pi \in S_n$ and $|\Phi\rangle \in (\mathbb{C}^d)^{\otimes n}$, we can first permute the systems and sort the basis vectors to obtain one of the states $|T_{t_1,\dots,t_d}\rangle$. You can think of the resulting sequence $t_1, \dots, t_d$ as a "generalized Hamming weight" of the original string of basis vectors, since t_i counts the number of appearances of $|i\rangle$. To obtain a basis of $\mathrm{Sym}^n(\mathbb{C}^d)$, 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. □

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.*

You will work out some examples in Practice Problem 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*

$$\dim\big(\mathrm{Sym}^n(\mathbb{C}^d)\big) = |\Lambda_{n,d}| = \binom{n+d-1}{n} = \frac{(n+d-1)!}{n!(d-1)!}$$

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 $\mathrm{Sym}^n(\mathbb{C}^d)$ 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 rest be stars, yielding the desired binomial coefficient. □

We will need the following result, which we state without proof.

Lemma 13.10. *Let $A \in L(\mathcal{H}^{\otimes n})$ for some Hilbert space $\mathcal{H}$ and $n \geqslant 1$. Then*

$$U^{\otimes n} A U^{\dagger \otimes n} = A, \qquad \forall U \in U(\mathcal{H}),$$

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 over pure quantum states.

Lemma 13.11. *For any $n \geqslant 1$ and $d \geqslant 2$,*

$$\Pi_n = \binom{n+d-1}{n} \int d\psi \, (|\psi\rangle\langle\psi|)^{\otimes n},$$

where $d\psi$ is the uniform probability measure on pure states in $\mathbb{C}^d$.

Proof. You will prove this in Practice Problem 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 taking $d = 2$ *and explicitly evaluating the integral from Lemma 13.11, one can check for* $n = 2$ *that*

$$\binom{n+d-1}{n}\frac{1}{4\pi}\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\rho(\theta,\varphi)^{\otimes n}\sin\theta\, d\theta\, d\varphi = \frac{1}{2}\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} = \Pi_2.$$

13.3 The quantum de Finetti theorem

If $|\Phi_{A_1\cdots A_n}\rangle \in \mathrm{Sym}^n(\mathcal{H})$, then all its two-party reduced density matrices $\Phi_{A_iA_j}$ for $i \neq j$ are identical. If a given mixed state $\rho_{AA'}$ can be extended to such symmetric pure state $|\Phi_{A_1\cdots A_n}\rangle$, for some large value of n, then $\rho_{AA'}$ must be very close to separable (in fact, the distance goes to zero as $n \to \infty$). This is made rigorous by the quantum de Finetti theorem.

Theorem 13.13 (Quantum de Finetti theorem)**.** *Let* $k \geqslant 1$, $n \geqslant 0$, *and consider* $k+n$ *systems* $A_1 \cdots A_{k+n}$, *each of dimension* $d \geqslant 2$. *For any* $|\Phi\rangle \in \mathrm{Sym}^{k+n}(\mathbb{C}^d)$, *there exists a probability density function* p *on pure states in* $\mathbb{C}^d$ *such that*

$$\frac{1}{2}\left\| \Phi_{A_1\cdots A_k} - \int d\psi\, p(\psi)\, (|\psi\rangle\langle\psi|)^{\otimes k} \right\|_1 \leqslant \sqrt{\frac{dk}{k+n}}, \tag{13.7}$$

where $\Phi_{A_1\cdots A_k} = \mathrm{Tr}_{A_{k+1}\cdots A_{k+n}}\big[|\Phi\rangle\langle\Phi|\big]$.

Proof. Recall from Eq. (13.5) that $(I_{A_1\cdots A_k} \otimes \Pi_n)|\Phi\rangle = |\Phi\rangle$, so

$$\begin{aligned}
\Phi_{A_1\cdots A_k} &= \mathrm{Tr}_{A_{k+1}\cdots A_{k+n}}\big[|\Phi\rangle\langle\Phi|\big] \\
&= \mathrm{Tr}_{A_{k+1}\cdots A_{k+n}}\big[(I_{A_1\cdots A_k} \otimes \Pi_n)|\Phi\rangle\langle\Phi|\big] \\
&= \binom{n+d-1}{n}\int d\psi\; \mathrm{Tr}_{A_{k+1}\cdots A_{k+n}}\Big[\Big(I_{A_1\cdots A_k} \otimes \big(|\psi\rangle^{\otimes n}\langle\psi|^{\otimes n}\big)_{A_{k+1}\cdots A_{k+n}}\Big)|\Phi\rangle\langle\Phi|\Big] \\
&= \binom{n+d-1}{n}\int d\psi\, \big(I_{A_1\cdots A_k} \otimes \langle\psi|^{\otimes n}\big)|\Phi\rangle\langle\Phi|\big(I_{A_1\cdots A_k} \otimes |\psi\rangle^{\otimes n}\big),
\end{aligned}$$

where we substituted the integral formula for Π_n from Lemma 13.11 and then used the following cyclic property of the partial trace:

$$\mathrm{Tr}_B\Big[\big(I_A \otimes |\psi\rangle\langle\psi|_B\big)\Phi_{AB}\Big] = \big(I_A \otimes \langle\psi|_B\big)\Phi_{AB}\big(I_A \otimes |\psi\rangle_B\big), \tag{13.8}$$

which holds for any $\Phi_{AB} \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$ and unit vector $|\psi\rangle_B \in \mathcal{H}_B$. To prove this identity, first choose $U_B \in U(\mathcal{H}_B)$ such that $U_B|j\rangle_B = |\psi\rangle_B$, where $|j\rangle_B$ is an arbitrary standard basis vector of

$\mathcal{H}_B$, and then use the unitary invariance of the partial trace:

$$\begin{aligned}\operatorname{Tr}_B\Big[(I_A\otimes|\psi\rangle\langle\psi|_B)\Phi_{AB}\Big] &= \operatorname{Tr}_B\Big[(I_A\otimes U_B)(I_A\otimes|j\rangle\langle j|_B)(I_A\otimes U_B^\dagger)\Phi_{AB}\Big]\\ &= \operatorname{Tr}_B\Big[(I_A\otimes|j\rangle\langle j|_B)(I_A\otimes U_B^\dagger)\Phi_{AB}(I_A\otimes U_B)\Big]\\ &= \operatorname{Tr}_B\Big[(I_A\otimes|j\rangle\langle j|_B)\hat{\Phi}_{AB}\Big],\end{aligned}$$

where $\hat{\Phi}_{AB} := (I_A\otimes U_B^\dagger)\Phi_{AB}(I_A\otimes U_B)$. Then we simply evaluate the partial trace according to Definition 2.7:

$$\begin{aligned}\operatorname{Tr}_B\Big[(I_A\otimes|j\rangle\langle j|_B)\hat{\Phi}_{AB}\Big] &= \sum_i (I_A\otimes\langle i|_B)(I_A\otimes|j\rangle\langle j|_B)\hat{\Phi}_{AB}(I_A\otimes|i\rangle_B)\\ &= \sum_i (I_A\otimes\langle i|j\rangle\langle j|_B)\hat{\Phi}_{AB}(I_A\otimes|i\rangle_B)\\ &= (I_A\otimes\langle j|_B)\hat{\Phi}_{AB}(I_A\otimes|j\rangle_B),\end{aligned}$$

After substituting back $\hat{\Phi}_{AB}$ and using $U_B|j\rangle_B$, we recover Eq. (13.8).

Let us continue with the proof and rewrite the integral as follows:

$$\Phi_{A_1\cdots A_k} = \int d\psi\, p(\psi)\,|\Phi_\psi\rangle\langle\Phi_\psi|, \tag{13.9}$$

where $|\Phi_\psi\rangle_{A_1\cdots A_k}\in(\mathbb{C}^d)^{\otimes k}$ and p are such that

$$\sqrt{p(\psi)}\,|\Phi_\psi\rangle := \sqrt{\binom{n+d-1}{n}}\,(I_{A_1\cdots A_k}\otimes\langle\psi|^{\otimes n})|\Phi\rangle. \tag{13.10}$$

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. (13.9). More specifically, $p(\psi)$ is given by

$$p(\psi) := \binom{n+d-1}{n}\left\|(I_{A_1\cdots A_k}\otimes\langle\psi|^{\otimes n})|\Phi\rangle\right\|^2.$$

Let us compare the integral in Eq. (13.9) with $\tilde{\Phi}_{A_1\cdots A_k} = \int d\psi\, p(\psi)\,|\psi\rangle^{\otimes k}\langle\psi|^{\otimes k}$, where p is the same probability density function. Using triangle inequality and the formula from Eq. (3.11) for the trace distance between pure states,

$$\begin{aligned}\frac{1}{2}\left\|\Phi_{A_1\cdots A_k}-\tilde{\Phi}_{A_1\cdots A_k}\right\|_1 &\leqslant \int d\psi\, p(\psi)\,\frac{1}{2}\left\||\Phi_\psi\rangle\langle\Phi_\psi| - |\psi\rangle^{\otimes k}\langle\psi|^{\otimes k}\right\|_1\\ &= \int d\psi\, p(\psi)\,\sqrt{1-\left|\langle\psi|^{\otimes k}|\Phi_\psi\rangle\right|^2}\\ &\leqslant \sqrt{\int d\psi\, p(\psi)\left(1-\left|\langle\psi|^{\otimes k}|\Phi_\psi\rangle\right|^2\right)}\\ &= \sqrt{1-\int d\psi\, p(\psi)\left|\langle\psi|^{\otimes k}|\Phi_\psi\rangle\right|^2},\end{aligned}$$

where we used Jensen's inequality from Eq. (5.4) 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. (13.10) that

$$\begin{aligned}\sqrt{p(\psi)}\,\langle\psi|^{\otimes k}|\Phi_\psi\rangle &= \sqrt{\binom{n+d-1}{n}}\,\big(\langle\psi|^{\otimes k}\otimes\langle\psi|^{\otimes n}\big)|\Phi\rangle\\ &= \sqrt{\binom{n+d-1}{n}}\,\langle\psi|^{\otimes k+n}|\Phi\rangle.\end{aligned}$$

Hence,

$$\begin{aligned}\int d\psi\, p(\psi)\,\big|\langle\psi|^{\otimes k}|\Phi_\psi\rangle\big|^2 &= \binom{n+d-1}{n}\int d\psi\,\langle\Phi|\big(|\psi\rangle\langle\psi|^{\otimes k+n}\big)|\Phi\rangle\\ &= \binom{n+d-1}{n}\binom{k+n+d-1}{k+n}^{-1}\int d\psi\,\langle\Phi|\Pi_{k+n}|\Phi\rangle\\ &= \binom{n+d-1}{n}\binom{k+n+d-1}{k+n}^{-1},\end{aligned}$$

where we used the integral formula from Lemma 13.11 and the assumption $|\Phi\rangle\in\mathrm{Sym}^{k+n}(\mathbb{C}^d)$ which implies that $\Pi_{k+n}|\Phi\rangle=|\Phi\rangle$.

The ratio of the two binomial coefficients can be expressed as follows:

$$\begin{aligned}\binom{n+d-1}{n}\binom{k+n+d-1}{k+n}^{-1} &= \frac{(n+d-1)!}{n!\cancel{(d-1)!}}\cdot\frac{(k+n)!\cancel{(d-1)!}}{(k+n+d-1)!}\\ &= \frac{(n+d-1)!}{n!}\cdot\frac{(k+n)!}{(k+n+d-1)!}\\ &= \frac{(n+d-1)\cdots(n+1)}{(k+n+d-1)\cdots(k+n+1)}.\end{aligned}$$

Note that $\frac{a+1}{b+1}-\frac{a}{b}=\frac{b-a}{b(b+1)}\geqslant 0$ when $b\geqslant a$, so $\frac{a+1}{b+1}\geqslant\frac{a}{b}$ and hence

$$\begin{aligned}\frac{n+d-1}{k+n+d-1}\cdots\frac{n+1}{k+n+1} &= \left(\frac{n+1}{k+n+1}\right)^{d-1}\\ &\geqslant\left(1-\frac{k}{k+n+1}\right)^{d-1}\\ &\geqslant 1-(d-1)\frac{k}{k+n+1}\\ &\geqslant 1-\frac{dk}{k+n},\end{aligned}$$

where we used $(1-\alpha)^x\geqslant 1-\alpha x$ for $\alpha\in(0,1)$ and $x\geqslant 1$. Putting everything together,

$$\begin{aligned}\frac{1}{2}\big\|\Phi_{A_1\cdots A_k}-\tilde{\Phi}_{A_1\cdots A_k}\big\|_1 &= \sqrt{1-\int d\psi\,p(\psi)\,\big|\langle\psi|^{\otimes k}|\Phi_\psi\rangle\big|^2}\\ &= \sqrt{1-\binom{n+d-1}{n}\binom{k+n+d-1}{k+n}^{-1}}\\ &\leqslant\sqrt{\frac{dk}{n+k}},\end{aligned}$$

which is the desired bound. □

Lecture 14

Quantum state merging

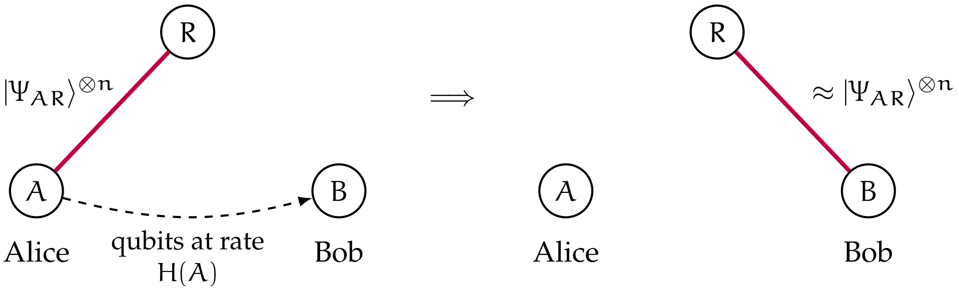
Last week, we introduced the notion of monogamy of entanglement, which can be stated informally as the fact that you cannot simultaneously share a lot of entanglement with many other people. To make this more formal, we defined the symmetric subspace that contains multipartite pure states that are invariant under any permutation of parties. These states have a very high degree of symmetry – in particular, all k-party reduced states are identical. Our main result was the quantum de Finetti theorem (Theorem 13.13), which shows that the k-party reduced state of a symmetric state is close to separable.

In this lecture, we will look at entanglement from yet another perspective. We will revisit Schumacher compression and entanglement distillation from Lectures 6 and 12, respectively, and consider a more general problem called *quantum state merging* that generalizes both.

14.1 Special cases of quantum state merging

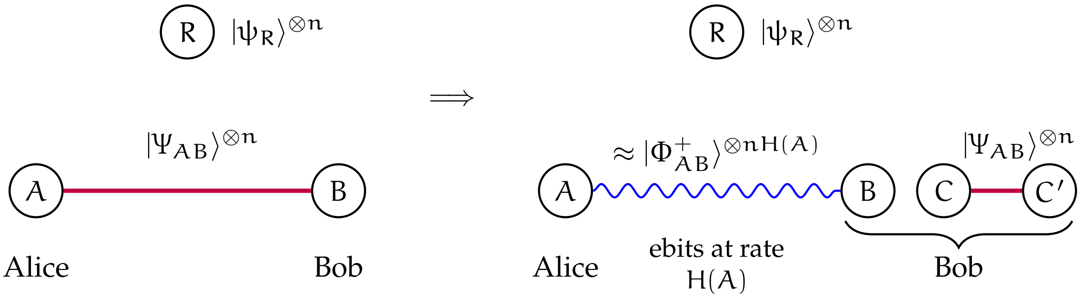
The general quantum state merging problem is concerned with a pure tripartite state $|\Psi_{ABR}\rangle$, where A and B are held by Alice and Bob, respectively, and R is a reference system that is not accessible to them. We assume that the reduced state Ψ_{AB} is known to Alice and Bob so that they are able to manipulate it. Before we discuss the general problem, let us consider two special cases you are already familiar with – Schumacher compression and entanglement distillation. They correspond to state merging for states of the form $|\Psi_{AR}\rangle \otimes |\psi_B\rangle$ and $|\Psi_{AB}\rangle \otimes |\psi_R\rangle$, respectively.

Recall from Theorem 6.8 that Schumacher compression can compress n copies of a mixed state ρ_A to roughly $nH(\rho_A)$ qubits. Just like in the classical case (see Lemma 6.3), quantum compression also has the property that it approximately preserves correlations with an external reference system R. Hence, if instead of ρ_A we compress the A subsystem of its purification $|\Psi_{AR}\rangle$, we recover the original correlations with R after decompressing the state. This scenario can be illustrated as follows:


R
$|\Psi_{AR}\rangle^{\otimes n}$
A
B
Alice
qubits at rate $H(A)$
Bob
$\Longrightarrow$
R
$\approx |\Psi_{AR}\rangle^{\otimes n}$
A
B
Alice
Bob

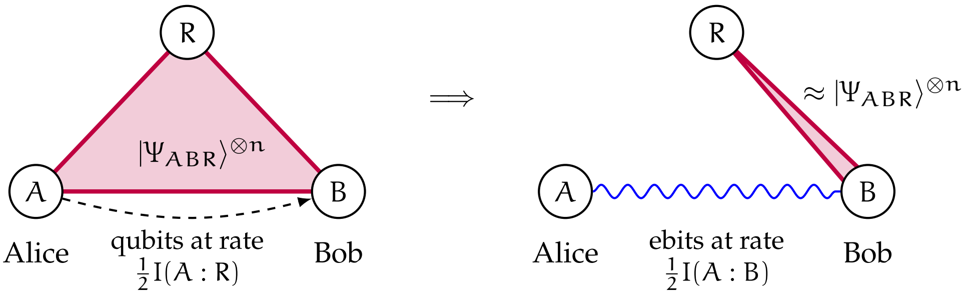
Using Schumacher's Theorem 6.8, Alice can compresses the registers $A_1 \cdots A_n$ of $|\Psi_{AR}\rangle^{\otimes n}$ and send them to Bob at rate $H(A) = H(\rho_A)$. After Bob decompresses, he approximately recovers the original state $|\Psi_{AR}\rangle^{\otimes n}$, which is still correlated with the reference system R. This is a special case of quantum state merging and is characterized by the property that the reduced state on AR is pure (i.e., B is in product with AR). Having no B register at all is a special case of this.

Another interesting special case is when there is no reference system R at all or, more generally, when the initial state factorizes as $|\Psi_{ABR}\rangle = |\Psi_{AB}\rangle \otimes |\psi_R\rangle$. In this case, we can ignore the reference system altogether and Bob can simply prepare n copies of the desired target state $|\Psi_{AB}\rangle$ all by himself on a new pair of registers CC'. Since he still shares the state $|\Psi_{AB}\rangle^{\otimes n}$ with Alice and they don't need it anymore, they might as well convert it by LOCC to as large number of maximally entangled states as possible and keep them for later use. According to Theorem 12.4, they can distill roughly $nH(A)$ copies of $|\Phi^+_{AB}\rangle$. The overall transformation in this case looks as follows:


R
$|\psi_R\rangle^{\otimes n}$
$|\Psi_{AB}\rangle^{\otimes n}$
A
B
Alice
Bob
$\Longrightarrow$
R
$|\psi_R\rangle^{\otimes n}$
$\approx |\Phi^+_{AB}\rangle^{\otimes nH(A)}$
$|\Psi_{AB}\rangle^{\otimes n}$
A
B
C
C′
Alice
ebits at rate
$H(A)$
Bob

14.2 General problem

Quantum stage merging generalizes these two problems to a case where all three parties holding the state $|\Psi_{ABR}\rangle$ are entangled in some non-trivial way. Given n copies of such state, the goal is for Alice to transmit some qubits to Bob so that he can recover the full state $|\Psi_{ABR}\rangle^{\otimes n}$, where systems A and B are now both on Bob's side and the state is still correlated to the reference system R the same way as before. Moreover, as the last example suggests, we should also account for the possibility that after achieving the desired transformation Alice and Bob may still be able to extract additional shared maximally entangled states which they can keep for later use. Here is an illustration of the general quantum state merging task:



Note that this task involves two rates:

- *quantum communication rate* – the rate at which Alice needs to send qubits to Bob,
- *distillation rate* – the rate at which Alice and Bob can produce entanglement by LOCC.

The values of these rates achieved by the quantum state merging protocol are $\frac{1}{2}I(A:R)$ and $\frac{1}{2}I(A:B)$, respectively. Note that quantum state merging in general cannot be achieved by LOCC since it requires quantum communication.

Before we discuss the communication and distillation rates, recall from Definition 7.2 that the quantum mutual information is defined as follows:

$$I(A:B) = H(A) + H(B) - H(AB).$$

Moreover, recall from Eq. (7.8) that

$$\frac{1}{2}I(A:B) \leqslant \min\{H(A), H(B)\},$$

with equality if and only if Ψ_{AB} is pure (this can be proved using the Araki-Lieb inequality from Eq. (7.4)). In particular,

$$\frac{1}{2}I(A:R) \leqslant H(R), \qquad\qquad \frac{1}{2}I(A:B) \leqslant H(B). \tag{14.1}$$

As a sanity check, let us verify that the desired communication and distillation rates do not contradict what we discussed in Section 14.1 for the special cases of Schumacher compression and entanglement distillation. Recall that in the two special cases the rates were as follows:

- *Schumacher compression*: $|\Psi_{AR}\rangle$ pure (or no B) – communication at rate $\frac{1}{2}I(A:R) = H(A)$,
- *entanglement distillation*: $|\Psi_{AB}\rangle$ pure (or no R) – distillation at rate $\frac{1}{2}I(A:B) = H(B)$.

This is consistent with the rates involved in state merging. Since Schumacher compression and (pure-state) entanglement distillation rates are optimal, the rates obtained through state merging cannot exceed them according to Eq. (14.1). In fact, state merging achieves optimal rates in these two special cases. Note that the communication rate will generally be lower than $H(A)$ since Bob already has some part of the state. Similarly, the distillation rate will also generally be lower than $H(B)$ since some of Bob's entanglement might be shared with R instead of Alice.

The general quantum state merging problem interpolates between Schumacher compression and entanglement distillation. While the rates achieved by state merging are not optimal in general, the quantum state merging protocol is fairly intuitive and it shines more light on bipartite mixed-state entanglement. In particular, it is curious to note that

$$\frac{I(A:R)}{2} + \frac{I(A:B)}{2} = \frac{1}{2}\big(H(A) + H(R) - H(AR) + H(A) + H(B) - H(AB)\big) = H(A)$$

since $H(R) = H(AB)$ and $H(B) = H(AR)$ as the global state is pure. Intuitively, the state merging protocol splits Alice's correlations into two kinds – ones she has with R (which need to be compressed and sent to Bob) and ones she has with Bob (which need to be distilled).

14.3 Perfect decoupling

Before we discuss the general quantum state merging protocol, let us consider another illustrative special case – the *perfectly decoupled* case. It illustrates the intuition behind the *decoupling* technique we will use in Section 14.4.

Assume that Alice's system is maximally entangled with Bob's. Since the overall state $|\Psi_{ABR}\rangle$ is pure, the reference system R cannot be entangled with Alice at all due to monogamy of entanglement. Moreover, to accommodate the other half of the maximally entangled state, Bob's system has to be at least as large as Alice's and it can be partitioned as $B = B'B''$, where $\dim B' = \dim A$ and B' is maximally entangled with Alice's system A while B'' is in product with A due to monogamy. In other words, by applying a suitable basis change on Bob's side, we can assume the overall state to be of the form

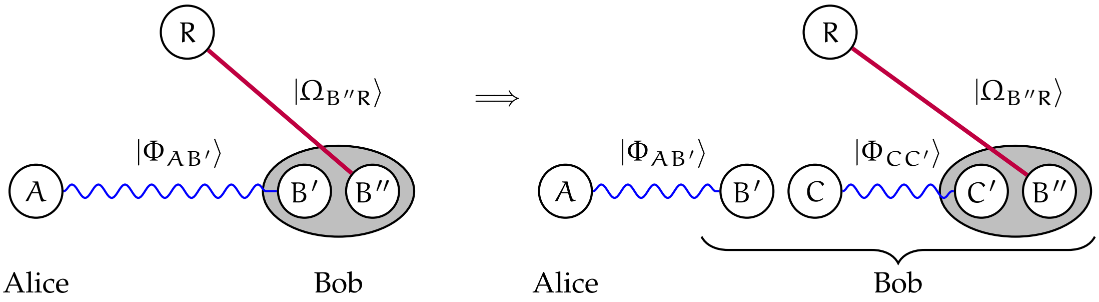
$$|\Psi_{ABR}\rangle = |\Phi_{AB'}\rangle \otimes |\Omega_{B''R}\rangle, \tag{14.2}$$

where $|\Phi_{AB'}\rangle$ is maximally entangled and $|\Omega_{B''R}\rangle$ is some remaining state that can still exhibit correlations with R. Equivalently, the reduced state on AR can be factorized as

$$\Psi_{AR} = \frac{I_A}{d_A} \otimes \Psi_R$$

where Alice's system is maximally mixed and Ψ_R is some arbitrary state on R.

To recover the desired state $|\Psi_{ABR}\rangle$ fully on Bob's side, he simply needs to create two fresh registers C and C' of the same dimension as A and prepare a local copy of the maximally entangled state $|\Phi_{CC'}\rangle$ in them:



Note that the final state is of the form

$$|\Phi_{AB'}\rangle \otimes |\Phi_{CC'}\rangle \otimes |\Omega_{B''R}\rangle,$$

where Bob holds all registers except A and R. In particular, note from Eq. (14.2) that the last two terms coincide with the desired target state once the registers C' and B'' are grouped together:

$$|\Phi_{CC'}\rangle \otimes |\Omega_{B''R}\rangle = |\Psi_{C,C'B'',R}\rangle.$$

In other words, Bob has managed to prepare the desired target state all by himself without Alice having to send him any qubits. Moreover, they even have a maximally entangled state $|\Phi_{AB'}\rangle$ left over that is not part of the target state and thus can be kept for later use. The reason we refer to this as the "perfectly decoupled" case is because Bob's system can be factorized to completely decouple A from R, and A is maximally entangled with the B' subsystem of B.

14.4 Decoupling

Appendix A

Practice Problems

Quantum Information Theory, Spring 2020

Practice problem set #1

You do **not** have to hand in these exercises, they are for your practice only.

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.

   (a) 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?

   i. $|\psi\rangle + \langle\phi|$
   ii. $|\psi\rangle\langle\phi|$
   iii. $A\langle\psi|$
   iv. $\langle\psi|A$
   v. $\langle\psi|A + \langle\psi|$
   vi. $|\psi\rangle\langle\phi| + A$
   vii. $|\psi\rangle\langle\phi|A$
   viii. $|\psi\rangle A\langle\phi|$
   ix. $\langle\psi|A|\phi\rangle$
   x. $\langle\psi|A|\phi\rangle + \langle\psi|\phi\rangle$
   xi. $\langle\psi|\phi\rangle\langle\psi|$
   xii. $\langle\psi|\phi\rangle A$

   (b) Let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$ be two pure states on the same system. Verify that

   $$\operatorname{Tr}[\rho\sigma] = |\langle\psi|\phi\rangle|^2.$$

   Hint: You may use that the trace is cyclic, i.e. $\operatorname{Tr}[ABC] = \operatorname{Tr}[CAB] = \operatorname{Tr}[BCA]$.

2. **Positive semidefinite operators:** Recall from class that an operator $A \in L(\mathcal{H})$ is called *positive semidefinite* if it is Hermitian and all its eigenvalues are nonnegative. We denote by $\mathrm{PSD}(\mathcal{H})$ the set of positive semidefinite operators on a Hilbert space $\mathcal{H}$. Argue that the following conditions are equivalent:

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

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

   (a) Show that $\mathrm{PSD}(\mathcal{H})$ is convex and closed under multiplication by $\mathbb{R}_{\geqslant 0}$ (i.e., a *convex cone*).
   (b) Show that $D(\mathcal{H})$ is convex.

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

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

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

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

(a) Show that $r_x = \mathrm{Tr}[\rho X]$, $r_y = \mathrm{Tr}[\rho Y]$, and $r_z = \mathrm{Tr}[\rho Z]$.

(b) Let σ be another qubit state, with Bloch vector $\vec{s}$. Verify that $\mathrm{Tr}[\rho\sigma] = \frac{1}{2}\left(1 + \vec{r} \cdot \vec{s}\right)$.

(c) Let $\{|\psi_i\rangle\}_{i=0,1}$ denote an orthonormal basis of $\mathbb{C}^2$, $\mu\colon \{0,1\} \to \mathrm{PSD}(\mathbb{C}^2)$ the corresponding basis measurement (i.e., $\mu(i) = |\psi_i\rangle\langle\psi_i|$ for $i \in \{0,1\}$), and $\vec{r}_i$ the Bloch vector of $|\psi_i\rangle\langle\psi_i|$. Show that the probability of obtaining outcome $i \in \{0,1\}$ when measuring ρ using μ is given by $\frac{1}{2}(1 + \vec{r} \cdot \vec{r}_i)$. Show that $\vec{r}_0 = -\vec{r}_1$. How can you visualize these two facts on the Bloch sphere?

(d) Now imagine that ρ is an unknown qubit state ρ whose Bloch vector $\vec{r}$ you would like to characterize completely. Consider the following measurement with six outcomes:

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

where $\sigma_x = X$, $\sigma_y = Y$, and $\sigma_z = Z$ are the three Pauli matrices. Show that μ is a valid measurement and that the probabilities of measurement outcomes are given by

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

How can you visualize this formula on the Bloch sphere? Describe how measuring many copies of ρ by using μ allows for estimating the entries of $\vec{r}$ to arbitrary accuracy.

Quantum Information Theory, Spring 2020

Practice problem set #2

You do **not** have to hand in these exercises, they are for your practice only.

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.

1. **Singular values and eigenvalues:** Recall that the *singular values* of an operator M are the square roots of the nonzero eigenvalues of $M^\dagger M$ or $MM^\dagger$ (which are always positive semidefinite).

   (a) Show that if M is Hermitian then its singular values are equal to its nonzero *absolute* eigenvalues. How about if the operator is positive semidefinite?

   (b) Argue that $\|M\|_1 = \mathrm{Tr}[M]$ if M is positive semidefinite. In particular, $\|\rho\|_1 = 1$ for any state.

2. **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. Compute the reduced states ρ_X and ρ_Y.

3. **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 $|\Psi_{AB}\rangle = \frac{1}{3}|0, 0\rangle + \frac{2}{3}|0, 1\rangle + \frac{2}{3}|1, 0\rangle$.

   Hint: If this calculation seems too painful to carry out, see the next problem.

4. **Schmidt decomposition:** Let $\rho_{AB} = |\Psi_{AB}\rangle\langle\Psi_{AB}|$ be an arbitrary pure state. Fix 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 $\rho_B = M^T\overline{M}$.

   (b) Let $M = \sum_i s_i|e_i\rangle\langle f_i|$ be a singular value decomposition. Show that $|\Psi_{AB}\rangle = \sum_i s_i|e_i\rangle \otimes |\overline{f_i}\rangle$ is a Schmidt decomposition.

   (c) Explain how to find a Schmidt decomposition of the following two-qubit pure state:

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

   Note: The transpose and the complex conjugate are computed with respect to the fixed bases.

5. **Partial trace trickery:** Verify the following calculational rules for the partial trace:

   (a) Let $X_A, Y_A \in L(\mathcal{H}_A)$ and $M_{AB} \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then,

   $$\mathrm{Tr}_B[(X_A \otimes I_B)M_{AB}(Y_A \otimes I_B)] = X_A\, \mathrm{Tr}_B[M_{AB}]Y_A.$$

   (b) Let $N_{ABC} \in L(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$. Then,

   $$\mathrm{Tr}_{AB}[N_{ABC}] = \mathrm{Tr}_A[\mathrm{Tr}_B[N_{ABC}]] = \mathrm{Tr}_B[\mathrm{Tr}_A[N_{ABC}]].$$

   What does this last rule look like if there is no C-system?

6. **Observables (for those of you who have taken a quantum mechanics course):** In this problem we discuss the relationship between 'measurements' as defined in the lecture and 'observables' as introduced in a first quantum mechanics course. An *observable* on a quantum system is by definition a Hermitian operator on the corresponding Hilbert space $\mathcal{H}$.

   (a) Let $\mu\colon \Omega \to \mathrm{PSD}(\mathcal{H})$ 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:

   $$\mathcal{O} = \sum_{\omega \in \Omega} \omega \, \mu(\omega) \tag{A.1}$$

   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. (A.1) 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 $\mathrm{Tr}[\rho\mathcal{O}]$.
   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?

Quantum Information Theory, Spring 2020

Practice problem set #3

You do **not** have to hand in these exercises, they are for your practice only.

1. **Positive semidefinite operators:** For all $Q \in \mathrm{PSD}(\mathcal{H})$, show that:

   (a) $A^\dagger Q A$ is positive semidefinite for all $A \in L(\mathcal{H})$.
   (b) If Q is invertible, its inverse Q^{-1} is again positive semidefinite.

   A positive semidefinite operator that is invertible is often called *positive definite*.

2. **Properties of the trace distance:** Show that the trace distance $T(\rho, \sigma) := \frac{1}{2}\|\rho - \sigma\|_1$ satisfies the following properties:

   (a) *Invariance:* $T(\rho, \sigma) = T(V\rho V^\dagger, V\sigma V^\dagger)$ for all states ρ, σ and any isometry V.
   (b) *Monotonicity:* $T(\rho_A, \sigma_A) \leqslant T(\rho_{AB}, \sigma_{AB})$ for all states ρ_{AB}, σ_{AB}.

   The fidelity $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1$ is likewise invariant under isometries (can you see why?). However, its monotonicity goes the opposite way (see the homework). Why is this intuitive?

3. **Quantum channels:** Show that the following maps Φ are quantum channels by directly verifying that they are trace-preserving and completely positive.

   (a) *Basis change:* $\Phi[M] = UMU^\dagger$ for a unitary U.
   (b) *Add state:* $\Phi[M_A] = M_A \otimes \sigma_B$ for a state σ_B.
   (c) *Partial trace:* $\Phi[M_{AB}] = \mathrm{Tr}_B[M_{AB}]$.
   (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. **Composing channels:** If $\Phi_{A\to B}, \Psi_{B\to C}$ are quantum channels, then so is $\Psi_{B\to C} \circ \Phi_{A\to B}$. If $\Phi_{A\to B}, \Xi_{C\to D}$ are quantum channels, then so is $\Phi_{A\to B} \otimes \Xi_{C\to D}$.

5. **Schmidt decomposition:** Let $\rho_A = \sum_{i=1}^r p_i |e_i\rangle\langle e_i|$ be an arbitrary eigendecomposition, where $p_1, \dots, p_r$ are the nonzero eigenvalues of ρ_A and the $|e_i\rangle$ corresponding 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) Conclude that any two purifications $|\Psi_{AB}\rangle$ and $|\Phi_{AB}\rangle$ are related by a unitary U_B – as claimed in Lecture 2.

   How about if we consider purifications on different Hilbert spaces?

Quantum Information Theory, Spring 2020

Practice problem set #4

You do **not** have to hand in these exercises, they are for your practice only.

1. **Functionals:** Let $\lambda\colon L(\mathcal{H}) \to \mathbb{C}$ be a linear function.

   (a) Show that there exists a unique $X \in L(\mathcal{H})$ such that $\lambda[M] = \operatorname{Tr}[X^\dagger M]$ for all $M \in L(\mathcal{H})$.
   (b) How about if $\lambda[M] \geqslant 0$ for all $M \geqslant 0$?

2. **Depolarizing and dephasing channels:** The *completely depolarizing channel* on $L(\mathcal{H})$ is given by

$$\mathcal{D}[M] = \operatorname{Tr}[M]\frac{I}{d} \qquad \forall M \in L(\mathcal{H}),$$

   where $d = \dim \mathcal{H}$. For $\mathcal{H} = \mathbb{C}^\Sigma$, the *completely dephasing channel* is defined by

$$\Delta[M] = \sum_x \langle x|M|x\rangle\, |x\rangle\langle x| \qquad \forall M \in L(\mathcal{H}).$$

   (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?

3. **Kraus and Stinespring:** Find Kraus and Stinespring representations for the following quantum channels:

   (a) *Trace:* $\Phi[M] = \operatorname{Tr}[M]$
   (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).

4. **Kraus and Stinespring:** Given Kraus or Stinespring representations of two channels $\Phi_{A\to B}$ and $\Psi_{B\to C}$, explain how to obtain the same representation for $\Psi_{B\to C} \circ \Phi_{A\to B}$.

5. **Stinespring with unitaries:** Use the Stinespring representation to prove that any quantum channel $\Phi_{A\to B}$ can be written in the following form:

$$\Phi_{A\to B}[M_A] = \operatorname{Tr}_E\left[U_{AC\to BE}(M_A \otimes \sigma_C)U^\dagger_{AC\to BE}\right] \qquad \forall M_A,$$

   where $\mathcal{H}_C$, $\mathcal{H}_E$ are auxiliary Hilbert spaces, $\sigma_C \in D(\mathcal{H}_E)$ is a *pure* state, and $U_{AC\to BE}$ a *unitary*.

6. **Adjoint superoperator:** Recall that the Hilbert-Schmidt inner product on $L(\mathcal{H})$ is given by $\langle M, N\rangle_{HS} = \operatorname{Tr}[M^\dagger N]$. This allows us to define the *adjoint* of a superoperator $\Phi \in L(L(\mathcal{H}_A), L(\mathcal{H}_B))$. Explicitly, this is the superoperator $\Phi^\dagger \in L(L(\mathcal{H}_B), L(\mathcal{H}_A))$ such that

$$\langle M_A, \Phi^\dagger[N_B]\rangle_{HS} = \langle \Phi[M_A], N_B\rangle_{HS} \qquad \forall M_A, N_B.$$

   (a) Given a Kraus representation of Φ, explain how to find one for $\Phi^\dagger$.
   (b) Show that Φ is completely positive if and only if $\Phi^\dagger$ is completely positive.
   (c) Show that Φ is trace-preserving if and only if $\Phi^\dagger$ is unital (i.e., $\Phi^\dagger[I_B] = I_A$).

Quantum Information Theory, Spring 2020

Practice problem set #5

You do **not** have to hand in these exercises, they are for your practice only.

1. **Fidelity and trace distance:** For any $\rho, \sigma \in D(\mathcal{H})$, prove that $T(\rho, \sigma) \leqslant \sqrt{1 - F(\rho, \sigma)^2}$. This is one of the Fuchs-van de Graaf inequalities.

   Hint: Use Uhlmann's theorem and the monotonicity property of the trace distance.

2. **Classical-quantum states:** Let $\mathcal{H}_X = \mathbb{C}^\Sigma$. We say that a state ρ_{XB} is *classical on subsystem* X or a *classical-quantum state* 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 $\rho_{B,x}$. By convention, we will always denote subsystems by $X, Y, \ldots$ if we know them to be classical (and $A, B, \ldots$ otherwise).

   (a) Discuss how this generalizes the notion of classical states.
   (b) 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.
   (c) Let $\Phi_{A \to X}$ be the channel corresponding to a measurement μ_A, as on the previous homework. Show that, for any system B and any state ρ_{AB}, the state $\rho_{XB} = (\Phi_{A \to X} \otimes \mathcal{I}_B)\,[\rho_{AB}]$ is a classical-quantum state and compute the probabilities $p(x)$ and the states $\rho_{B,x}$.

3. **How to compress it?** Suppose you would like 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?

4. **Lossy vs. lossless compression:** In class, we mostly discussed *lossy* compression protocols which compress any input sequence into a fixed number of bits but may fail with some small probability. In practice, it is also interesting to consider *lossless* compression protocols that use a variable number of bits (depending on the input sequence) and never fail.

   Given an (n, R, δ)-code, which achieves lossy compression, can you construct a lossless compression protocol with average rate $\approx R$ (for large n and small δ)?

5. **Lexicographic order (for the bonus problem):** The lexicographic order $\leqslant_{\text{lex}}$ on $\{0,1\}^n$ is defined as follows: Given bitstrings x^n and y^n, we let $x^n \leqslant_{\text{lex}} y^n$ 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` $\leqslant_{\text{lex}}$ `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* $1 \leqslant k \leqslant n - 1$.

Quantum Information Theory, Spring 2020

Practice problem set #6

You do **not** have to hand in these exercises, they are for your practice only.

1. **Trace distance of probability distributions:** In today's lecture, we defined the *(normalized) trace distance* between $p, q \in P(\Sigma)$ by $T(p, q) := \frac{1}{2} \sum_{x\in\Sigma} |p(x) - q(x)|$. This quantity is also known as the *total variation distance* between p and q.

   (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

   $$T(p, q) = \max_{S\subseteq\Sigma} \big(\Pr(X \in S) - \Pr(Y \in S)\big).$$

   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 $T(p, q) \leqslant \Pr(X \neq Y)$. This beautiful inequality is known as the *coupling inequality*.

2. **Compression vs correlations:** In today's lecture we characterized (n, R, δ)-codes in terms of how well they preserve correlations with an auxiliary random variable. Revisit the proof sketch in light of the results from Problem 1 and make sure you understand each step.

3. **On the definition of quantum codes:** The definition of an (n, R, δ)-quantum code in the lecture was perhaps surprising. Why did we not simply demand that $F(\mathcal{D}[\mathcal{E}[\rho^{\otimes n}]], \rho^{\otimes n}) \geqslant 1 - \delta$? Argue that such a definition would not correspond to a reliable compression protocol. What is the probability theory analog of this condition?

4. **Converse of Schumacher's theorem:** In this problem you can try to prove part 2 of Schumacher's Theorem 6.8 yourself. Fix $\rho \in D(\mathcal{H}_A)$, $\delta \in (0, 1)$, and $R < H(\rho)$.

   (a) Show that there exists $\varepsilon > 0$ such that, for any orthogonal projection P of rank $\leqslant 2^{nR}$,

   $$\mathrm{Tr}[P\rho^{\otimes n}] \leqslant 2^{-\varepsilon n} + \big(1 - \mathrm{Tr}[\Pi_{n,\varepsilon}\rho^{\otimes n}]\big)\,.$$

   (b) Show that, in any (n, R, δ)-code for ρ, the Kraus operators of $\mathcal{D} \circ \mathcal{E}$ have rank $\leqslant 2^{nR}$.

   (c) Show that there exist (n, R, δ)-codes for ρ for at most finitely many values of n. *Hint: Use the formula for the channel fidelity from class and the Cauchy-Schwarz inequality for operators.*

Quantum Information Theory, Spring 2020

Practice problem set #7

You do **not** have to hand in these exercises, they are for your practice only.

1. **Mutual information upper bound:** From class we know that $I(A:B) \leqslant \log d_A + \log d_B$, where $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$. Give a simple proof of this fact.

2. **Weak monotonicity:** Use a purification to deduce the weak monotonicity inequality $H(AC) + H(BC) \geqslant H(A) + H(B)$ from the strong subadditivity inequality, and vice versa.

3. **Strict concavity of the von Neumann entropy:** In Homework Problem 6.4 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 from today's lecture.

4. **Equality condition for monotonicity:** In Homework Problem 5.3, you proved that the Shannon entropy satisfies the following monotonicity inequality: $H(XY) \geqslant H(X)$ for any probability distribution p_{XY}. (Warning: This inequality is in general *false* for quantum states!) Show that equality holds if and only if $p_{XY}(x,y) = p_X(x)\delta_{f(x),y}$ for a function $f\colon \Sigma_X \to \Sigma_Y$.

   In terms of random variables, this means that $Y = f(X)$, i.e., the second is a function of the first!

5. **Binary entropy function:** The Shannon entropy of a probability distribution with two possible outcomes is given by the so-called *binary entropy function*:

$$h(p) := H(\{p, 1-p\}) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p},$$

   (a) Sketch this function.

   (b) Does there exists a constant $L > 0$ such that $|h(p) - h(q)| \leqslant L|p-q|$ for all $0 \leqslant p, q \leqslant 1$? That is, is h Lipschitz continuous?

Quantum Information Theory, Spring 2020

Practice problem set #8

You do **not** have to hand in these exercises, they are for your practice only.

1. **Warmup:**

   (a) Show that, if ρ and σ are both pure states, $D(\rho\|\sigma) \in \{0, \infty\}$.

   (b) Find a state ρ and a channel Φ such that $H(\Phi[\rho]) < H(\rho)$.

   (c) Compute the relative entropy $D(\rho\|\sigma)$ for $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ and $\sigma = \frac{1}{4}|+\rangle\langle +| + \frac{3}{4}|-\rangle\langle -|$.

2. **Matrix logarithm:** Recall that the logarithm of a positive definite operator with eigendecomposition $Q = \sum_i \lambda_i |e_i\rangle\langle e_i|$ is defined as $\log(Q) = \sum_i \log(\lambda_i)|e_i\rangle\langle e_i|$ (as always, our logarithms are to base 2). Verify the following properties:

   (a) $\log(cI) = \log(c)I$ for every $c \geqslant 0$.

   (b) $\log(Q \otimes R) = \log(Q) \otimes I_B + I_A \otimes \log(R)$ for all positive definite $Q \in L(\mathcal{H}_A), R \in L(\mathcal{H}_B)$.

   (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 $\rho_x \in D(\mathcal{H}_B)$.

   Warning: It is in general not true that $\log(QR) = \log(Q) + \log(R)$*!*

3. **From relative entropy to entropy and mutual information:** Use Problem 2 to verify the following claims from class:

   (a) $D(\rho\|\frac{I}{d}) = \log d - H(\rho)$ for every $\rho \in D(\mathcal{H})$, where $d = \dim \mathcal{H}$.

   (b) $D(\rho_{AB}\|\rho_A \otimes \rho_B) = I(A : B)_{\rho_{AB}}$ for every $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $\rho_A = \mathrm{Tr}_B[\rho_{AB}]$ and $\rho_B = \mathrm{Tr}_A[\rho_{AB}]$. You may assume that all three operators are positive definite.

4. **Entropy and ensembles:** In this problem, you will prove the upper bound on the Holevo information that we discussed in class: For every ensemble $\{p_x, \rho_x\}$,

$$\chi(\{p_x, \rho_x\}) \leqslant H(p) \quad \text{or, equivalently,} \quad H(\sum_x p_x \rho_x) \leqslant H(p) + \sum_x p_x H(\rho_x).$$

   Moreover, show that equality holds if and only if the ρ_x with $p_x > 0$ have pairwise orthogonal image.

   In terms of the cq-state corresponding to the ensemble, the above inequality can also be written as $H(XB) \geqslant H(B)$. This confirms our claim in Lecture 7. In Homework Problem 6.4 you showed that $H(XB) \geqslant H(X)$, but since the situation is not symmetric (X is classical but B is not) we now need a different argument.

   (a) First prove these claims assuming that each ρ_x is a pure state, i.e., $\rho_x = |\psi_x\rangle\langle\psi_x|$.

   Hint: Consider the pure state $|\Phi\rangle = \sum_x \sqrt{p_x}|x\rangle \otimes |\psi_x\rangle$ *and compare the entropy of the first system before and after measuring in the standard basis.*

   (b) Now prove the claims for general ρ_x.

   Hint: Apply part (a) to a suitable ensemble obtained from the eigendecompositions of the ρ_x.

Quantum Information Theory, Spring 2020

Practice problem set #9

You do **not** have to hand in these exercises, they are for your practice only.

1. **Warmup:**

   (a) Show that every classical state ρ_{XY} is separable.

   (b) Let $|\psi_{AB}\rangle$ be a pure state. Show that the following are equivalent: (i) $|\psi_{AB}\rangle$ is separable, (ii) its Schmidt rank is one, (iii) its entanglement entropy is zero. Recall from Lemma 2.12 that the Schmidt rank is the number of non-zero coefficients in the Schmidt decomposition.

   (c) In the lecture we defined *separable operators* to be those that can be written as

   $$\sum_i P_{A,i} \otimes Q_{B,i}$$

   where $P_{A,i}$ and $Q_{B,i}$ are positive semidefinite. Show that the restriction of this definition to density matrices coincides with the definition of separable states from the lecture. Show that restricting this further to pure states also coincides with the definition from the lecture.

   (d) Recall that the four *Bell states* are defined by

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

   where $z, x \in \{0,1\}$ and $|\Phi^+\rangle$ is the canonical two-qubit maximally entangled state. Verify that

   $$|\Phi^{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \qquad |\Phi^{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),$$
   $$|\Phi^{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \qquad |\Phi^{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle).$$

   Verify also that

   $$|\Phi^{zx}\rangle = (I \otimes X^x Z^z)|\Phi^+\rangle.$$

2. **Maximally entangled state tricks:** Let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^\Sigma$ and

   $$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{|\Sigma|}} \sum_{x\in\Sigma} |x\rangle \otimes |x\rangle.$$

   a maximally entangled state.

   (a) Show that, for any $M \in L(\mathcal{H}_A)$,

   $$(M \otimes I)|\Phi^+\rangle = (I \otimes M^T)|\Phi^+\rangle.$$

   (b) Show that for $M, N \in L(\mathcal{H}_B)$

   $$\operatorname{Tr}(M^\dagger N) = |\Sigma| \langle \Phi^+ | \overline{M} \otimes N | \Phi^+ \rangle.$$

3. **Two-qubit pure states (product vs entangled):** Let

$$|\psi\rangle = \begin{pmatrix} \psi_{00} \\ \psi_{01} \\ \psi_{10} \\ \psi_{11} \end{pmatrix} \in \mathbb{C}^4.$$

be an arbitrary pure state on two qubits. Define

$$\Delta(|\psi\rangle) = \psi_{00}\psi_{11} - \psi_{01}\psi_{10}.$$

The goal of this exercise is to show that $\Delta(|\psi\rangle) = 0$ if and only if $|\psi\rangle$ is a product state.

(a) Assume that $|\psi\rangle$ is a product state, i.e., $|\psi\rangle = |\alpha\rangle \otimes |\beta\rangle$, for some single-qubit states

$$|\alpha\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}, \qquad |\beta\rangle = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

Show that in such case $\Delta(|\psi\rangle) = 0$.

(b) Conversely, let $|\psi\rangle$ be an arbitrary two-qubit state and assume that $\Delta(|\psi\rangle) = 0$. Find two single-qubit states $|\alpha\rangle$ and $|\beta\rangle$ such that $|\psi\rangle = |\alpha\rangle \otimes |\beta\rangle$.

The quantity $\Delta(|\psi\rangle)$ can not only be used to determine if a pure two-qubit state is entangled or not but is also a meaningful measure of the amount of entanglement.

4. **Pauli matrices and the swap:** Let $\Sigma = \{0, 1\}$ and $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^\Sigma$. The two-qubit *swap operation* $W \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is defined on the computational basis states as follows:

$$W|a, b\rangle = |b, a\rangle,$$

for all $a, b \in \{0, 1\}$. Recall that the four *Pauli matrices* are

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(a) Verify that

$$W = \frac{1}{2}(I \otimes I + X \otimes X + Y \otimes Y + Z \otimes Z).$$

(b) Verify that

$$W = \frac{1}{2} \sum_{z,x \in \{0,1\}} Z^z X^x \otimes X^x Z^z.$$

Quantum Information Theory, Spring 2020

Practice problem set #10

You do **not** have to hand in these exercises, they are for your practice only.

1. **Vectorization:**

   (a) Let $\mathcal{H}_A = \mathbb{C}^\Sigma, \mathcal{H}_B = \mathbb{C}^\Gamma$, and $M = |b\rangle\langle a| \in L(\mathcal{H}_A, \mathcal{H}_B)$, for some $a \in \Sigma$ and $b \in \Gamma$. Recall from Definition 10.2 that the vectorization of M is given by $|M_{AB}\rangle = |a, b\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. Prove the vectorization identity

   $$(A \otimes B)|M\rangle = |BMA^T\rangle$$

   when $A = |c\rangle\langle a|$ and $B = |d\rangle\langle b|$, for some standard basis states $|c\rangle$ and $|d\rangle$.

   (b) Argue why this implies the same identity for arbitrary $A \in L(\mathcal{H}_A, \mathcal{H}_C)$, $B \in L(\mathcal{H}_B, \mathcal{H}_D)$, $M \in L(\mathcal{H}_A, \mathcal{H}_B)$.

2. **Separable maps:**

   (a) Let $\Xi \in \mathrm{CP}(\mathcal{H}_A \otimes \mathcal{H}_B, \mathcal{H}_C \otimes \mathcal{H}_D)$. Show that $\Xi \in \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$ if and only if there exist $A_x \in L(\mathcal{H}_A, \mathcal{H}_C)$ and $B_x \in L(\mathcal{H}_B, \mathcal{H}_D)$ such that

   $$\Xi(X) = \sum_{x \in \Sigma} (A_x \otimes B_x) X (A_x \otimes B_x)^\dagger,$$

   for all $X \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$.

   (b) Let $\Xi_1 \in \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_C : \mathcal{H}_B, \mathcal{H}_D)$ and $\Xi_2 \in \mathrm{SepCP}(\mathcal{H}_C, \mathcal{H}_E : \mathcal{H}_D, \mathcal{H}_F)$. Show that their composition is also separable:

   $$\Xi_2 \circ \Xi_1 \in \mathrm{SepCP}(\mathcal{H}_A, \mathcal{H}_E : \mathcal{H}_B, \mathcal{H}_F).$$

3. **Examples of separable maps:** Show that the following maps jointly implemented by Alice and Bob are separable:

   (a) Alice and Bob share a random variable distributed according to $p \in P(\Sigma \times \Gamma)$, where Σ labels Alice's register and Γ labels Bob's register. Moreover, Alice has a quantum register A and Bob has a quantum register B. They both observe their halves of the random variable. If Alice's value is $x \in \Sigma$, she applies a channel Φ_x on her register A. Similarly, if Bob's value is $y \in \Gamma$, he applies a channel Ψ_y on his register B.

   (b) Alice has a register A that she measures in the standard basis. She sends the measurement outcome $x \in \Sigma$ to Bob who applies a channel Ψ_x on his register B.

   (c) Any LOCC channel.

4. **Instruments:** Recall from Definition 10.7 that an instrument is a collection $\{\Phi_\omega : \omega \in \Omega\} \subset \mathrm{CP}(\mathcal{H}_A, \mathcal{H}_B)$ such that $\sum_{\omega \in \Omega} \Phi_\omega \in \mathrm{C}(\mathcal{H}_A, \mathcal{H}_B)$. When applied to a state $\rho \in D(\mathcal{H}_A)$, it produces outcome $\omega \in \Omega$ with probability $\mathrm{Tr}[\Phi_\omega[\rho]]$ and changes ρ to $\rho_\omega = \Phi_\omega[\rho]/\mathrm{Tr}[\Phi_\omega[\rho]]$. Show that any instrument can be implemented by a quantum channel, followed by an orthonormal measurement.

Quantum Information Theory, Spring 2020

Practice problem set #11

You do **not** have to hand in these exercises, they are for your practice only.

1. **Majorization examples:**

   (a) Let $p = (0.1, 0.7, 0.2)$ and $q = (0.3, 0.2, 0.5)$. Determine whether $p \prec q$ or $q \prec p$.
   (b) Find a sequence of Robin Hood transfers that converts 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.
   (e) Find a pair of probability distributions p and q such that neither $p \prec q$ nor $q \prec p$.

2. **Alternative definitions of majorization:** Let $u = (u_1, \dots, u_n)$ be a vector and let r denote *reverse sorting* and s denote *sorting*:

$$\begin{aligned} r_1(u) \geqslant r_2(u) \geqslant \cdots \geqslant r_n(u), \\ s_1(u) \leqslant s_2(u) \leqslant \cdots \leqslant s_n(u), \end{aligned}$$

   such that $\{r_i(u) : i = 1, \dots, n\} = \{s_i(u) : i = 1, \dots, n\} = \{u_i : i = 1, \dots, n\}$ as multisets. Let u and v be two probability distributions over $\Sigma = \{1, \dots, n\}$, i.e., $u_i \geqslant 0$, $v_i \geqslant 0$, and $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i = 1$. Show that the following ways of expressing $v \prec u$ are equivalent:

   (a) $\sum_{i=1}^m r_i(v) \leqslant \sum_{i=1}^m r_i(u)$, for all $m \in \{1, \dots, n-1\}$.

   (b) $\sum_{i=1}^m s_i(v) \geqslant \sum_{i=1}^m s_i(u)$, for all $m \in \{1, \dots, n-1\}$.

   (c) $\forall t \in \mathbb{R} : \sum_{i=1}^n \max(v_i - t, 0) \leqslant \sum_{i=1}^n \max(u_i - t, 0)$.

3. **Vectorization and partial trace:**

   (a) Show that, for all $L, R \in L(\mathcal{H}_A, \mathcal{H}_B)$,

$$\mathrm{Tr}_A \left[|L\rangle\langle R| \right] = LR^\dagger.$$

   (b) Let $\Xi \in \mathrm{SepC}(A : B)$ be given by

$$\Xi(M) = \sum_{a \in \Sigma} (A_a \otimes B_a) M (A_a \otimes B_a)^\dagger,$$

   for all $M \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$. Show that, for all $X \in L(\mathcal{H}_A, \mathcal{H}_B)$,

$$\mathrm{Tr}_A \Big[\Xi(|X\rangle\langle X|) \Big] = \sum_{a \in \Sigma} B_a X A_a^\mathsf{T} \bar{A}_a X^\dagger B_a^\dagger.$$

Quantum Information Theory, Spring 2020

Practice problem set #12

You do **not** have to hand in these exercises, they are for your practice only.

1. **Maximally entangled states:** A pure state $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is *maximally entangled* if

$$\mathrm{Tr}_B\left[|\Psi\rangle\langle\Psi|\right] = \frac{I_A}{\dim(\mathcal{H}_A)} \qquad \text{and} \qquad \mathrm{Tr}_A\left[|\Psi\rangle\langle\Psi|\right] = \frac{I_B}{\dim(\mathcal{H}_B)}.$$

   (a) Show that it must be the case that $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_B)$.

   (b) Let $|\Psi_{AB}\rangle, |\Psi'_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be two maximally entangled states. Show that there exist local unitaries $U_A \in U(\mathcal{H}_A)$ and $V_B \in U(\mathcal{H}_B)$ such that $(U_A \otimes V_B)|\Psi_{AB}\rangle = |\Psi'_{AB}\rangle$.

   (c) Let $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\Phi_{A'B'}\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ be maximally entangled. Show that $|\Psi_{AB}\rangle \otimes |\Phi_{A'B'}\rangle$ is also maximally entangled with respect to the partition $\mathcal{H}_A \otimes \mathcal{H}_{A'} : \mathcal{H}_B \otimes \mathcal{H}_{B'}$.

   (d) Let $|\Psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a maximally entangled state with $\dim(\mathcal{H}_A) = \dim(\mathcal{H}_B) = d$, and let

$$|\Phi^+_{AB}\rangle = \frac{|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B}{\sqrt{2}}$$

   be the canonical two-qubit maximally entangled state. Show that an exact copy of $|\Psi_{AB}\rangle$ can be obtained by LOCC from $|\Phi^+_{AB}\rangle^{\otimes n}$, for some large enough n. What is the smallest value of n for which this holds?

2. **Fidelity and composition of channels:** Let $\tau_1 \in D(\mathcal{H})$, $\sigma \in D(\mathcal{K})$, $\tau_2 \in D(\mathcal{L})$ be quantum states and let $\Phi \in C(\mathcal{H}, \mathcal{K})$ and $\Psi \in C(\mathcal{K}, \mathcal{L})$ be quantum channels. Assuming that

$$F\big(\Phi(\tau_1), \sigma\big) > 1 - \varepsilon, \qquad F\big(\Psi(\sigma), \tau_2\big) > 1 - \varepsilon, \tag{A.2}$$

   for some $\varepsilon > 0$, show that

$$F\big((\Psi \circ \Phi)(\tau_1), \tau_2\big) > 1 - 4\varepsilon, \tag{A.3}$$

   where $\Psi \circ \Phi$ denotes the composition of the two channels.

   Hint: Recall from Homework Problem 4.1 that fidelity is monotonic under any quantum channel. Moreover, you will show in Homework Problem 12.1 that $F(\rho_1, \sigma)^2 + F(\rho_2, \sigma)^2 \leqslant 1 + F(\rho_1, \rho_2)$, for any states $\rho_1, \rho_2, \sigma \in D(\mathcal{H})$.

3. **From any 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. How many copies of the state σ can be distilled per copy of ρ by LOCC?

Quantum Information Theory, Spring 2020

Practice problem set #13

You do **not** have to hand in these exercises, they are for your practice only.

1. **Symmetric subspace:**

   (a) Write out Π_2 and Π_3.

   (b) In class we wrote down a basis for $\mathrm{Sym}^2(\mathbb{C}^2)$. Write down bases of $\mathrm{Sym}^2(\mathbb{C}^d)$ and $\mathrm{Sym}^3(\mathbb{C}^2)$.

   (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.

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) Recall the following important fact from Lemma 13.10: *If* $A \in L(\mathcal{H}^{\otimes n})$ *is an operator such that* $[A, U^{\otimes n}] = 0$ *for all unitaries* $U \in U(\mathcal{H})$, *then* A *is a linear combination of permutation operators* R_π, $\pi \in S_n$. Use this fact 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 $\tilde{\Pi}_n = \Pi_n$.

3. **Haar measure:** There is a unique probability measure dU on the unitary operators $U(\mathcal{H})$ that is invariant under $U \mapsto VUW$ 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 $\int f(U)\, dU = \int f(VUW)\, dU$. Now let $A \in L(\mathcal{H})$.

   (a) Argue that $\int UAU^\dagger dU$ commutes with all unitaries.

   (b) Deduce that $\int UAU^\dagger dU = \frac{\mathrm{Tr}[A]}{d} I$, where $d = \dim \mathcal{H}$.

4. ⚗ **De Finetti theorem and quantum physics (optional):** 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 $\frac{E_0}{n} \leqslant \frac{1}{n}\langle\psi^{\otimes n}|H|\psi^{\otimes n}\rangle = \langle\psi^{\otimes 2}|h|\psi^{\otimes 2}\rangle$ 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 $|E_0\rangle \in \mathrm{Sym}^n(\mathbb{C}^d)$ (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* $\psi^{\otimes n}$.

Appendix B

Homework Problems

Quantum Information Theory, Spring 2020

Homework problem set #1 **due February 10, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **Trace distance between pure states:** Let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$ be two pure states on an arbitrary Hilbert space $\mathcal{H}$. Show that

$$\frac{1}{2}\|\rho - \sigma\|_1 = \sqrt{1 - |\langle\psi|\phi\rangle|^2}.$$

Here, $\|\cdot\|_1$ denotes the *trace norm*, which for a Hermitian operator A with eigenvalues (a_i) is defined by $\|A\|_1 := \sum_i |a_i|$.

Hint: Argue that the eigenvalues of $\rho - \sigma$ *are of the form* $(\lambda, -\lambda, 0, \dots, 0)$ *for some* $\lambda \in \mathbb{R}$. *Compute* $\|\rho - \sigma\|_1$ *and* $\mathrm{Tr}[(\rho - \sigma)^2]$ *in terms of* λ. *Can you relate the latter to* $|\langle\psi|\phi\rangle|^2$*?*

2. *(4 points)* **Uncertainty relation:** Given a measurement $\mu\colon \{0,1\} \to \mathrm{PSD}(\mathcal{H})$ with two outcomes and a state $\rho \in \mathrm{D}(\mathcal{H})$, define the *bias* by

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

(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 $\beta_{\mathrm{Std}}(\rho)$ and $\beta_{\mathrm{Had}}(\rho)$ in terms of the Bloch vector of the qubit state ρ.

(c) Show that $\beta^2_{\mathrm{Std}}(\rho) + \beta^2_{\mathrm{Had}}(\rho) \leqslant 1$. Why is this called an *uncertainty relation*?

3. *(4 bonus points)* 🖩 **Practice:** In the exercise class, you discussed how to estimate an unknown qubit state ρ by performing the following measurement on many copies of ρ:

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

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

The file `01-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.

Quantum Information Theory, Spring 2020

Homework problem set #2 **due February 17, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **Nayak's bound:** Alice wants to communicate m bits to Bob by sending n qubits. She chooses one state $\rho(x) \in D(\mathcal{H})$, 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 $\mu\colon \{0,1\}^m \to \mathrm{PSD}(\mathcal{H})$ to decode the message.

   (a) 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 $\{0,1\}^m$.
   (b) 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 points)* **Trace distance and Helstrom's theorem:** The *(normalized) trace distance* between two quantum states ρ, σ on a Hilbert space $\mathcal{H}$ is defined as

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

   in terms of the *trace norm* $\|\cdot\|_1$, which you know from the lecture and the previous homework.

   (a) Show that $T(\rho, \sigma) \in [0, 1]$.
   (b) Show that $T(\rho, \sigma) = \max_{0 \leqslant Q \leqslant I} \mathrm{Tr}[Q(\rho - \sigma)]$, and that the maximum can be achieved by a projection Q. *Hint: Consider the spectral decomposition of* $\rho - \sigma$.

   Now suppose we want to distinguish ρ and σ by a measurement $\mu : \{0,1\} \to \mathrm{PSD}(\mathcal{H})$. By convention, outcome '0' corresponds to state ρ, while outcome '1' corresponds to state σ. Assuming both states occur with 50% probability, the probability of success using μ is given by

$$p_{\text{success}} = \frac{1}{2}\mathrm{Tr}[\rho\mu(0)] + \frac{1}{2}\mathrm{Tr}[\sigma\mu(1)].$$

   (c) Use (b) to prove *Helstrom's theorem*, which states that the *maximal* probability of success (over all possible measurements) is $\frac{1}{2} + \frac{1}{2}T(\rho, \sigma)$ and can be achieved by a *projective* measurement.

3. *(4 points)* **Extensions of pure states:** Let $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_C$ be arbitrary Hilbert spaces.

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

   Notation: Just like in class, if ρ_{AB} is a state then we write ρ_A and ρ_B for its reduced states obtained by taking suitable partial traces (likewise for ρ_{ABC} and its reduced states).

Quantum Information Theory, Spring 2020

Homework problem set #3 **due February 24, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(3 points)* **Pretty good measurement:** Let $\rho_1, \dots, \rho_n$ be a collection of quantum states on some Hilbert space $\mathcal{H}$ with the property that $I \in \operatorname{span}\{\rho_1, \dots, \rho_n\}$.

   (a) Show that $Q := \sum_{i=1}^n \rho_i$ is positive semidefinite and invertible. *Hint: For the latter, assume* Q *has an eigenvector* $|\psi\rangle$ *with eigenvalue* 0. *Use the span property to get a contradiction.*
   (b) Define $\mu\colon \{1, \dots, n\} \to L(\mathcal{H})$ by $\mu(i) = Q^{-1/2} \rho_i Q^{-1/2}$. Show that μ is a measurement.

2. *(3 points)* **Properties of the fidelity:** Use Uhlmann's theorem to prove the following properties of the fidelity. As always, $\mathcal{H}_A$ and $\mathcal{H}_B$ denote arbitrary Hilbert spaces.

   (a) *Monotonicity:* $F(\rho_{AB}, \sigma_{AB}) \leqslant F(\rho_A, \sigma_A)$ for any two states $\rho_{AB}, \sigma_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$.
   (b) *Joint concavity:* $\sum_{i=1}^n p_i F(\rho_i, \sigma_i) \leqslant F(\sum_{i=1}^n p_i \rho_i, \sum_{i=1}^n p_i \sigma_i)$, where $(p_i)_{i=1}^n$ is an arbitrary probability distribution and $\rho_1, \dots, \rho_n$ and $\sigma_1, \dots, \sigma_n$ are states in $D(\mathcal{H}_A)$.

   Hint: For both (a) and (b), try to find suitable purifications.

3. *(3 points)* **Quantum channels:** Show that the following maps Φ are quantum channels by verifying that they are completely positive and trace-preserving:

   (a) *Mixture of unitaries:* $\Phi[M] = \sum_{i=1}^n p_i U_i M U_i^\dagger$, where $(p_i)_{i=1}^n$ is an arbitrary probability distribution and $U_1, \dots, U_n$ arbitrary unitaries.
   (b) *State replacement:* $\Phi[M] = \operatorname{Tr}[M]\, \sigma$, 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 $\mathbb{C}^\Sigma$ and σ_x is an arbitrary state for each $x \in \Sigma$.

4. *(3 points)* **No cloning:** In this problem, you will show that it is not possible to perfectly clone an unknown state – even if we restrict to classical or to pure states. Let $\mathcal{H} = \mathbb{C}^2$ be a qubit. We say that a channel $\Phi \in C(\mathcal{H}, \mathcal{H} \otimes \mathcal{H})$ *clones* a state $\rho \in D(\mathbb{C}^2)$ if $\Phi[\rho] = \rho \otimes \rho$.

   (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.

5. *(2 bonus points)* 🖩 **Practice:** Here you can verify that the measurement from Problem 1 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 $\sigma(t)$. *Hint: Helstrom's theorem.*
   (b) The probability of distinguishing ρ and $\sigma(t)$ by using the measurement from Problem 1.

Quantum Information Theory, Spring 2020

Homework problem set #4 **due March 2, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(2 points)* **Monotonicity of distance measures:** Use the Stinespring representation to deduce the following monotonicity properties. For all states ρ_A, σ_A and channels $\Phi_{A\to B}$,

$$T(\Phi_{A\to B}[\rho_A], \Phi_{A\to B}[\sigma_A]) \leqslant T(\rho_A, \sigma_A) \quad \text{and} \quad F(\Phi_{A\to B}[\rho_A], \Phi_{A\to B}[\sigma_A]) \geqslant F(\rho_A, \sigma_A).$$

2. *(3 points)* **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:

$$\mathcal{D}_\lambda[M] = \lambda M + (1-\lambda) \operatorname{Tr}[M] \frac{I}{d}$$

   (a) Compute the Choi operator of $\mathcal{D}_\lambda$ for any value of λ.
   (b) For which values of λ is $\mathcal{D}_\lambda$ a quantum channel?

3. *(5 points)* **Kraus and Stinespring:** Find Kraus and Stinespring representations for the following quantum channels:

   (a) *Partial trace:* $\Phi[M_{AE}] = \operatorname{Tr}_E[M_{AE}]$
   (b) *Add state:* $\Phi[M_A] = M_A \otimes \sigma_B$ for a state σ_B.
   (c) *Measure and prepare:* $\Phi[M] = \sum_{x\in\Sigma} \langle x|M_A|x\rangle\, \sigma_{B,x}$, where $|x\rangle$ denotes the standard basis of $\mathcal{H}_A = \mathbb{C}^\Sigma$ and $\sigma_{B,x}$ is an arbitrary state for each $x \in \Sigma$.

4. *(2 points)* **Quantum to classical channels:** Let $\mathcal{H}_A$ be an arbitrary Hilbert space and $\mathcal{H}_X = \mathbb{C}^\Omega$. Assume that $\Phi_{A\to X}$ is a quantum channel such that $\Phi_{A\to X}[\rho_A]$ is classical for every state ρ_A. Show that there exists a measurement $\mu_A\colon \Omega \to \mathrm{PSD}(\mathcal{H}_A)$ such that

$$\Phi_{A\to X}[\rho_A] = \sum_{x\in\Omega} \operatorname{Tr}\left[\mu_A(x)\rho_A\right] |x\rangle\langle x| \qquad \forall \rho_A.$$

   Hint: Use Practice Problem 4.1.

Quantum Information Theory, Spring 2020

Homework problem set #5 — due March 9, 2020

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(2 points)* **Fidelity between classical-quantum states:** 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
$$F(\rho_{XB}, \sigma_{XB}) = \sum_{x\in\Sigma} \sqrt{p(x)q(x)}\, F(\rho_{B,x}, \sigma_{B,x}).$$

2. *(3 points)* **Gentle measurement lemma:** This useful technical result states that if $\rho \in D(\mathcal{H})$ is a state and $0 \leqslant Q \leqslant I$ an operator such that $\mathrm{Tr}[Q\rho] \geqslant 1-\varepsilon$, then the following inequalities hold:
$$F\Big(\rho, \frac{\sqrt{Q}\rho\sqrt{Q}}{\mathrm{Tr}[Q\rho]}\Big) \geqslant \sqrt{1-\varepsilon} \quad \text{and} \quad T\Big(\rho, \frac{\sqrt{Q}\rho\sqrt{Q}}{\mathrm{Tr}[Q\rho]}\Big) \leqslant \sqrt{\varepsilon} \tag{B.1}$$
   (a) Prove that $\mathrm{Tr}\sqrt{\sqrt{\rho}\sqrt{Q}\rho\sqrt{Q}\sqrt{\rho}} = \mathrm{Tr}[\sqrt{Q}\rho]$ and $\sqrt{Q} \geqslant Q$.
   Hint: The square root $\sqrt{A}$ of a PSD operator A is the unique PSD operator that squares to A.
   (b) Prove the first inequality in Eq. (B.1) using part (a), and deduce the second inequality from the first by using a result from the practice problems.

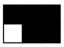
3. *(4 points)* **Properties of the Shannon entropy:** Given a joint distribution, we write $H(XY)$ for its Shannon entropy and $H(X)$, $H(Y)$ for the entropies of its marginal distributions.
   (a) *Monotonicity:* Show that $H(XY) \geqslant H(Y)$.
   (b) *Subadditivity:* Show that $H(X) + H(Y) \geqslant H(XY)$.
   (c) 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.

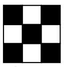
4. *(3 points)* **Optimality of the Shannon entropy:** In this problem, you will prove the converse part of Shannon's source coding theorem which states that it is impossible to compress at rates below the entropy of the source. Given a probability distribution p on a finite set Σ, recall that an (n, R, δ)-*code* consists of functions $E\colon \Sigma^n \to \{0,1\}^{\lfloor nR \rfloor}$ and $D\colon \{0,1\}^{\lfloor nR \rfloor} \to \Sigma^n$ such that $\sum_{x^n\in\Sigma^n : D(E(x^n))=x^n} p(x_1)\cdots p(x_n) \geqslant 1-\delta$. 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.

5. *(2 bonus points)* 🖩 **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 $m \in \{0, 1, \ldots, \binom{rs}{k}-1\}$

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 $(r, s, k, m) = (7, 8, 8, 243185306)$?

Quantum Information Theory, Spring 2020

Homework problem set #6 due March 16, 2020

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(3 points)* **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 $F(\mathcal{D} \circ \mathcal{E}, \rho^{\otimes n}) \geqslant 1 - \delta$ then

$$\sum_{x^n} p(x_1)\cdots p(x_n)\, F\big(\mathcal{D}\big[\mathcal{E}[\rho_{x_1} \otimes \cdots \otimes \rho_{x_n}]\big], \rho_{x_1} \otimes \cdots \otimes \rho_{x_n}\big) \geqslant 1 - \delta.$$

Show that the converse is not necessarily true.

Hint: There are even counterexamples for $n = 1$ and $\delta = 0$. Consider measure-and-prepare channels.

2. *(2 points)* **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.

3. *(3 points)* **Subadditivity of the von Neumann entropy:** Use Schumacher's theorem to show that, for all states ρ_{AB},

$$H(A) + H(B) \geqslant H(AB),$$

where we use the same notation as in the previous problem. Thus, *subadditivity* still holds.

Hint: You may use the following 'triangle inequality' for the fidelity (without proof): For any three states $\alpha, \beta, \gamma \in D(\mathcal{H})$, if $F(\alpha, \beta) \geqslant 1 - \delta$ and $F(\beta, \gamma) \geqslant 1 - \delta$ then $F(\alpha, \gamma) \geqslant 1 - 4\delta$.

4. *(4 points)* **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 $\mathcal{H}_B = \mathcal{H}$.

   (a) Show that $H(XB) = H(p) + \sum_{x\in\Sigma} p(x) H(\rho_x)$.
   (b) Conclude that $H(XB) \geqslant H(X)$. 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 Problem 3 for a classical-quantum state.

Quantum Information Theory, Spring 2020

Homework problem set #7 **due March 23, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(2 points)* **Quantum mutual information:** From class, we know that $I(A:B) \leqslant 2\log d$ 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 ρ_{AB} is a pure state with $\rho_A = \rho_B = I/d$ (such states are called *maximally entangled*). Write down the Schmidt decomposition of a general state of this form.

   Hint: In the exercise class you gave simple proof of the above inequality.

2. *(2 points)* **Classical mutual information:** From class, we know that $I(X:Y) \leqslant \log d$ 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 $f\colon \Sigma_X \to \Sigma_Y$ (such p_{XY} are called *maximally correlated*).

   Hint: In the exercise class you characterized the probability distributions with $H(XY) = H(X)$.

3. *(8 points)* **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:

$$H(p_{\text{Std}}) + H(p_{\text{Had}}) \geqslant H(\rho) + 1 \tag{B.2}$$

   (a) Why is it appropriate to call (B.2) 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 $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and $Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

   (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,

$$\omega_{ABC} = \frac{1}{4}\sum_{a=0}^{1}\sum_{b=0}^{1} |a\rangle\langle a| \otimes |b\rangle\langle b| \otimes X^a Z^b \rho Z^b X^a,$$

   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 that you proved last week.
   (f) Use part (e) and the strong subadditivity inequality to deduce (B.2).

4. *(2 bonus points)* 🖩 **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 $\varepsilon = 0.01$:

   $$r(n) = \frac{1}{n} \log \dim S_{n,\varepsilon}, \quad p(n) = \operatorname{Tr}[\Pi_{n,\varepsilon} \rho^{\otimes n}],$$

   where $\Pi_{n,\varepsilon}$ denotes the orthogonal projection onto the typical subspace $S_{n,\varepsilon}$ of ρ.

Quantum Information Theory, Spring 2020

Homework problem set #8 **due March 30, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **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 $\mu\colon \Omega \to \mathrm{PSD}(\mathcal{H})$, let $p, q \in P(\Omega)$ denote the corresponding probability distributions of measurement outcomes.

   (a) Prove that $T(p, q) \leqslant T(\rho, \sigma)$.
   (b) Show that, for any ρ and σ, there exists a measurement μ such that equality holds.

   Hint: Recall Helstrom's theorem. You can choose Ω.

2. *(4 points)* **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 $\mu\colon \Sigma \to \mathrm{PSD}(\mathcal{H})$ 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.

   Hint: In Practice Problem 8.4 you discussed when an ensemble satisfies $\chi(\{p_x, \rho_x\}) = H(p)$.

3. *(4 points)* **Applications of monotonicity:** Prove the following two inequalities by using the monotonicity of the quantum relative entropy:

   (a) *Entropy increase:* $H(\Phi[\rho]) \geqslant H(\rho)$ for every $\rho \in D(\mathcal{H})$ and *unital* channel $\Phi \in C(\mathcal{H}, \mathcal{H}')$. Recall that a channel is *unital* if $\Phi[I_{\mathcal{H}}] = I_{\mathcal{H}'}$.
   (b) *Joint convexity* of relative entropy: $D(\sum_{x\in\Sigma} p_x\rho_x \| \sum_{x\in\Sigma} p_x\sigma_x) \leqslant \sum_{x\in\Sigma} p_x D(\rho_x\|\sigma_x)$, where $(p_x)_{x\in\Sigma}$ is an arbitrary finite probability distribution and $(\rho_x)_{x\in\Sigma}$, $(\sigma_x)_{x\in\Sigma}$ families of states in $D(\mathcal{H})$. You may assume that the operators ρ_x and σ_x are positive definite.

   Hint: In the exercise class, we computed the logarithm of a cq-state.

Quantum Information Theory, Spring 2020

Homework problem set #9 **due April 6, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **Teleportation and entanglement swapping:**

   (a) Let $|\psi\rangle \in \mathbb{C}^2$ be an arbitrary pure single-qubit state. Verify the teleportation identity

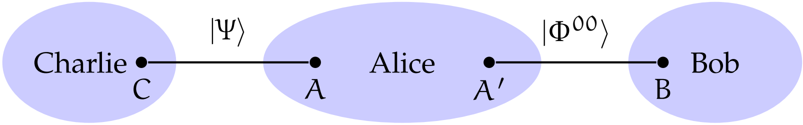
$$|\psi\rangle \otimes |\Phi^{00}\rangle = \frac{1}{2} \sum_{z,x\in\{0,1\}} |\Phi^{zx}\rangle \otimes X^x Z^z |\psi\rangle,$$

   where $|\Phi^{zx}\rangle$ are the four Bell states and X and Z are the Pauli matrices.
   Hint: Some identities from the exercise class might be handy.

   (b) Consider the following generalization of the teleportation identity involving three friends Alice, Bob and Charlie who share four qubits: Alice has two qubits in systems A and A′, Bob has a qubit in system B, while Charlie has a qubit in system C. Their joint state is

$$|\Psi\rangle_{CA} \otimes |\Phi^{00}\rangle_{A'B}$$

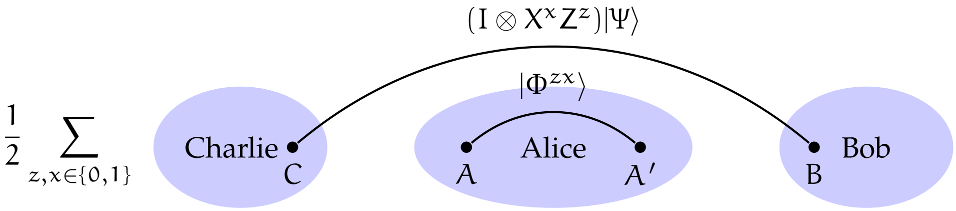
   where $|\Psi\rangle_{CA} \in \mathbb{C}^4$ is an arbitrary pure two-qubit state. This can be depicted as follows:



   Show that

$$|\Psi\rangle_{CA} \otimes |\Phi^{00}\rangle_{A'B} = \frac{1}{2} \sum_{z,x\in\{0,1\}} |\Phi^{zx}\rangle_{AA'} \otimes (I \otimes X^x Z^z)|\Psi\rangle_{CB}.$$

   Note that pictorially the right-hand side looks as follows:


$(I \otimes X^x Z^z)|\Psi\rangle$
$|\Phi^{zx}\rangle$
$\frac{1}{2}\sum_{z,x\in\{0,1\}}$
Charlie
C
A
Alice
A′
B
Bob

   (c) Using the equation from the previous part, explain what happens if Alice and Bob perform the usual teleportation protocol. More specifically, what are the probabilities of outcomes if Alice measures her two qubits in the Bell basis? For each outcome, what is the corresponding post-measurement state for Bob and Charlie? What Pauli correction operation should Bob apply at the end of the protocol?

   (d) What could be a potential application of the protocol performed by Alice, Bob and Charlie?

2. *(3 points)* **Superdense coding:** Assume that Alice and Bob share the Bell state $|\Phi^{00}\rangle_{AB}$. Assume that Alice wants to send two bits $z, x \in \{0, 1\}$ to Bob. They perform the following protocol: (i) Alice applies some unitary operation on her qubit, (ii) sends her qubit to Bob, and (iii) Bob performs an orthogonal measurement to recover z and x.

   (a) What operation should Alice apply?
   (b) What measurement should Bob perform?
   (c) Formulate this procedure as a resource trade-off.

3. *(5 points)* **Partial transpose test:** Let Σ be an alphabet, $n = |\Sigma|$, and $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^\Sigma$. Let

$$|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{n}} \sum_{a \in \Sigma} |a\rangle_A \otimes |a\rangle_B$$

   be an n-dimensional *maximally entangled state* on registers AB. Let $t \in [0, 1]$ and define $\rho_0, \rho_1, \rho(t) \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$ as follows:

$$\rho_0 = |\Phi^+\rangle\langle\Phi^+|, \qquad \rho_1 = \frac{I \otimes I - \rho_0}{n^2 - 1}, \qquad \rho(t) = (1 - t)\rho_0 + t\rho_1.$$

   (a) Show that $\rho(t)$ is a quantum state when $t \in [0, 1]$.
   (b) Let $\mathcal{T} \in L(L(\mathcal{H}_A), L(\mathcal{H}_A))$ be the *transpose* map defined as $\mathcal{T}[X] = X^T$, for all $X \in L(\mathcal{H}_A)$. Show that

$$(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_0] = \frac{1}{n} W,$$

   where $W \in L(\mathcal{H}_A \otimes \mathcal{H}_B)$ is the *swap operator* defined as

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

   for all $x, y \in \Sigma$.
   (c) Compute $(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho(t)]$.
   (d) For what range of t can we say that $\rho(t)$ is entangled?

4. *(2 bonus points)* 🖩 **Practice** In this problem, you will play around with the partial transpose test (part a) and the entanglement entropy (part b).

   (a) In the files `A.txt`, `B.txt` and `C.txt`, you will find density matrices of the following dimensions:

$$A \in D(\mathbb{C}^2 \otimes \mathbb{C}^2), \qquad B \in D(\mathbb{C}^2 \otimes \mathbb{C}^3), \qquad \text{and} \qquad C \in D(\mathbb{C}^2 \otimes \mathbb{C}^4).$$

   For each of the states A, B and C, compute its partial transpose (where the transpose is applied to the second register) 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: Recall that the partial transpose test is always conclusive when the total dimension is at most 6.
   (b) In the file `D.txt`, you will find the *pure* state $D \in D(\mathbb{C}^2 \otimes \mathbb{C}^2)$. Calculate its entanglement entropy.

Quantum Information Theory, Spring 2020

Homework problem set #10 **due April 13, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(2 points)* **Discriminating Bell states by LOCC:** Recall that the Bell states are given by

$$|\Phi^{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \qquad |\Phi^{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),$$
$$|\Phi^{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \qquad |\Phi^{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle).$$

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 $|\Phi^{01}\rangle$.
(b) Find an LOCC protocol that can perfectly discriminate between $|\Phi^{00}\rangle$ and $|\Phi^{10}\rangle$.

2. *(6 points)* **One-way LOCC struggle:** Let $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. Consider a two-qubit system where Alice holds the first qubit and Bob holds the second qubit. These two qubits are initialized in one of the following four states:

$$|\Psi_1\rangle = |0\rangle \otimes |0\rangle,$$
$$|\Psi_2\rangle = |0\rangle \otimes |1\rangle,$$
$$|\Psi_3\rangle = |1\rangle \otimes |+\rangle,$$
$$|\Psi_4\rangle = |1\rangle \otimes |-\rangle.$$

(a) Show that if μ is a separable measurement, and μ perfectly distinguishes an orthonormal basis, then this basis must consist of product states.
(b) Write down a measurement that perfectly distinguishes the above four states and show that it is separable.
(c) Find a one-way LOCC measurement from Alice to Bob that perfectly determines which of the four states they share.
(d) Show that there is no one-way LOCC measurement from Bob to Alice that can perfectly determine which of the four states they share.
Hint: Show that the choice of measurement for Alice can not depend on the outcome of Bob if she wants to perfectly distinguish the remaining states on her qubit.

3. *(4 points)* **Operations on PPT states:** Recall from Corollary 9.9 that a state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ has positive partial transpose (PPT) if $(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_{AB}] \geqslant 0$ where $\mathcal{T}[X] = X^T$ is the transpose map. Suppose that Alice and Bob share such PPT state.

(a) Show that if they apply a separable channel Ξ, the resulting state $\Xi[\rho_{AB}]$ is again PPT.
(b) Show that they cannot get a maximally entangled state

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{|\Sigma|}} \sum_{a \in \Sigma} |a\rangle_A \otimes |a\rangle_B$$

with any $|\Sigma| > 1$ by applying an LOCC operation on a PPT state ρ_{AB}.

Quantum Information Theory, Spring 2020

Homework problem set #11 **due May 11, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **Entanglement entropy and separable maps:** Let $|\Psi_1\rangle_{AB}$ be a pure state on registers A and B, and assume that it can be perfectly transformed to another state $|\Psi_2\rangle_{AB}$ by a separable operation. Show that such transformation cannot make the state more entangled in the sense of increasing its entanglement entropy. That is, show that
$$H(\rho_1) \geqslant H(\rho_2),$$
where $\rho_i = \mathrm{Tr}_B\left[|\Psi_i\rangle\langle\Psi_i|_{AB}\right]$ denotes the reduced state of $|\Psi_i\rangle_{AB}$ on Alice and $H(\rho)$ denotes the von Neumann entropy of ρ. *Hint: Entropy is a concave function.*

2. *(4 points)* **Local conversion with *no* communication:** Show that a pure state $|\Psi_1\rangle_{AB}$ shared by Alice and Bob can be converted to another pure state $|\Psi_2\rangle_{AB}$ using *only* local unitary operations (and *no* communication) if and only if
$$\rho_1 \prec \rho_2 \qquad \text{and} \qquad \rho_2 \prec \rho_1$$
where $\rho_i = \mathrm{Tr}_B\left[|\Psi_i\rangle\langle\Psi_i|_{AB}\right]$ denotes the reduced state of $|\Psi_i\rangle_{AB}$ on Alice. Show both directions of the implication.

3. *(4 points)* **Nielsen's theorem in action:** According to Nielsen's theorem, a maximally entangled state $|\Psi_1\rangle_{AB}$ shared between Alice and Bob can be transformed to any other shared pure state $|\Psi_2\rangle_{AB}$ of the same local dimensions by a one-way LOCC protocol from Bob to Alice. For each case below, devise an explicit one-way LOCC protocol that transforms $|\Psi_1\rangle_{AB}$ to $|\Psi_2\rangle_{AB}$ and succeeds with 100% probability. Write down the Kraus operators of Bob's instrument and the unitary corrections that Alice must apply after she receives Bob's measurement outcome.

   (a) Let $p \in [0,1]$ and
$$|\Psi_1\rangle_{AB} = \frac{1}{\sqrt{2}}\,|0\rangle \otimes |0\rangle + \frac{1}{\sqrt{2}}\,|1\rangle \otimes |1\rangle,$$
$$|\Psi_2\rangle_{AB} = \sqrt{p}\,|0\rangle \otimes |0\rangle + \sqrt{1-p}\,|1\rangle \otimes |1\rangle.$$

   (b) Let $p \in P(\mathbb{Z}_d)$ be an arbitrary probability distribution over $\mathbb{Z}_d = \{0,\dots,d-1\}$ and
$$|\Psi_1\rangle_{AB} = \sum_{i\in\mathbb{Z}_d} \frac{1}{\sqrt{d}}\,|i\rangle \otimes |i\rangle,$$
$$|\Psi_2\rangle_{AB} = \sum_{i\in\mathbb{Z}_d} \sqrt{p(i)}\,|i\rangle \otimes |i\rangle.$$

   Hint: Let $S : \mathbb{C}^{\mathbb{Z}_d} \to \mathbb{C}^{\mathbb{Z}_d}$ denote the cyclic shift operator that acts as $S|i\rangle = |i+1\rangle$ where "+" denotes addition modulo d. Notice that $\frac{1}{d}\sum_{a\in\mathbb{Z}_d} S^a p = u$ where p is the original probability distribution and $u = (1,\dots,1)/d$ is the uniform distribution on $\mathbb{Z}_d$.

4. *(2 bonus points)* 🖩 **Practice:** Implement a subroutine that, given two probability distributions p and q (not necessarily of the same length) determines whether $p \prec q$.

   (a) The file `abc.txt` contains three probability distributions: a, b, and c. Compare the distributions a and b using your subroutine and output `"a < b"`, `"b < a"`, or `"incomparable"`.

   (b) Use your subroutine to compare the distributions $a \otimes c$ and $b \otimes c$. Output `"a*c < b*c"`, `"b*c < a*c"`, or `"incomparable"`.

   (c) How can you interpret this outcome?

   (d) The files `psi1.txt` and `psi2.txt` contain bipartite pure states

   $$|\Psi_1\rangle_{AB} \in \mathbb{C}^5 \otimes \mathbb{C}^7 \qquad \text{and} \qquad |\Psi_2\rangle_{AB} \in \mathbb{C}^5 \otimes \mathbb{C}^9,$$

   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 $|\Psi_1\rangle_{AB}$ can be perfectly transformed into $|\Psi_2\rangle_{AB}$ by LOCC.

Quantum Information Theory, Spring 2020

Homework problem set #12 **due May 18, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **Fidelity inequality:** Let $\mathcal{H}$ be a Hilbert space with $\dim(\mathcal{H}) \geqslant 2$.

   (a) Let $|u_1\rangle, |u_2\rangle, |v\rangle \in \mathcal{H}$ be arbitrary pure quantum states. Show that

   $$|\langle u_1|v\rangle|^2 + |\langle u_2|v\rangle|^2 \leqslant 1 + |\langle u_1|u_2\rangle|.$$

   Hint: Upper bound the left-hand side by the largest eigenvalue of some rank-2 matrix. Compute this eigenvalue to get the right-hand side.

   (b) Let $\rho_1, \rho_2, \sigma \in D(\mathcal{H})$ be arbitrary states. Show that

   $$F(\rho_1, \sigma)^2 + F(\rho_2, \sigma)^2 \leqslant 1 + F(\rho_1, \rho_2).$$

2. *(4 points)* **Entanglement cost using compression and teleportation:**
   In this exercise you will give an alternative proof for the fact that the entanglement cost is at most the entanglement entropy for a pure state. Let $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a pure state.

   (a) Let ρ_A and ρ_B be its reduced density matrices and let $\alpha > H(\rho_A) = H(\rho_B)$. Show, using compression, that for all $\delta > 0$ and all but finitely many n there exists an LOCC protocol which converts $\phi^{\otimes \lfloor \alpha n \rfloor}$ into a state $\tilde{\rho}_n$ with $F(\rho_{AB}^{\otimes n}, \tilde{\rho}_n) > 1 - \delta$.
   Hint: Use teleportation!

   (b) Use (a) to show that $E_C(\rho_{AB}) \leqslant H(\rho_A)$, for every pure state $\rho_{AB} \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$.

3. *(4 points)* **Entanglement rank and the fidelity with the maximally entangled state:**
   Let $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^{\otimes n}$ and $\phi_n \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ be the canonical maximally entangled state of dimension n, i.e., $\phi_n = |\Phi_n^+\rangle\langle\Phi_n^+|$ where

   $$|\Phi_n^+\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |i\rangle \otimes |i\rangle.$$

   Show that $F(\sigma, \phi_n)^2 \leqslant r/n$, for any (mixed) state $\sigma \in D(\mathcal{H}_A \otimes \mathcal{H}_B)$ of entanglement rank r.

Quantum Information Theory, Spring 2020

Homework problem set #13 **due May 27, 2020**

Rules: Always explain your solutions carefully. You can work in groups, but must write up your solutions alone. You must submit your solutions before the Monday lecture (in person or by email).

1. *(4 points)* **Rényi-2 entropy:** In this problem you will study a new entropy measure called the *Rényi-2 entropy*. It is defined by $H_2(\rho) := -\log \operatorname{Tr}[\rho^2]$ for any quantum state $\rho \in D(\mathbb{C}^d)$.

   (a) Find a formula for $H_2(\rho)$ in terms of the eigenvalues of ρ.
   (b) Show that $H_2(\rho) \leqslant H(\rho)$ by using Jensen's inequality.
   (c) Show that $\operatorname{Tr}[\rho^2] = \operatorname{Tr}[F\rho^{\otimes 2}]$, where $F : |i\rangle \otimes |j\rangle \mapsto |j\rangle \otimes |i\rangle$ for all $i, j \in \{1, \ldots, d\}$, is the *swap operator*.

2. *(4 points)* **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
   $$\int |\psi_{AB}\rangle^{\otimes 2} \langle \psi_{AB}|^{\otimes 2} \, d\psi_{AB} = \frac{1}{d_A d_B (d_A d_B + 1)} \left(I_{AA} \otimes I_{BB} + F_{AA} \otimes F_{BB} \right).$$
   (b) Verify that $\int \operatorname{Tr}[\rho_A^2] \, d\psi_{AB} = \frac{d_A + d_B}{d_A d_B + 1}$.
   (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 Problem 1 and Jensen's inequality.

3. *(4 points)* **Haar measure:** In the exercise class, 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_π, $\pi \in S_n$.
   (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 $\mathcal{H}^{\otimes 2}$, $\alpha = \frac{d}{d^3 - d} \operatorname{Tr}[A] - \frac{1}{d^3 - d} \operatorname{Tr}[FA]$, and $\beta = \frac{d}{d^3 - d} \operatorname{Tr}[FA] - \frac{1}{d^3 - d} \operatorname{Tr}[A]$.