Quantum Information Theory (NMMB537)

Peter Zeman

January 8, 2026

Abstract

These are lecture notes for the course NMMB537 Quantum Information Theory, taught at Faculty of Mathematics and Physics, Charles University. The lecture notes are mostly inspired by the course NMAK14020U Quantum Information Theory, offered at University of Copenhagen.

Contents

Chapter 1

Quantum States and Measurements

We start by describing a precise mathematical framework for quantum states.

1.1 Finite-dimensional Hilbert Spaces

Axiom 1 (Hilbert space)**.** To every quantum system we asocciate a Hilbert space $\mathcal{H}$.

We only consider finite-dimensional Hilbert spaces in this course. If $\dim(\mathcal{H}) = d$, then $\mathcal{H}$ is isomorphic to

$$\mathbb{C}^d = \left\{ \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_d \end{pmatrix} : \psi_i \in \mathbb{C} \right\}$$

with the standard inner product

$$\langle \psi | \phi \rangle = \sum_{i=1}^{d} \overline{\psi_i} \phi_i.$$

Bra-ket notation. We write $|\psi\rangle \in \mathcal{H}$ for any vector. By $\langle \psi | \in \mathcal{H}^\dagger$, we mean the linear map $\mathcal{H} \to \mathbb{C}$ defined by $\langle \psi | (|\phi\rangle) = \langle \psi | \phi \rangle$. In other words, $\langle \psi |$ is the row vector corresponding to the conjugate transpose of $|\psi\rangle$. We denote the standard basis of $\mathbb{C}^d$ by $|0\rangle, |1\rangle, \ldots, |d\rangle$. That is, $|i\rangle$ has 1 on the i^{th} coordinate and 0 elsewhere.

Linear algebra notation. Every basis that we will consider will be orthogonal. If there is no confusion, we will use the terms linear map and matrix interchangeably.

By $\mathrm{Lin}(\mathcal{H}, \mathcal{K})$, we denote the set of linear maps $\mathcal{H} \to \mathcal{K}$. When we fix the standard bases, we can express $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ as

$$M = \sum_{i,j} M_{ij} |i\rangle\langle j|.$$

Note that $\langle i | M | j \rangle = M_{ij} \in \mathbb{C}$.

Exercise 1.1. Write $\sum_{i,j \in \{0,1\}} ij |i\rangle\langle j|$ as a matrix.

If $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{H}) = \mathrm{Lin}(\mathcal{H})$, then the *trace* with respect to the standard basis is

$$\mathrm{tr}[M] = \sum_i M_{ii} = \sum_i \langle i|M|i\rangle.$$

An important fact about trace is that it is in fact independent of the basis. Recall also that $\mathrm{tr}[MN] = \mathrm{tr}[NM]$.

Here are some important types of matrices:

- $M \in \mathrm{Lin}(\mathcal{H})$ is *Hermitian* if $M^\dagger = M$.
- $P \in \mathrm{Lin}(\mathcal{H})$ is *positive* if $\langle\psi|P|\psi\rangle \geq 0$, for all $|\psi\rangle \in \mathcal{H}$. $\mathrm{PSD}(\mathcal{H})$ is the class of all positive matrices. We also use the notation $P \geq 0$ to indicate that P is positive. If $M, N \in \mathrm{Lin}(\mathcal{H})$, then we define $M \geq N$ if $M - N \geq 0$, that is, if $M - N$ is positive.
- $U \in \mathrm{Lin}(\mathcal{H})$ is *unitary* if $U^\dagger U = UU^\dagger = I$. $\mathcal{U}(\mathcal{H})$ is the class of all unitary matrices.
- $V \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ is an *isometry* if $V^\dagger V = I$.
- $P \in \mathrm{Lin}(\mathcal{H})$ is a *projection* if $P^\dagger = P$ and $P^2 = P$.

Exercise 1.2. Are the following matrices $\{\text{Hermitian}, \text{PSD}, \text{unitary}, \text{projections}\}$?

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \begin{pmatrix} e^{i\frac{\pi}{4}} & 0 \\ 0 & e^{-i\frac{\pi}{4}} \end{pmatrix}.$$

Theorem 1.3 (Spectral theorem for Hermitian matrices)**.** *If $M \in \mathrm{Lin}(\mathcal{H})$ is Hermitian $d \times d$ matrix, then M has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and there is a basis of eigenvectors $\{|\psi_i\rangle\}_{i=1}^d$ such that*

$$M = \sum_{i=1}^d \lambda_i |\psi_i\rangle\langle\psi_i|.$$

Recall that the eigenvalues can have multiple occurences and the basis of eigenvectors may not be unique. Indeed, for instance,

$$I = \sum_{i=1}^d |\psi_i\rangle\langle\psi_i|$$

is true for any basis $\{|\psi_i\rangle\}_{i=1}^d$.

This allows us to define functions of Hermitian matrices. If $f\colon \mathbb{R} \to \mathbb{R}$ is a function, then we can define

$$f(M) = \sum_{i=1}^d f(\lambda_i)|\psi_i\rangle\langle\psi_i|.$$

For instance,

$$\sqrt{M} = \sum_{i=1}^d \sqrt{\lambda_i}|\psi_i\rangle\langle\psi_i|.$$

Note that if P is a projection, then the condition $P^2 = P$ implies that the eigenvalues are in $\{0, 1\}$. Thus, we can write $P = \sum_{i=1}^r |\psi_i\rangle\langle\psi_i|$. Here, $r = \mathrm{rank}(P)$ of the projection P.

Theorem 1.4 (Characterization of positive matrices)**.** *If $P \in \mathrm{Lin}(\mathcal{H})$, then the following are equivalent:*

(a) P is positive, i.e., $\langle\psi|P|\psi\rangle \geq 0$, for all $|\psi\rangle \in \mathcal{H}$.

(b) $P^\dagger = P$ and all eigenvalues are nonnegative.

(c) There is $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ such that $P = M^\dagger M$.

(d) $\mathrm{tr}[PQ] \geq 0$, for every $Q \in \mathrm{PSD}(\mathcal{H})$.

Exercise 1.5. Prove that if $P \in \mathrm{PSD}(\mathcal{H})$ and $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, then $MPM^\dagger$ is positive.

1.2 Quantum States

Definition 1.6. A *density matrix* is a positive operator $\rho \in \mathrm{PSD}(\mathcal{H})$ with $\mathrm{tr}[\rho] = 1$. We put $S(\mathcal{H}) = \{\rho \in \mathrm{PSD}(\mathcal{H}) : \mathrm{tr}[\rho] = 1\}$.

Axiom 2. The state of a quantum system with Hilbert space $\mathcal{H}$ is described by a density matrix ρ. We will refer to ρ as a quantum state.

Example 1.7 (classical states)**.** In probability theory we have a finite set of outcomes Σ with a probability distribution p such that $p(x) \geq 0$, for $x \in \Sigma$, and $\sum_{x\in\Sigma} p(x) = 1$. If $\mathcal{H}$ has a basis $\{|x\rangle\}_{x\in\Sigma}$, then

$$\rho = \sum_{x\in\Sigma} |x\rangle\langle x|$$

is a density matrix. Conversely, any diagonal density matrix corresponds to a probability distribution. □

Example 1.8 (pure states)**.** Let $|\psi\rangle \in \mathcal{H}$ be a unit vector. Then $\rho = |\psi\rangle\langle\psi|$ is a density matrix since $\mathrm{tr}[|\psi\rangle\langle\psi|] = \mathrm{tr}[\langle\psi|\psi\rangle] = 1$. If $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, then

$$|\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha|^2 & \alpha\overline{\beta} \\ \overline{\alpha}\beta & |\beta|^2 \end{pmatrix}.$$

Note that replacing $|\psi\rangle$ by $e^{i\theta}|\psi\rangle$ does not change the density matrix. □

Exercise 1.9. Is $\rho = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ a pure state?

Solution. Yes, since $\rho = |+\rangle\langle+|$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. □

Example 1.10 (general state)**.** Let ρ be a general quantum state. By Theorem 1.4, we can write

$$\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i|,$$

where $p_i \geq 0$. Since ρ is diagonal in the basis $\{|\psi_i\rangle\}_{i=1}^d$ and trace is independent of the chosen basis, $\mathrm{tr}[\rho] = 1$ implies that $\sum_{i=1}^d p_i = 1$. So, the eigenvalues p_i's define a probability distribution. We may interpret ρ as if we have the pure state $|\psi_i\rangle\langle\psi_i|$ with probability p_i. We call ρ a *mixed state* if it is not pure.

However, note that the decomposition of ρ given above is not unique. It is possible that ρ may be written as a sum over some different probabilities and different states, even different number of states. □

Exercise 1.11. What is the state corresponding to 50% chance of $|0\rangle$ and 50% chance of $|+\rangle$?

Solution.

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}.$$

Note that the eigenvalues of ρ are $1/2 + \sqrt{5}/4$ and $1/2 - \sqrt{5}/4$, so the spectral decomposition of ρ gives a completely different decomposition of ρ compared to the one we started with. □

Definition 1.12. The state $\tau = \frac{1}{d}I$ is called *the maximally mixed state.*

1.3 Bloch Sphere

How does the set $S(\mathcal{H})$ look like? For two states $\rho_1, \rho_2 \in S(\mathcal{H})$ and $t \in [0, 1]$, the convex combination $t\rho_1 + (1-t)\rho_2$ is also a state. In fact, we have the following lemma.

Lemma 1.13. *The set $S(\mathcal{H}) \subseteq \mathrm{Lin}(\mathcal{H})$ is convex. Moreoever, the extreme points of $S(\mathcal{H})$ are exactly the pure states.*

For the particular case of one qubit, there is a visualization known as the *Bloch sphere*. Hermitian 2×2 matrices have the following (real) basis:

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

The last three matrices are known as *Pauli matrices*. Here are some of their properties:

$$X^2 = Y^2 = Z^2 = I, \quad XY = iZ = -YX, \quad \mathrm{tr}[X] = \mathrm{tr}[Y] = \mathrm{tr}[Z] = 0.$$

In particular, Pauli matrices are unitary and have eigenvalues ± 1.

Every 2×2 Hermitian matrix can be written in a unique way as a linear combination of the matrices I, X, Y, Z with all coefficients being real numbers. An arbitrary 2×2 Hermitian matrix ρ with $\mathrm{tr}[\rho] = 1$ can be expressed as

$$\rho = \frac{1}{2}(I + xX + yY + zZ) = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}. \tag{1.3.1}$$

The coefficient $1/2$ in front of the I is fixed by the condition $\mathrm{tr}[\rho] = 1$. Further, $\mathrm{tr}[\rho] = 1$ implies that the eigenvalues of ρ are λ and $1 - \lambda$, for some $\lambda \in \mathbb{R}$. If we want ρ to be a quantum state, it also has to be positive, which is equivalent to requiring that $\lambda \geq 0$ and $1 - \lambda \geq 0$. Now, λ and $1 - \lambda$ are nonnegative if and only if $\det(\rho) = \lambda(1 - \lambda) \geq 0$. We compute

$$\det(\rho) = \frac{1}{4}((1+z)(1-z) - (x+iy)(x-iy)) = \frac{1}{4}(1 - x^2 - y^2 - z^2).$$

If we let $r = (x, y, z) \in \mathbb{R}^3$, then $\rho \geq 0$ if and only if $\|r\| \leq 1$. If ρ is a pure state, then $\mathrm{rank}(\rho) = 1$ and $\det(\rho) = 0$, which is equivalent to $\|r\| = 1$. Thus, pure states correspond exactly to the unit sphere in $\mathbb{R}^3$ and mixed states correspond to the open unit ball.

Exercise 1.14. What is the pure state $|\psi\rangle$ and its density matrix $|\psi\rangle\langle\psi|$ corresnponding to $r_1 = (0, 1, 0)$ and to $r_2 = (0, -1, 0)$.

Solution. By plugging r_1 into Eq. (1.3.1), we get

$$\rho_1 = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}.$$

We can easily calculate, that the eigenvector corresnponding to the eigenvalue 1 of ρ_1 is $|\psi_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$. By Theorem 1.3, we get $\rho_1 = |\psi_1\rangle\langle\psi_1|$. Similarly, we get that $\rho_2 = |\psi_2\rangle\langle\psi_2|$, where $|\psi_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle)$. □

We have the following important bases of pure states that correspond to the intersection points of x, y, z-axes with the Bloch sphere:

- Z-basis: $\{|0\rangle, |1\rangle\}$, $r = (0, 0, 1)$, $-r = (0, 0, -1)$.
- X-basis: $\{|+\rangle, |-\rangle\}$, $r = (1, 0, 0)$, $-r = (-1, 0, 0)$
- Y-basis: $\{|+i\rangle, |-i\rangle\}$, $r = (0, 1, 0)$, $-r = (0, -1, 0)$.

1.4 Measurements

Definition 1.15. A *measurement* or a *positive operator valued measure (POVM)* on a Hilbert space $\mathcal{H}$ with a finite set of outcomes Ω is a function $\mu\colon \Omega \to \mathrm{PSD}(\mathcal{H})$ such that $\sum_{x\in\Omega} \mu(x) = I$. If $\mu(x)$ are projections, then the measurement is called *projective*.

Axiom 3. If we measure a state ρ using μ, then the probability of getting an outcome $x \in \Omega$ is $p(x) = \mathrm{tr}[\mu(x)\rho]$.

Remark 1.16. In general in quantum mechanics, it is not possible to perform measurements on a state without changing it. However, for now, we will not consider what happens with a state after the measurement (we can think that it is destroyed). We only get classical information about which outcome occured. We will discuss post-measurement states later in the course.

Exercise 1.17. Prove that if ρ is a quantum state and μ is a measurement, then $p(x) = \mathrm{tr}[\mu(x)\rho]$ defines a probability distribution on Ω.

Solution. By Theorem 1.4(d), we have $\mathrm{tr}[\mu(x)\rho] \geq 0$ since both $\mu(x)$ and ρ are positive. By the linearity of trace,

$$\sum_{x\in\Omega} \mathrm{tr}[\mu(x)\rho] = \mathrm{tr}[\sum_{x\in\Omega} \mu(x)\rho] = \mathrm{tr}[I\rho] = 1.$$

□

Example 1.18 (basis measurement)**.** Let $\{|\psi_i\rangle\}_{i=1}^d$ be an orthonormal basis and let $\mu(i) = |\psi_i\rangle\langle\psi_i|$ be the projection onto $|\psi_i\rangle$. We get

$$p(i) = \mathrm{tr}[|\psi_i\rangle\langle\psi_i|\rho] = \mathrm{tr}[\langle\psi_i|\rho|\psi_i\rangle] = \langle\psi_i|\rho|\psi_i\rangle.$$

In particular, if $\rho = |\phi\rangle\langle\phi|$ is a pure state, then we get

$$p(i) = \langle\psi_i||\phi\rangle\langle\phi||\psi_i\rangle = |\langle\psi_i|\phi\rangle|^2.$$

□

Exercise 1.19. What are the outcome probabilities when we measure $|0\rangle$ the X-basis?

Solution. We get $|+\rangle$ with probability $|\langle+|0\rangle|^2 = 1/2$ and $|-\rangle$ with probability $|\langle-|0\rangle|^2 = 1/2$. □

1.5 Bloch Sphere Revisited

Given $r = (x, y, z)$ with $\|r\| = 1$, the positive matrices

$$\mu(0) := \rho(r) = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} \quad \text{and} \quad \mu(1) := \rho(-r) = \frac{1}{2}\begin{pmatrix} 1-z & -x+iy \\ -x-iy & 1+z \end{pmatrix}$$

define a general qubit basis measurement. Performing this measurement on a quantum state with Bloch vector $s = (x', y', z')$, then the probability of obtaining the outcome 0 is

$$p(0) = \frac{1}{2} + \frac{1}{2}(r \cdot s) = \frac{1}{2} + \frac{1}{2}(xx' + yy' + zz').$$

Geometrically $r \cdot s$ is the projection of the Bloch vector s of a quantum state onto the axis defining the measurment.

1.6 Observables

Sometimes it is convenient to reformulate measurements in terms of *observables*. A *projective measurement* is a measurement where all measurement operators are projections, that is, we have a set of outcomes Ω and projections $\{P_x : x \in \Omega\}$ such that $\sum_{x\in\Omega} P_x = I$. Suppose that $\Omega \subseteq \mathbb{R}$. The *observable* associated to the projective measurment is an operator $O \in \mathrm{Lin}(\mathcal{H})$ given by

$$O = \sum_{x\in\Omega} xP_x.$$

Since x are real, O is Hermitain, and, conversely, each Hermitian operator O has a spectral decomposition of this form, thereby defining a projective measurement.

Remark 1.20. In physics, observables are often the preferred way to reason about measurements. Measuring the observable O is the same as performing the projective measurement defined by the spectral decomposition of O.

Example 1.21. Since $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$, the observable Z corresponds to measuring in the Z-basis with outcomes $\{\pm 1\}$. Similarly, since $X = |+\rangle\langle +| - |-\rangle\langle -|$, measuring X corresponds to measuring in the X-basis with outcomes $\{\pm 1\}$. More generally, we may measure a qubit in the basis given by the antipodal points $r = (x, y, z)$ and $-r$ on the Bloch sphere with outcomes $\{\pm 1\}$. This gives an observable

$$\begin{aligned} O(r) = \rho(r) - \rho(-r) &= \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1-z & -x+iy \\ -x-iy & 1+z \end{pmatrix} \\ &= \begin{pmatrix} z & x-iy \\ x+iy & -z \end{pmatrix}. \end{aligned}$$

□

An observable provides a compact formula for the expectation value of the measurement outcome. If $O = \sum_{x\in\Omega} xP_x$ is an observable and $\rho \in S(\mathcal{H})$, then

$$\mathbb{E}(\text{outcome}) = \sum_{x\in\Omega} x\mathbb{P}(\text{outcome } x) = \mathrm{tr}[O\rho].$$

In case $\rho = |\psi\rangle\langle\psi|$, then the right-hand hand side is equal to $\langle\psi|O|\psi\rangle$.

1.7 Uncertainty Relation

Let $|\psi\rangle$ be a pure state. Note that since $X = |+\rangle\langle+| - |-\rangle\langle-|$, we get the following:

$$|\langle\psi|X|\psi\rangle| = |p_X(1) - p_X(-1)| = |2p_X(1) - 1| = 2\max\{p_X(1), p_X(-1)\} - 1,$$

where $p_X(x)$ denotes the probability of getting outcome x after measuring in the X-basis. Clearly,

$$0 \leq |\langle\psi|X|\psi\rangle| \leq 1.$$

The upper bound is attained precisely when either $p_X(1) = 1$ or $p_X(-1) = 1$, that is, when the measurement outcome is certain. The lower bound is attained precisely when $p_X(1) = p_X(-1) = 1/2$, which means that the measurement is completely uncertain. Thus, $|\langle\psi|X|\psi\rangle|$ provides a meaningful way to quantify our uncertainty about the measurement outcome.

Using $Z = |0\rangle\langle0| - |1\rangle\langle1|$, we similarly obtain $|\langle\psi|Z|\psi\rangle| = 2\max\{p_Z(1), p_Z(-1)\} - 1$. Combining the two, we get

$$|\langle\psi|X|\psi\rangle| + |\langle\psi|Z|\psi\rangle| \leq 2.$$

Note that by the previous discussion, the equality can never be attained. However, there is a significant strenghtening:

Theorem 1.22 (Uncertainty relation for Pauli matrices)**.** *For every state $|\psi\rangle$, we have*

$$|\langle\psi|X|\psi\rangle| + |\langle\psi|Z|\psi\rangle| \leq \sqrt{2}.$$

Proof. Let $s_X, s_Z \in \{\pm1\}$ and let $A = s_X X + s_Z Z$. We need to show that

$$s_X\langle\psi|X|\psi\rangle + s_Z\langle\psi|Z|\psi\rangle = \langle\psi|A|\psi\rangle \leq \sqrt{2}.$$

Using Cauchy-Schwarz inequality, we get

$$\langle\psi|A|\psi\rangle \overset{\text{CS}}{\leq} \|A|\psi\rangle\| \leq \|A\|,$$

where $\|A\| = \sup_{\||\psi\rangle\|=1} \|A|\psi\rangle\|$ is the operator norm of A. Further,

$$A^\dagger A = A^2 = (s_X X + s_Z Z)(s_X X + s_Z Z) = I + s_X s_Z (XZ + ZX) + I = 2I.$$

This calculation shows that $A/\sqrt{2}$ is unitary. Since the operator norm of a unitary is one, we get

$$\|A\| = \sqrt{2}\left\|A/\sqrt{2}\right\| = \sqrt{2}.$$

□

We can interpret the quantity $\max\{p_X(1), p_X(-1)\}$ as the *guessing probability* $p_{\text{guess},X}$, that is, the maximal probability of guessing the outcome of X-basis measurement on the state $|\psi\rangle$ – the best option is to just guess the outcome with larger probability. Using this notation, we can rewrite the uncertainty relation from Theorem 1.22 as follows:

$$\begin{aligned} |\langle\psi|X|\psi\rangle| + |\langle\psi|Z|\psi\rangle| &\leq \sqrt{2} \\ 2\max\{p_X(1), p_X(-1)\} - 1 + 2\max\{p_Z(1), p_Z(-1)\} - 1 &\leq \sqrt{2} \\ p_{\text{guess},X} + p_{\text{guess},Z} &\leq 1 + \frac{\sqrt{2}}{2}. \end{aligned}$$

So the uncertainty relation from Theorem 1.22 gives a bound on the sum of probabilities of guessing the two measurement outcomes correctly.

1.8 TODO

- simultaneously diagonalizable operators
- more about uncertainty principle

Chapter 2

Multiple Quantum Systems

2.1 Multiple Systems

If we have classical random variables with outcome sets $\Sigma_1, \ldots, \Sigma_n$, then their joint distribution is a probability distribution on the product set

$$\Sigma_1 \times \cdots \times \Sigma_n = \{(x_1, \ldots, x_n) : x_j \in \Sigma_j\}.$$

So n bits are represented by an n-tuples.

Tensor product is precisely the quantum version of this. If we have bases $\Sigma_1, \ldots, \Sigma_n$ for the Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$, then the Hilbert space

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

is the *tensor product* with basis

$$\{|x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle : |x_j\rangle \in \Sigma_j\}.$$

We will sometimes use the shorthand $|x_1\rangle|x_2\rangle \cdots |x_n\rangle$ or $|x_1 \cdots x_n\rangle$. Clearly, we have $\dim(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n) = \dim(\mathcal{H}_1) \cdots \dim(\mathcal{H}_n)$ since the basis elements are labelled by the elements of $\Sigma_1 \times \cdots \times \Sigma_n$.

Axiom 4. If we have multiple quantum systems with Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$, then the joint system has associated Hilbert space $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$.

Example 2.1. If we have two qubits, the joint Hilbert space is $\mathbb{C}^2 \otimes \mathbb{C}^2$, which is 4-dimensional and has the standard basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$.

The Hilbert space of n qubits is $(\mathbb{C}^2)^{\otimes n}$, which is 2^n-dimensional and has the standard basis $\{|x_1 \cdots x_n\rangle\}_{x_1,\ldots,x_n \in \{0,1\}}$. □

Notation. We will often label different quantum systems by $A, B, C, \ldots$ and denote the associated Hilbert spaces as $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C, \ldots$. The bases of these Hilbert spaces are usually denoted by $\Sigma_A, \Sigma_B, \Sigma_C, \ldots$. We will also write $\mathrm{Lin}(A) = \mathrm{Lin}(\mathcal{H}_A)$, $S(A) = S(\mathcal{H}_A)$, $\mathrm{PSD}(A) = \mathrm{PSD}(\mathcal{H}_A)$. We also write for example AB instad of $\mathcal{H}_A \otimes \mathcal{H}_B$. So, we may write $\rho_{AB} \in S(AB)$ for a quantum state shared by Alice and Bob and μ_A for a measurement on Alice's quantum system.

Tensor product of operators. If $M_A \in \mathrm{Lin}(A)$ and $N_B \in \mathrm{Lin}(AB)$, then $M_A \otimes N_B \in \mathrm{Lin}(AB)$ is defined by extending the following formula by linearity:

$$(M_A \otimes N_B)|\psi_A\rangle \otimes |\phi_B\rangle = (M_A|\psi_A\rangle) \otimes (N_B|\phi_B\rangle),$$

where $|\psi_A\rangle = \sum_a \psi_a|a\rangle$, $|\phi_B\rangle = \sum_b \phi_b|b\rangle$, and $|\psi_A\rangle|\psi_B\rangle = \sum_{a,b} \psi_a\phi_b|ab\rangle$.

Example 2.2. If $M_A = |a\rangle\langle a'|$ and $N_B = |b\rangle b'$, then $M_A \otimes N_B = |a\rangle\langle a'| \otimes |b\rangle\langle b'| = |ab\rangle\langle a'b'|$.

Definition 2.3. Let A and B be quantum systems. States of the form $\rho = \rho_A \otimes \rho_B \in S(AB)$, for $\rho_A \in S(A)$ and $\rho_B \in S(B)$ are called *product states*. A state, which is not a product state is called *correlated*.

Note that the previous definition makes sense since the tensor product of positive operators is positive and $\mathrm{tr}[M_A \otimes N_B] = \mathrm{tr}[M_A]\,\mathrm{tr}[N_B]$.

Example 2.4 (classical states)**.** A *joint probability distribution* $p_{XY} \in \mathbb{P}(XY)$ associates a probability $p_{XY}(x, y)$ to each pair $(x, y) \in \Sigma_X \times \Sigma_Y$. The classical state corresponding to XY has the density matrix

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

Moreover, any classical joint state is of this form. The state ρ_{XY} is a product state if and only if X and Y are indpendent under the probability distribution p_{XY}. Thus, one can think of product states as the quantum generalization of independence in probability theory. Most of the quantum states are neither classical nor product states. □

Example 2.5 (maximally correlated state)**.** A classical state that is not a product state is for example the *maximally correlated state*:

$$\sigma_{AB} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) = \frac{1}{2}(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|).$$

Writing this in the basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ gives

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

□

Exercise 2.6. Let

$$\rho_A = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \rho_B = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Express $\rho_A \otimes \rho_B$ in the standard basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$.

Solution.

$$\rho_A \otimes \rho_B = \frac{1}{4}\begin{pmatrix} 1\rho_B & 1\rho_B \\ 1\rho_B & 1\rho_B \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 & & 1 & \\ & 1 & & 1 \\ 1 & & 1 & \\ & 1 & & 1 \end{pmatrix}.$$

□

Example 2.7 (pure product states)**.** If $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$, then pure product states are for example the basis states $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$. Another example is the state $|+\rangle \otimes |+\rangle$. □

We will use abbreviation $|\psi_A\rangle|\phi_B\rangle = |\psi_A\rangle \otimes |\phi_B\rangle$ for pure states.

Example 2.8 (maximally entangled state)**.** A state that is neither classical nor a product state is the *maximally entangled state*:

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

The density matrix is

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

Writing this in the basis $|00\rangle, |01\rangle, |10\rangle, |11\rangle$ gives

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

Recall that $|xy\rangle = |x\rangle \otimes |y\rangle$, so we can also write

$$\rho_{AB} = \frac{1}{2}(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|).$$

□

2.2 The Partial Trace

If Alice does a measurement $\mu_A \colon \Omega \to \mathrm{PSD}(A)$, how to describe this on AB? Define

$$\mu_A \otimes I_B := \{\mu_A \otimes I_B : x \in \Omega\}.$$

This is a measurement on the whole system since tensor products of positive operators are positive and

$$\sum_{x\in\Omega} \mu_A(x) \otimes I_B = \left(\sum_{x\in\Omega} \mu_A(x)\right) \otimes I_B) = I_A \otimes I_B = I_{AB}.$$

More generally, suppose that Alice and Bob have measurements $\mu_A \colon \Omega_1 \to \mathrm{PSD}(A)$ and $\mu_B \colon \Omega_2 \to \mathrm{PSD}(B)$. This corresponds to a measurement on AB given by

$$\mu_A \otimes \mu_B := \{\mu_A(x_1) \otimes \mu_B(x_2) : (x_1, x_2) \in \Omega_1 \times \Omega_2\}.$$

The outcome probabilities for Alice do not depend on the choice of measurement for Bob.

Given a state $\rho_{AB} \in S(AB)$, what is the state ρ_A of Alice? We would like

$$\mathrm{tr}[\mu_A(x)\rho_A] = \mathrm{tr}[(\mu_A(x) \otimes I_B)\rho_{AB}],$$

for any measurement operator $\mu_A(x)$. We can proceed by computing the trace on the RHS. We chose bases $\{|a\rangle\}_{a\in\Sigma_A}$ and $\{|b\rangle\}_{b\in\Sigma_B}$ of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, which gives the product basis $\{|a\rangle\otimes|b\rangle\}$ for $\mathcal{H}_A\otimes\mathcal{H}_B$. We can expand the RHS above

$$\begin{aligned}\operatorname{tr}[(\mu_A(x)\otimes I_B)\rho_{AB}] &= \sum_{\substack{a\in\Sigma_A\\ b\in\Sigma_B}} \langle ab|(\mu_A(x)\otimes I_B)\rho_{AB}|ab\rangle \quad \text{use } |a\rangle\otimes|b\rangle = (I_A\otimes|b\rangle)|a\rangle \\ &= \sum_{a,b}\langle a|(I_A\otimes\langle b|)(\mu_A(x)\otimes I_B)\rho_{AB}(I_A\otimes|b\rangle)|a\rangle \\ &= \sum_{a,b}\langle a|\mu_A(x)(I_A\otimes\langle b|)\rho_{AB}(I_A\otimes|b\rangle)|a\rangle \\ &= \sum_{a}\langle a|\mu_A(x)\left(\sum_b (I_A\otimes\langle b|)\rho_{AB}(I_A\otimes|b\rangle)\right)|a\rangle.\end{aligned}$$

Finally, we define

$$\rho_A = \sum_b (I_A\otimes\langle b|)\rho_{AB}(I_A\otimes|b\rangle).$$

Definition 2.9 (partial trace)**.** Let A and B be systems with Hilbert spaces $\mathcal{H}_A$ nad $\mathcal{H}_B$ and choose basis $\{|b\rangle\}_{b\in\Sigma_B}$ for $\mathcal{H}_B$. Let $M_{AB}\in\operatorname{Lin}(AB)$. Then the *partial trace over B* of M_{AB} is

$$\operatorname{tr}_B[M_{AB}] = \sum_b (I_A\otimes\langle b|)M_{AB}(I_A\otimes|b\rangle).$$

For $\rho_{AB}\in S(AB)$, we call $\rho_A = \operatorname{tr}_B[\rho_{AB}]\in S(A)$ the *reduced state* of ρ_{AB} on A.

If we choose a basis $|a\rangle$ for $\mathcal{H}_A$, then the entries of the partial trace are given by

$$\langle a|\operatorname{tr}_B[M_{AB}]|a'\rangle = \langle a|\sum_b (I_A\otimes\langle b|)M_{AB}(I_A\otimes|b\rangle)|a'\rangle = \sum_b \langle ab|M_{AB}|a'b\rangle.$$

Further, we can write

$$M_{AB} = \sum_{a,a'\in\Sigma_A}\sum_{b,b'\in\Sigma_B} M_{ab,a'b'}|ab\rangle\langle a'b'| = \sum_{a,a'\in\Sigma_A}\sum_{b,b'\in\Sigma_B} M_{ab,a'b'}|a\rangle\langle a'|\otimes|b\rangle\langle b'|.$$

The partial trace is then

$$\operatorname{tr}_B[M_{AB}] = \sum_{a,a',b} M_{ab,a'b}|a\rangle\langle a'| = \sum_{a,a'}\left(\sum_b M_{ab,a'b}\right)|a\rangle\langle a'|.$$

For tensor product of operators $M_{AB} = N_A\otimes O_B$, the partial trace is given by

$$\operatorname{tr}_B[N_A\otimes O_B] = N_A\operatorname{tr}[O_B] = \operatorname{tr}[O_B]N_A.$$

This follows from the above since $M_{ab,a'b'} = N_{a,a'}O_{b,b'}$ and $\sum_b M_{ab,a'b} = M_{a,a'}\operatorname{tr}[O_B]$. Every operator can be written as a linear combination of tensor product operators, so this formula is sufficient to compute partial traces of arbitrary operators. Moreover, it shows that the notation of the reduced state is compatible with the notation for product states: $\rho_{AB} = \rho_A\otimes\rho_B$, then ρ_A and ρ_B are reduced states of A and B, respectively.

Lemma 2.10 (properties of partial trace)**.** *(a) The map* $\mathrm{tr}_B\colon \mathrm{Lin}(AB) \to \mathrm{Lin}(A)$ *is linear.*

(b) For $N_A \in \mathrm{Lin}(A)$ *and* $M_{AB} \in \mathrm{Lin}(AB)$*, we have* $\mathrm{tr}[(N_A \otimes I_B)M_{AB}] = \mathrm{tr}[N_A \,\mathrm{tr}_B[M_{AB}]]$.

(c) The partial trace does not depend on the choice of basis Σ_B.

(d) For $M_{AB} \in \mathrm{Lin}(AB)$*, we have* $\mathrm{tr}[\mathrm{tr}_B[M_{AB}]] = \mathrm{tr}[M_{AB}]$.

(e) If $P_{AB} \in \mathrm{PSD}(AB)$*, then* $\mathrm{tr}_B[P_{AB}] \in \mathrm{PSD}(A)$.

Proof. (a) The formula in the definition is linear in M_{AB}.

(b) We can follow the same calculation that we used in the derivation of partial trace with M_{AB} instead of ρ_{AB} and N_A instead of $\mu_A(x)$.

(c) Consider the formula in (b). Since LHS does not depend on the choice of basis, neither does the right-hand side. From linear algebra, we know that $\mathrm{tr}[AX] = \mathrm{tr}[BX]$ for all $X \in \mathrm{Lin}(\mathcal{H})$ if and only if $A = B$. Since the equation in (b) holds for all N_A, using the fact, $\mathrm{tr}_B[M_{AB}]$ is determined.

(d) Immedietelly from (b) by using $N_A = I_A$.

(e) By Theorem 1.4(d) it suffices to check that

$$\mathrm{tr}[Q_A \,\mathrm{tr}_B[P_{AB}]] \geq 0, \quad \text{for all } Q_A \in \mathrm{PSD}(A).$$

If $Q_A \geq 0$, then $Q_A \otimes I_B \geq 0$. So we can apply first (b) and then Theorem 1.4(d):

$$\mathrm{tr}[Q_A \,\mathrm{tr}_B[P_{AB}]] = \mathrm{tr}[(Q_A \otimes I_B)P_{AB}] \geq 0.$$

□

Example 2.11. Let $|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and $\rho_{AB} = |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|$. Then

$$\rho_A = \mathrm{tr}_B[\rho_{AB}] = \mathrm{tr}_B[\frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}] = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This is the maximally mixed state. This also discribes the situation where Alice has a classical bit equal to zero or one with equal probability. If Alice cannot communicate with Bob, she cannot see the difference. □

We have three sources of mixed states in the quantum formalism: probabilistic mixtures (we receive ρ_x with probability p(x)), restricting to subsystem (even if $|\psi_{AB}\rangle$ is pure, the reduced state ρ_A can be mixed), measurement (if we perform measurement μ on ρ, we have some probability distribution on Ω, which can be described by a classical state).

Example 2.12 (marginal distributions)**.** For classical states, the partial trace reduces to marginal distirbutions. Let

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

where p_{XY} is a joint probability distribution on $\Sigma_X \times \Sigma_Y$. The reduced state

$$\rho_X = \mathrm{tr}_Y[\rho_{XY}] = \sum_x \left(\sum_y p_{XY}(x,y)\right)|x\rangle\langle x| = \sum_x p_X(x)|x\rangle\langle x|,$$

where p_X is the *marginal distribution* of random variable X given by

$$p_X(x) = \sum_y p_{XY}(x, y).$$

The two random variables X and Y are idependent, that is, $p_{XY}(x, y) = p_X(x)p_Y(y)$, if and only if ρ_{XY} is a product state, that is, $\rho_{XY} = \rho_X \otimes \rho_Y$.

Another important concept in classical probability is the notion of conditional probability. In quantum information, there is no direct generalization of conditional probabilities for general quantum states. We come to this when dealing with quantum entorpy later.

2.3 Purification

The situation when with probability $p(x)$ we output a state ρ_x is described by the density matrix $\rho_A = \sum_{x\in\Omega} p(x)\rho_x$. However, given the density matrix, the interpretation can be non-unique.

Suppose now that we want to model a different situation, namely with probability $p(x)$ we output a state ρ_x, but now we also receive the value of x as well. This can be described by introducing a reference system X with Hilbert space $\mathbb{C}^\Omega$ and taking the joint state

$$\rho_{AX} = \sum_{x\in\Omega} p(x)\rho_x \otimes |x\rangle\langle x|.$$

It is easy to check that $\mathrm{tr}_A[\rho_{AX}] = \rho_A$. Once we actually recive outcome x, the state must be ρ_x (like in probability if we see outcome of die to be 6, then the die is in the state 6 with probability 1). We may consider X to be side information. It turns out that it is usefull to have quantum side information.

Definition 2.13. Given $\rho_A \in S(\mathcal{H}_A)$, a *purification* of ρ_A is a pure state $|\phi_{AR}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R$ such that

$$\mathrm{tr}_R[|\phi_{AR}\rangle\langle\phi_{AR}|] = \rho_A.$$

The system R is called a *reference* or *purifying system*. We will refer to both $|\phi_{AR}\rangle$ and $\rho_{AR} = |\phi_{AR}\rangle\langle\phi_{AR}|$ as purification of ρ_A.

Lemma 2.14. *Every $\rho_A \in S(A)$ has a purification. The dimension $|R|$ of the purifying system can be taken to be* $\mathrm{rank}(\rho_A)$.

Proof. Let $r = \mathrm{rank}(\rho_A)$ and $\rho_A = \sum_{j=0}^{r-1} p_j|e_j\rangle\langle e_j|$ be a spectral decomposition. Put $\mathcal{H}_R = \mathbb{C}^r$ and

$$|\phi_{AR}\rangle = \sum_{j=0}^{r-1} \sqrt{p_j}|e_j\rangle \otimes |j\rangle.$$

Then

$$\mathrm{tr}_R[|\phi_{AR}\rangle\langle\phi_{AR}|] = \mathrm{tr}_R\left[\sum_{j,k=0}^{r-1} \sqrt{p_j p_k}|e_j\rangle\langle e_k| \otimes |j\rangle\langle k|\right] = \sum_j^{r-1} p_j|e_j\rangle\langle e_j| = \rho_A.$$

□

The proof does notu seo orthogonality given by the spectral decomposition. So we may take any decomposition $\rho_A = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ and take $\sum_j \sqrt{p_j}|\psi_j\rangle \otimes |j\rangle$ as purification. This does not give the optimal dimension of the reference system.

Exercise 2.15. Compute the purification of $\rho_A = \frac{1}{3}(|0\rangle\langle 0| + |+\rangle\langle +| + |1\rangle\langle 1|)$.

Solution.

$$|\phi_{AR}\rangle = \frac{1}{\sqrt{3}}(|0\rangle|0\rangle + |+\rangle|1\rangle + |1\rangle|2\rangle).$$

□

Lemma 2.16. *If $|\phi_{AR}\rangle$ and $|\phi_{AS}\rangle$ are purifications of ρ_A and $|R| \leq |S|$, then there is an isometry $V_{R\to S} \in \mathrm{Isom}(R, S)$ such that*

$$(I_A \otimes V_{R\to S})|\phi_{AR}\rangle = |\phi_{AS}\rangle.$$

In particular, when $S = R$, then the purificaton is unique up to a unitary.

Proof. Uses Schmidt decomposition. □

2.4 Schmidt decomposition

A standard result in linear algebra is that every matrix has a *singular value decomposition (SVD)*. Let $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, there are bases $\{|e_j\rangle\}$ and $\{|f_j\rangle\}$ of $\mathcal{K}$ and $\mathcal{H}$, respectively, and $s_1 \geq \cdots \geq s_r > 0$, the *singular values* of M, such that

$$M = \sum_{j=1}^{r} s_j |e_j\rangle\langle f_j|.$$

The number of singular values equals $r = \mathrm{rank}(M)$.

To find SVD, we can take $M^\dagger M$, which is positive. It has eigenvalues s_j^2. The eigenvectors of $M^\dagger M$ and $MM^\dagger$ are the bases $\{|e_j\rangle\}$ and $\{|f_j\rangle\}$.

We may interpret a pure state $\psi_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ as a linear map $M \in \mathrm{Lin}(\mathcal{H}_A^*, \mathcal{H}_B)$ and apply the singular value decomposition to M. This leads to the Schmidt decomposition.

Theorem 2.17 (Schmidt decomposition)**.** *Let $|\psi_{AB}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a pure quantum state. Then there are bases $\{|e_j\rangle\}$ and $\{|f_j\rangle\}$ of $\mathcal{H}_A$ and $\mathcal{H}_B$, and positive number $s_1 \geq \cdots \geq s_r > 0$, where $r \leq \min(|A|, |B|)$, such that $\sum_j s_j^2 = 1$ and*

$$|\psi_{AB}\rangle = \sum_{j=1}^{r} s_j |e_j\rangle \otimes |f_j\rangle.$$

The numbers $s_1, \ldots, s_r$ are called Schmidt coefficients and r is calleed the Schmidt rank.

Proof. Let $\{|a\rangle\}$ and $\{|b\rangle\}$ be arbitrary basis for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Let M be the $|A| \times |B|$-matrix defined by $M_{ab} = \langle ab|\psi_{AB}\rangle$. We apply SVD to M and get

$$M = \sum_{j=1}^{r} s_j |e_j\rangle\langle f_j|.$$

Thus,

$$\begin{aligned} M_{ab} = \langle a|M|b\rangle &= \sum_{j=1}^{r} s_j \langle a|e_j\rangle\langle g_j|b\rangle \\ &= \sum_{j=1}^{r} s_j \langle a|e_j\rangle\overline{\langle b|g_j\rangle} \\ &= \sum_{j=1}^{r} s_j \langle a|e_j\rangle\langle b|f_j\rangle, \end{aligned}$$

here $|f_j\rangle$ denotes the vectors whose entries with respect to the basis $\{|b\rangle\}$ are the complex conjugate of the entries of $|g_j\rangle$, i.e., $\langle b|f_j\rangle = \overline{\langle b|g_j\rangle} = \langle g_j|b\rangle$, for all b. Note that $\{|f_j\rangle\}$ is also an orthonormal basis. We have

$$\begin{aligned} |\psi_{AB}\rangle &= \sum_{a,b} M_{ab}|a\rangle \otimes |b\rangle \\ &= \sum_{a,b}\sum_{j=1}^{r} s_j \langle a|e_j\rangle\langle b|f_j\rangle|a\rangle \otimes |b\rangle \\ &= \sum_{j=1}^{r}\sum_{a,b} s_j |a\rangle\langle a|e_j\rangle \otimes |b\rangle\langle b|f_j\rangle \\ &= \sum_{j=1}^{r} s_j |e_j\rangle \otimes |f_j\rangle. \end{aligned}$$

The state is normalized, so $\sum_j s_j^2 = 1$. □

Exercise 2.18. Compute the Schmidt decomposition of

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 \\ 1 \\ 2 \\ -2 \end{pmatrix}.$$

Solution. To do it systematically, reorganize the vector into the matrix M, e.g., we take $|01\rangle$ to $|0\rangle\langle 1|$. We get

$$M = \frac{1}{\sqrt{10}} \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$$

Then we compute the SVD.

To get the left singular vectors, we take

$$MM^\dagger = \begin{pmatrix} 1/5 & 0 \\ 0 & 4/5 \end{pmatrix}.$$

The eigenvalues of $MM^\dagger$ are $1/5$ and $4/5$, so the Schmidt coefficients are

$$s_1 = \frac{1}{\sqrt{5}} \quad \text{and} \quad s_2 = \frac{2}{\sqrt{5}}$$

and the eigenvectors are just the computational basis $|0\rangle$ and $|1\rangle$.

To get the right singular vectors, we take

$$M^\dagger M = \begin{pmatrix} 1/2 & -3/10 \\ -3/10 & 1/2 \end{pmatrix}.$$

We have eigenvalue $1/5$ with eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $4/5$ with eigenvector $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. After normalization, we get the right singular vectors $|+\rangle$ and $|-\rangle$. Finally, we get

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{5}}|0\rangle \otimes |+\rangle + \frac{2}{\sqrt{5}}|1\rangle \otimes |-\rangle.$$

□

Lemma 2.19. *If $|\phi_{AB}\rangle$ is a pure state with Schmidt decomposition*

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

then the reduced density matrices are given by

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

In particular, Schmidt rank and Schmidt coefficients are uniquely determined by the rank and the non-zero eigenvalues of the reduced states, respectively.

Proof. We compute

$$\rho_A = \mathrm{tr}_B[|\psi_{AB}\rangle\langle\psi_{AB}|] = \mathrm{tr}_B \left[\sum_{j,k=1}^{r} s_j s_k |e_j\rangle\langle e_k| \otimes |f_i\rangle\langle f_k| \right] = \sum_{j=1}^{r} s_j^2 |e_j\rangle\langle e_j|.$$

Similarly for ρ_B. □

Corollary 2.20. *A pure state $|\phi_{AB}\rangle$ with density matrix ρ_{AB} is a product state if and only if ρ_A is pure if and only if ρ_B is pure.*

2.5 Entanglement

Two random variables X and Y are independent if the corresponding classical state is a product state $\rho_{XY} = \rho_X \otimes \rho_Y$. If X and Y are not independent, the are correlated. For instance the maximally correlated state on two qubits

$$\rho_{XY} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|).$$

Non-classical correlations, so-called *entanglement*, creates fundamental difference between classical and quantum information theory.

Definition 2.21. A pure state $|\psi_{AB}\rangle \in \mathcal{H}_{AB}$ is called *entangled* if it is not a product state, i.e., there are no $|\psi_A\rangle$ and $|\psi_B\rangle$ such that $|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$.

The following lemma follows directly from Theorem 2.20.

Lemma 2.22. *A pure state $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ is entangled if and only if the reduced density matrix ρ_A (or ρ_B) is not pure.*

Definition 2.23. A state $\rho_{AB} \in S(AB)$ is *maximally entangled state* if ρ_{AB} is pure and both reduced states are maximally mixed, i.e., $\rho_A = \frac{1}{|A|} I_A$ and $\rho_B = \frac{1}{|B|} I_B$.

The pervious lemma implies that maximally entangled states are indeed entangled since their reduced states are maximally mixed. We are not going to make this precise at this moment, but the "more mixed" the reduced states are, the "more entangled" the state is. We will come back to this later.

By Schmidt decomposition a pure state is maximally entangled if and only if its Schmidt rank is d and its Schmidt coefficients are all equal to $1/\sqrt{d}$, where $|A| = |B| = d$. In particualr maximally entangled states exists if and only if $|A| = |B|$, and they are of the form

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^{d} |e_j\rangle \otimes |f_j\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

For example using the standard basis for both, we get

$$|\Phi^+_{AB}\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |ii\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d.$$

Lemma 2.24 (properties of maximally entangled states)**.** *Let $|\Phi^+_{AB}\rangle$ be the maximally entangled state and let ρ_{AB} be its density matrix.*

(a) The reduced density matrices are maximally mixed: $\rho_A = \frac{1}{d} I_A$.

(b) For any $d \times d$ matrix M, we have $(M_A \otimes I_B)|\Phi^+_{AB}\rangle = (I_A \otimes M_B^T)|\Phi^+_{AB}\rangle$.

(c) For two $d \times d$ matrices M and N, we have $\langle\Phi^+_{AB}|M_A \otimes N_B|\Phi^+_{AB}\rangle = \frac{1}{d}\operatorname{tr}[M^T N] = \frac{1}{d}\operatorname{tr}[MN^T]$.

When $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^d$ and we use the standard basis, then the notation is clear. In general, the operators M_A, M_B^T, etc., are defined with respect to the same basis as those obtained in the Schmidt decomposition.

Exercise 2.25. Show that $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|{+}{+}\rangle + |{-}{-}\rangle)$.

Solution. Use Theorem 2.24(b). We have $(U \otimes U)|\Phi^+_{AB}\rangle = |\Phi^+_{AB}\rangle$. Now just apply the unitary

$$U = H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

□

If we have classical state on systems X and Y, it is of the form

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

It is a convex combination of the state $|x\rangle\langle x| \otimes |y\rangle\langle y|$. The next definition capters wider class of states where correlations between A and B are of classical nature and defines entanglement as its complement.

Definition 2.26. A state $\rho_{AB} \in S(AB)$ is *separable* if there is a collection of states $\rho_{A,x} \in S(A)$, $\rho_{B,x} \in S(B)$, for $x \in \Omega$, and a probability distribution p on Ω, for soem set Ω, such that

$$\rho_{AB} = \sum_{x\in\Omega} p(x)\rho_{A,x} \otimes \rho_{B,x}.$$

A state is called *entangled* if it is not separable.

Clearly classical states are separable. If a state is entangled, there is no choice of basis for A and B such that the state is classical in that basis. We can interpret separable states as follows: (1) Alice and Bob generate some shared classical random variable with outcome $x \in \Omega$. (2) Based on the outcome, Alice prepares $\rho_{A,x}$ and Bob prepares $\rho_{B,x}$. Thus, spearable states form a class of states where the correlations between Alice and Bob are of classical nature.

The definition of entanglement for pure states is consistent with the latter definition. Indeed, a pure product state is separable, and conversely we know that if a pure state is a convex combination of product states, then it must be product state itself. Also, if $|\psi_{AB}\rangle\langle\psi_{AB}| = \rho_A \otimes \rho_B$, then ρ_A and ρ_B must be pure since $1 = \operatorname{rank}(\rho_{AB}) = \operatorname{rank}(\rho_A)\operatorname{rank}(\rho_B)$.

2.6 TODO

- revise the proof Schmidt decomposition
- prove Theorem 2.16
- the no-communication theorem
- add problem about tensor product of observables

Chapter 3

Non-local Games and Quantum Foudnations

Non-local games provide a mathematical framework to study the power of entanglement.

3.1 Basic definitions

Definition 3.1. A *non-local game* is a 6-tuple $\mathcal{G} = (I_A, I_B, O_A, O_B, \pi, V)$, where I_A and I_B are finite sets of *questions*, O_A and O_B are finite sets of *answers*, π is a probability distribution on $I_A \times I_B$, and $V \colon O_A \times O_B \times I_A \times I_B \to \{0, 1\}$ is a *winning predicate*.

The game is played as follows: the referee selects a pair of questions $(x, y) \in I_A \times I_B$ according to π and sends x to Alice and y to Bob. Alice responds with $a \in O_A$ and Bob with $b \in O_B$. The players win if $V(a, b, x, y) = 1$. The players can agree on a strategy before the game starts, but cannot communicate afterwards.

Definition 3.2. th Given a non-local game $\mathcal{G} = (I_A, I_B, O_A, O_A, \pi, V)$, a *strategy* is an element $p = (p(a, b|x, y))_{a,b,x,y} \in [0, 1]^{O_A \times O_B \times I_A \times I_B}$ such that for each $(x, y) \in I_A \times I_B$, we have

$$\sum_{(a,b) \in O_A \times O_B} p(a, b|x, y) = 1.$$

Given a strategy S for the game $\mathcal{G}$, the *succes probability* is defined as

$$\omega(\mathcal{G}, S) = \sum_{x,y} \pi(x, y) \sum_{a,b} V(a, b, x, y) p(a, b|x, y).$$

We say that a strategy S is *perfect* if $\omega(\mathcal{G}, S) = 1$. For a fixed collection of possible strategies $\mathcal{S}$, then we define the associated *value* of $\mathcal{G}$ by

$$\omega(\mathcal{G}, \mathcal{S}) = \sup_{S \in \mathcal{S}} \omega(\mathcal{G}, S).$$

Definition 3.3. A *deterministic strategy* for a non-local game $\mathcal{G} = (I_A, I_B, O_A, O_A, \pi, V)$ is given by a pair of functions $f \colon I_A \to O_A$ and $g \colon I_B \to O_B$.

Definition 3.4. A strategy p for a non-local game $\mathcal{G}$ is called *classical/randomized* if there exists a probability space (Ω, μ) and for all a, x and b, y measurable functions $p_A(a|x,\cdot), p_B(b|y,\cdot)\colon \Omega \to [0,1]$ such that for all x, y, ω,

$$\sum_a p_A(a|x,\omega) = \sum_b p_B(b|y,\omega) = 1,$$

and

$$p(a,b|x,y) = \int_\omega p_A(a|x,\omega) p_B(b|y,\omega) d\mu(w).$$

The *classical value* of a non-local game $\mathcal{G}$ is denoted by $\omega_c(\mathcal{G})$.

Proposition 3.5. *If $\mathcal{G}$ is a non-local game, then $\omega_c(\mathcal{G})$ is attained by a deterministic strategy.*

Definition 3.6. A strategy p for a game $\mathcal{G}$ is *non-signalling* if every for every a, b, x, x', y, y' it holds

$$\sum_a p(a,b|x,y) = \sum_a p(a,b|x',y), \quad \text{and} \quad \sum_b p(a,b|x,y) = \sum_b p(a,b|x,y').$$

Definition 3.7. In a *(finite-dimensional) quantum strategy* Alice and Bob share a quantum state $\rho_{AB} \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$, and their answers are the result of a measurement on their system. For each $x \in I_A$, there is a measurement $\mu_A^x\colon O_A \to \mathrm{PSD}(\mathcal{H}_A)$ on Alice's system and Alice answers the outcome of the measurement. Similalry, for each $y \in I_B$, Bob has measurement $\mu_B^y\colon O_B \to \mathrm{PSD}(\mathcal{H}_B)$ and answers the outcome of his measurement. This means that Alice and Bob answer a and b, when asked x and y, with probability

$$p(a,b|x,y) = \mathrm{tr}\left[(\mu_A^x(a) \otimes \mu_B^y(b))\rho_{AB}\right].$$

The *quantum value* of a game $\mathcal{G}$ is denoted by $\omega_q(\mathcal{G})$.

3.2 CHSH game

The *Clauser-Horne-Shimony-Holt (CHSH) game* has $I_A = I_B = O_A = O_B = \{0,1\}$. The winning predicate is given by

$$x \cdot y = (a+b) \mod 2 \iff x \wedge y = a \oplus b.$$

The probability distribution π is just the uniform distribution on $\{0,1\}^2$.

Classical strategy. We first consider classical strategies. Suppose that there exist a deterministic strategy, given by functions $f, g\colon \{0,1\} \to \{0,1\}$ such that $a = f(x)$ and $b = f(y)$, that wins for all pairs of questions. The winning condition implies that

$$\sum_{x,y\in\{0,1\}} f(x) + g(y) \mod 2 = \sum_{x,y\in\{0,1\}} x \cdot y = 1.$$

On the other hand,

$$\sum_{x,y\in\{0,1\}} f(x) + g(y) = 2 \sum_{x\in\{0,1\}} f(x) + 2 \sum_{y\in\{0,1\}} g(y),$$

which is even. Thus, $\omega_c(\mathcal{G}) \leq 3/4$. There is a deterministic strategy that achives this value, e.g., always answer $a = b = 0$. Therefore $\omega_c(\mathcal{G}) = 3/4$.

Quantum strategy. We will use observables with outcomes ± 1 corresponding to answers a and b as $(-1)^a$ and $(-1)^b$. We encode Alice's projective measurement with operators $\mu_A^x(a) = P_a^x$ corresponding to the question x by the observable $A^x = P_0^x - P_1^x$. Similarly, we encode Bob's projective measurement with operators $\mu_B^x(b) = Q_b^y$ corresponding to the question y by the observable $B^y = Q_0^y - Q_1^y$.

A quantum strategy is determined by a state $\rho_{AB} = |\psi\rangle\langle\psi|$, Alice's obserables $\{A^0, A^1\}$, and Bob's observables $\{B^0, B^1\}$. The probabilty that Alice and Bob give the same answer to x and y minus the probability that hey give different answers is given by

$$\mathbb{P}(a \neq b) - \mathbb{P}(a = b) = \langle\psi|A^x \otimes B^y|\psi\rangle.$$

Further,

$$\beta := 2\mathbb{P}(\text{win}) - 1 = \mathbb{P}(\text{win}) - \mathbb{P}(\text{lose}) = \frac{1}{4}\langle\psi|A^0 \otimes B^0 + A^0 \otimes B^1 + A^1 \otimes B^0 - A^1 \otimes B^1|\psi\rangle.$$

The quantity β is called the *bias* of the strategy.

Consider the states

$$|\psi_0(\theta)\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle, \qquad |\psi_0(\theta)\rangle = -\sin\theta|0\rangle + \cos\theta|1\rangle.$$

Clearly these form an orthonormal basis. The corresponding vectors on the Bloch sphere are $\vec{r}_\theta = (\sin 2\theta, 0, \cos 2\theta)$ and $-\vec{r}_\theta$, respectively. The corresponding observable is

$$O_\theta(\vec{r}) = \rho(\vec{r}) - \rho(-\vec{r}) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}.$$

With the notation in place, we are ready to choose a strategy. As a quantum state, we pick the maximally entangled state $|\psi\rangle := |\Phi_{AB}^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The Alice's measurement is given by the observables

$$A^0 = O_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = Z, \qquad A^1 = O_{\frac{\pi}{4}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = X.$$

Bob's measurement is the basis measurement corresponding to $\vec{r} = \frac{1}{\sqrt{2}}(1, 0, 1)$, for $y = 0$, and $\vec{s} = \frac{1}{\sqrt{2}}(-1, 0, 1)$, for $y = 1$. This corresponds to the measurement in the bases

$$\begin{aligned} \Big\{\cos(\frac{\pi}{8})|0\rangle + \sin(\frac{\pi}{8})|1\rangle, -\sin(\frac{\pi}{8})|0\rangle + \cos(\frac{\pi}{8})|1\rangle\Big\}, &\quad \text{for } y = 0, \\ \Big\{\cos(-\frac{\pi}{8})|0\rangle + \sin(-\frac{\pi}{8})|1\rangle, -\sin(\frac{\pi}{8})|0\rangle + \cos(-\frac{\pi}{8})|1\rangle\Big\}, &\quad \text{for } y = 1. \end{aligned}$$

In terms of observables, we have

$$B^0 = O_{\frac{\pi}{8}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{Z + X}{\sqrt{2}}, \qquad B^1 = O_{-\frac{\pi}{8}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix} = \frac{Z - X}{\sqrt{2}}.$$

We have $\vec{r} = (x, y, z)$ and $\vec{s} = (x', y', z')$ in the Bloch sphere. Recall that

$$O(\vec{r}) = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix}.$$

If $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, by Theorem 2.24(c), we get

$$\langle\psi|O(\vec{r}) \otimes O(\vec{s})|\psi\rangle = \frac{1}{2}\operatorname{tr}[O(\vec{r})^T O(\vec{s})].$$

By directly computing this, trace, we get $\langle\psi|O(\vec{r}) \otimes O(\vec{s})|\psi\rangle = xx' - yy' + zz$. We use this to calculate β. Note that

$$\begin{aligned}\beta &= \frac{1}{4}\langle\psi|O_0 \otimes O_{\pi/8} + O_0 \otimes O_{-\pi/8} + O_{\pi/4} \otimes O_{\pi/8} - O_{\pi/4} \otimes O_{-\pi/8}|\psi\rangle \\ &= \frac{1}{4}\left((0 - 0 + \frac{1}{\sqrt{2}}) + (0 - 0 + \frac{1}{\sqrt{2}}) + (\frac{1}{\sqrt{2}} - 0 + 0) - (-\frac{1}{\sqrt{2}} + 0 - 0)\right) \\ &= \frac{1}{\sqrt{2}}.\end{aligned}$$

From this, we Immedietelly get $\mathbb{P}(\text{win}) = \frac{1}{2} + \frac{1}{2\sqrt{2}} = \cos^2(\pi/8) \approx 0.85$. We have proven the following theorem.

Theorem 3.8. *If $\mathcal{G}$ is the CHSH game, then*

$$\omega_q(\mathcal{G}) \geq \frac{1}{2} + \frac{1}{2\sqrt{2}} > \omega_c(\mathcal{G}).$$

The following *Tsirelson bound* shows that in fact we cannot do better.

Theorem 3.9 (Tsirelson bound)**.** *Let $\mathcal{G}$ be the CHSH game. Then*

$$\omega_q(\mathcal{G}) = \frac{1}{2} + \frac{1}{2\sqrt{2}}.$$

Proof. It remains to prove the upper bound. The key is that the bias β, defined above, is derived for an arbitrary quantum strategy. Let $|\psi_{AB}\rangle$ be the shared quantum state, which we assume to be pure and the measurements to be projective (we will later see that this is possible).

When we construct the observables A^x and B^y from a projective two-outcome measurement, we see that A^x and B^y are Hermitian with eigenvalues ± 1. Consequently $(A^x)^2 = I_A$ and $(B^y)^2 = I_B$. Let

$$M_{AB} = A^0 \otimes B^0 + A^0 \otimes B^1 + A^1 \otimes B^0 - A^1 \otimes B^1$$

We have

$$4\beta = \langle\psi_{AB}|M_{AB}|\psi_{AB}\rangle.$$

Using CS-inequality, we get

$$\begin{aligned}|\langle\psi_{AB}|M_{AB}|\psi_{AB}\rangle| &\leq \sqrt{\langle\psi_{AB}|M_{AB}M_{AB}^\dagger|\psi_{AB}\rangle}\sqrt{\langle\psi_{AB}|\psi_{AB}\rangle} \\ &= \sqrt{\langle\psi_{AB}|M_{AB}^2|\psi_{AB}\rangle}\end{aligned}$$

since M_{AB} is Hermitian and $|\psi_{AB}\rangle$ is normalized. We will now bound the right-hand side. To that end, we rewrite

$$\begin{aligned}
M_{AB}^2 &= \left(A^0 \otimes (B^0 + B^1) + A^1 \otimes (B^0 - B^1)\right)^2 \\
&= \left(A^0 \otimes (B^0 + B^1) + A^1 \otimes (B^0 - B^1)\right)\left(A^0 \otimes (B^0 + B^1) + A^1 \otimes (B^0 - B^1)\right) \\
&= (A^0)^2 \otimes (B^0 + B^1)^2 + (A^1)^2 \otimes (B^0 - B^1)^2 \\
&\quad + A^0A^1 \otimes (B^0 + B^1)(B^0 - B^1) + A^1A^0 \otimes (B^0 - B^1)(B^0 + B^1) \\
&= I_A^2 \otimes 4I_B - [A^0, A^1] \otimes [B^0, B^1].
\end{aligned}$$

Note that, in general, for hermitian matrices M and N with $M^2 = N^2 = I$, we have by triangle inequality and submultiplicativity

$$\|[M, N]\| = \|MN - NM\| \leq \|MN\| + \|NM\| \leq 2\|M\|\|N\| = 2.$$

Now, using the CS-inequality, we get

$$\begin{aligned}
|\langle\psi_{AB}|[A^0, A^1] \otimes [B^0, B^1]|\psi_{AB}\rangle| &\leq \big\|[A^0, A^1] \otimes [B^0, B^1]|\psi_{AB}\rangle\big\|\||\psi_{AB}\rangle\| \\
&\leq \big\|[A^0, A^1] \otimes [B^0, B^1]\big\|\||\psi_{AB}\rangle\|\||\psi_{AB}\rangle\| \\
&= \big\|[A^0, A^1]\big\|\big\|[B^0, B^1]\big\| \leq 4.
\end{aligned}$$

So, $\langle\psi_{AB}|M_{AB}^2|\psi_{AB}\rangle \leq 8$ and we get

$$4\beta = \langle\psi_{AB}|M_{AB}|\psi_{AB}\rangle \leq \sqrt{8} = 2\sqrt{2} \quad \Longrightarrow \quad \omega_q(\mathcal{G}) \leq \frac{1}{2} + \frac{1}{2\sqrt{2}}.$$

□

3.3 Mermin-Peris magic square game

Consider the magic square given in the Fig. 3.1 on the left. Each variable can be assigned values in $\{\pm 1\}$. We have six equations: each row and column corresponds to an equation $x_ix_jx_k = b$. So we have six Boolean linear equations.

The corresponding non-local game is defined as follows. The sets of questions are $X = Y = \{1, 2, 3\}$. For a pair of quesitons $(x, y) \in X \times Y$, Alice and Bob must respond with $a = (a_1, a_2, a_3), b = (b_1, b_2, b_3) \in \{-1, 1\}^3$, respectively. Alice and Bob win the game if the following three conditions are satisfied:

$$a_1 \cdot a_2 \cdot a_3 = r_x, \quad b_1 \cdot b_2 \cdot b_3 = c_y, \quad a_x = b_y.$$

In other words, their answers could form a part of the solution to the system of linear equations.

Classical strategy. First note that if there is a solution, then Alice and Bob can just agree beforehand, defining a deterministic strategy. On the other hand if $\omega(\mathcal{G}) = 1$, then there must be a determinstc strategy attaining this value. After Alice and Bob receive questions (x, y), then they must agree on the value of the (x, y) element, by the winning condition $a_x = b_y$. Doing this for every (x, y), we see that the strategy fixes a specific filling of the magic square

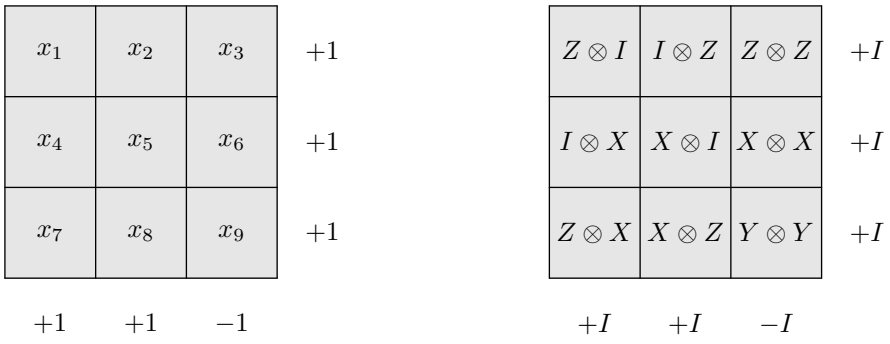


Figure 3.1: The Mermin-Peris magic square on the left. A 4-dimensional operator solution on the right.

which Alice and Bob are both using. Moreover, the other winning conditions require this filling to be a solution.

However, the system clearly has no solution. Indeed, suppose we have a solution such that the row products are all 1, the first two column products are 1 and the third column prodcut is -1. We get contradiction by taking all row products and obtaining 1, and taking all column products and obtaining -1.

Quantum strategy. To obtain a quantum strategy, we will make use of the *operator solution* given in Fig. 3.1 on the right. Each element of the operator solution is a tensor product of Pauli matrices on $\mathbb{C}^2 \otimes \mathbb{C}^2$. It is easy to see that each operator squares to the identity, operators in each row and each column pairwise commute, the product of each row and of the first two columns is the identity, and the product of the third column is the minus identity. In particular, the operators in each column and each row can be simultaneously diagonalized.

Alice and Bob share the following entangled state in $(\mathcal{H}_{A_1} \otimes \mathcal{H}_{B_1}) \otimes (\mathcal{H}_{A_2} \otimes \mathcal{H}_{B_2})$:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_{A_1} \otimes |0\rangle_{B_1} + |1\rangle_{A_1} \otimes |1\rangle_{B_1}) \otimes \frac{1}{\sqrt{2}}(|0\rangle_{A_2} \otimes |0\rangle_{B_2} + |1\rangle_{A_2} \otimes |1\rangle_{B_2}). \tag{3.3.1}$$

We can use the Schmidt decomposition to check that $|\psi\rangle$ is in fact an entangled state. By multiplying and reordering the factors of the tensor product, we get

$$\frac{1}{2}(|00\rangle_A|00\rangle_B + |01\rangle_A|01\rangle_B + |10\rangle_A|10\rangle_B + |11\rangle_A|11\rangle_B), \tag{3.3.2}$$

which is already a Schmidt decomposition with all $s_i = 1/2$ and the standard bases. By Theorem 2.19, we have

$$\rho_A = \frac{1}{4}(|00\rangle_A + |01\rangle_A + |10\rangle_A + |11\rangle_A),$$

which is not pure. So by Theorem 2.22 it is entangled.

Upon receiving questions from the referee, Alice and Bob each measure their two quibits with the operator from the corresponding row or column in the magic square to determine their outputs. For example, if Alice receives $x = 3$, she measures $Z \otimes X$, $X \otimes Z$, $Y \otimes Y$, and answers (a_1, a_2, a_3) according to the outcomes. From the properties of the operators we discussed above it directly follows that their answers will satisfy the corresponding equations.

It remains to argue that their answers will agree on the common entry. We can check this individually, for example, consider the entry $Z \otimes X$. Using the equations

$$Z|0\rangle = +1|0\rangle, \quad Z|1\rangle = -|1\rangle, \quad X|0\rangle = |1\rangle, \quad X|1\rangle = |0\rangle,$$

and the Eq. (3.3.2), we can conclude that

$$(Z \otimes X)_A \otimes (Z \otimes X)_B |\psi\rangle = |\psi\rangle.$$

Another way to see this is to rearrange $(Z \otimes X) \otimes (Z \otimes X)$ to $(Z_{A_1} \otimes Z_{B_1}) \otimes (X_{A_1} \otimes X_{A_2})$ ans use the fact that for maximally entangled state $|\Phi^+_{AB}\rangle$, we have

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

Then we can deduce this from Eq. (3.3.1).

3.4 TODO

- Bell's theorem formulate
- Kochen-Specker thoerem, contenxtuality

Chapter 4

Quantum Channels

So far we have considered a static picture. In this chapter, we will discuss the possible dynamics in a quantum system, i.e., the class of operations that can be applied to a quantum system.

4.1 Classical channels

Given X and Y with outcome sets Ω_X and Ω_Y, suppose that Y is the result of applying some operation to X. If we start with a fixed $x \in \Omega_X$, we get $y \in \Omega_Y$ with probability $q(y|x)$. If we start with a distribution p_X on X, we get outcome y with probability

$$p_Y(y) = \sum_x q(y|x)p_X(x)$$

.

Definition 4.1. A *classical channel* from X to Y is a map

$$Q\colon \mathcal{P}(X) \to \mathcal{P}(Y)$$

where $p_Y = Q(p_X)$, for $p_X \in \mathcal{P}(X)$ is given by

$$p_Y(y) = \sum_x q(y|x)p_X(x)$$

for some collection of $q(y|x) \in \mathbb{R}_{\geq 0}$ such that

$$\sum_y q(y|x) = 1,$$

for all x.

Note that the numbers $q(y|x)$ define a *stochastic matrix*, i.e., a matrix with non-negative real entries such that columns sum up to 1. A stochastic matrix is *deterministic* if each column contains exactly one entry equal to 1 and others equal to 0. Deterministic matrices are in one-to-one correspondance with functions $\Omega_X \to \Omega_Y$. One way to view the following lemma is that each stochastic matrix is a convex combination of deterministic matrices.

Lemma 4.2. *Let $Q\colon \mathcal{P}(X) \to \mathcal{P}(Y)$ be a classical channel, defined by conditional probabilities $q(y|x)$. Then there are deterministic channels Q_i and a probability distribution such that $Q(p_X) = \sum_i p_i Q_i(p_X)$.*

Proof. Let p_f be the probability that the channel Q chooses function $f\colon \Omega_X \to \Omega_Y$, i.e.,

$$p_f = \prod_x q(f(x)|x).$$

Let D_f be the deterministic channel (matrix) corresponding to the function f. We claim that

$$Q = \sum_f p_f D_f.$$

First of all, we have

$$\sum_f p_f = \sum_f \prod_x q(f(x)|x) = \prod_x \sum_y q(y|x) = \prod_x 1 = 1,$$

so, p_f is a probability distribution. Next, we have

$$\begin{aligned}
\sum_{f:f(x)=y} p_f &= \sum_{f:f(x)=y} \prod_x q(f(x)|x) \\
&= q(y|x) \sum_{f:f(x)=y} \prod_{x'\neq x} q(f(x')|x') \\
&= q(y|x) \prod_{x'\neq x} \sum_y q(y|x') \\
&= q(y|x).
\end{aligned}$$

Thus, Q is a convex combination of the D_f's. □

4.2 Unitary operations

The quantum channel representing a unitary operation U on A is

$$\Phi[\rho] = U\rho U^\dagger,$$

that is, the conjugation by U. This description is consistent with the fact the density matrix that representas given quantum state vector $|\psi\rangle$ is $|\psi\rangle\langle\psi|$. In particular if the unitary operation U is performed on $|\psi\rangle$, then the output state is represented by $U|\psi\rangle$, and so the density matrix describing this state is $(U|\psi\rangle)(U|\psi\rangle)^\dagger = U|\psi\rangle\langle\psi|U^\dagger$. Once we know what actian the quantum channel has taken on pure states, we can conclude it must work as specified above on any density matrix ρ. To see that ϕ is completely positive, we can write $\rho = M^\dagger M$ and then

$$(\mathbb{1}_R \otimes U)\rho(\mathbb{1}_R \otimes U^\dagger) = (\mathbb{1}_R \otimes U)M^\dagger M(\mathbb{1}_R \otimes U^\dagger) = (M(\mathbb{1}_R \otimes U^\dagger)^\dagger(M(\mathbb{1}_R \otimes U^\dagger)),$$

which shows complete positivity. Moreover,

$$\mathrm{tr}[\phi[\rho]] = \mathrm{tr}[U\rho U^\dagger] = \mathrm{tr}[U^\dagger U\rho] = 1,$$

so it is also trace-preserving.

4.3 Quantum channels

Quantum channels will be defined as linear operators acting on a quantum state and producing some new quantum state. Since quantum states are linear opertors themselves, it will be usefull to make a distinciton.

Definition 4.3. If A and B are quantum systems with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, a *superoperator* from A to B is a linear map $\Phi_{A\to B}\colon \mathrm{Lin}(A) \to \mathrm{Lin}(B)$. If $A = B$, we write $\Phi_{A\to A}$.

Example 4.4. We list some essential superoperators.

- The *identity superoperator* $\mathcal{I}_A$, defined by $\mathcal{I}[M_A] = M_A$, for every $M_A \in \mathrm{Lin}(A)$.
- The *partial trace* tr_B. If we have two quantum systems A and B, then taking the partial trace over B defines a superoperator $\mathrm{tr}_B\colon \mathrm{Lin}(AB) \to \mathrm{Lin}(A)$.
- If $V \in \mathrm{Isom}(A, B)$, then $\mathcal{V}_{A\to B}\colon \mathrm{Lin}(A) \to \mathrm{Lin}(B)$ is a superoperator defined by $M_A \mapsto VM_AV^\dagger$.

Note that if $\Phi_{A\to B}$ and $\Psi_{B\to C}$ are superoperators, then the composition $\Psi_{B\to C} \circ \Phi_{A\to B}$ is also a superoperator from A to C. Also, if $\Phi_{A\to B}$ and $\Psi_{C\to D}$ are superoperators, then the tensor product superoperator $\Phi_{A\to B} \otimes \Psi_{C\to D}$ from AC to BD is defined by extendning the following formula by linearity:

$$\Phi_{A\to B} \otimes \Psi_{C\to D}[M_A \otimes M_C] = \Phi_{A\to B}[M_A] \otimes \Psi_{C\to D}[M_C],$$

for all $M_A \in \mathrm{Lin}(A)$ and $M_C \in \mathrm{Lin}(C)$.

Example 4.5. Note that the trace is a superoperator $\mathrm{tr}\colon \mathrm{Lin}(B) \to \mathrm{Lin}(\mathbb{C})$. For quantum systems A and B the partial trace $\mathrm{tr}_B\colon \mathrm{Lin}(AB) \to \mathrm{Lin}(A)$ can be expressed as $\mathrm{tr}_B = \mathcal{I}_A \otimes \mathrm{tr}$.

Definition 4.6. A superoperator $\Phi_{A\to B}$ is a *positivity-preserving* (or just *positive*) map if it maps every positive operator $P_A \in \mathrm{PSD}(A)$ to a positive operator $\Phi_{A\to B}[P_A] \in \mathrm{PSD}(B)$.

Positivity here *does not mean* that $\Phi \in \mathrm{PSD}(\mathrm{Lin}(\mathcal{H}))$.
All the maps in Example 4.4 are positive.

Lemma 4.7. *If $\Phi_{A\to B}$ and $\Psi_{B\to C}$ are positive maps, then the composition $\Psi_{B\to C} \circ \Phi_{A\to B}$ is a positive map.*

Example 4.8 (tensor product of positive maps is not positive)**.** Suppose that $\Phi_{A\to A'}$ and $\Psi_{B\to B'}$ represent some quantum dynamics. Then we would expect their tensor product to represent the dynamics on the joint system AB. However, in general the tensor prodcut of positive maps does not have to be a positive map, so positivity is not enough.

To see this, let A be qubit system. We define the superoperator $\mathcal{T}_A$ by $\mathcal{T}_A[M_A] = M_A^T$. We have that $\mathcal{T}_A$ is positive and in particular we also have $\mathcal{T}(\rho_A) \in S(A)$, for all $\rho_A \in S(A)$. Let B be another qubit system and let $\rho_{AB} = |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|$ be the maximally entangled state. We

have

$$\begin{aligned}(\mathcal{T}_A \otimes \mathcal{I}_B)[\rho_{AB}] &= \frac{1}{2}(\mathcal{T}_A \otimes \mathcal{I}_B)[|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|] \\ &= \frac{1}{2}(\mathcal{T}_A \otimes \mathcal{I}_B)[|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|] \\ &= \frac{1}{2}(\mathcal{T}_A \otimes \mathcal{I}_B)[|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 0| \otimes |0\rangle\langle 1| + |0\rangle\langle 1| \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|] \\ &= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.\end{aligned}$$

This matrix is not positive.

Definition 4.9. A superoperator $\Phi_{A\to B}$ is *completely positive (CP)* map if for any reference sytem R, the superoperator $\Phi_{A\to B} \otimes \mathcal{I}_R$ is positive. The set of completely positive maps from A to B will be denoted by $\mathrm{CP}(A, B)$ and we let $\mathrm{CP}(A) = \mathrm{CP}(A, A)$.

Lemma 4.10. *If $\Phi_{A\to B} \in \mathrm{CP}(A, B)$ and $\Psi_{B\to C} \in \mathrm{CP}(B, C)$, then $\Psi_{B\to C} \circ \Phi_{A\to B} \in \mathrm{CP}(A, C)$. If $\Phi_{A\to B} \in \mathrm{CP}(A, B)$ and $\Psi_{C\to D} \in \mathrm{CP}(C, D)$, then $\Phi_{A\to B} \otimes \Psi_{C\to D} \in \mathrm{CP}(AC, BD)$.*

Proof. Since composition of positive maps is positive, the first part follows from

$$(\Psi_{B\to C} \circ \Phi_{A\to B}) \otimes \mathcal{I}_R = (\Psi_{B\to C} \otimes \mathcal{I}_R) \circ (\Phi_{A\to B} \otimes \mathcal{I}_R).$$

For the second part, it follows from the definition that $\Phi_{A\to B} \otimes \mathcal{I}_C$ and $\mathcal{I}_B \otimes \Psi_{C\to D}$ are completely positive. Then from the first part it follows that

$$\Phi_{A\to B} \otimes \Psi_{C\to D} = (\mathcal{I}_B \otimes \Psi_{C\to D}) \circ (\Phi_{A\to B} \otimes \mathcal{I}_C)$$

is completely positive. □

Definition 4.11. A superoperator $\Phi_{A\to B}$ is *trace preserving (TP)* if

$$\mathrm{tr}[\Phi_{A\to B}[M_A]] = \mathrm{tr}[M_A]$$

for all $M_A \in \mathrm{Lin}(A)$. The set of trace preserving maps from A to B will be denoted by $\mathrm{TP}(A, B)$ and we let $\mathrm{TP}(A) = \mathrm{TP}(A, A)$.

Definition 4.12. *Quantum channel* is a superoperator which is both completely positive and trace preserving (CPTP). The set of quantum channels, also called *CPTP maps*, from A to B will be denoted by

$$\mathrm{C}(A, B) = \{\Phi_{A\to B} \in \mathrm{CP}(A, B) \cap \mathrm{TP}(A, B)\}$$

and we let $\mathrm{C}(A) = \mathrm{C}(A, A)$.

The maps in Example 4.4 are all quantum channels.

Axiom 5. The set of possible operations from an input quantum system A to an output quantum system B is given by the set of quantum channels $\mathrm{C}(A, B)$.

Example 4.13. Here are some examples of quantum channels.

- *Depolarizing channel.* Let $\mathcal{D}_p\colon \mathrm{Lin}(A) \to \mathrm{Lin}(A)$ be defined by
$$\mathcal{D}_p[M_A] = (1-p)M_A + p\,\mathrm{tr}[M_A]\tau_A,$$
where $\tau_A = \frac{1}{\dim \mathcal{H}_A}\mathbb{1}_A$. It models the situation when the state is unchanged with probability $(1-p)$ and with probability p the state gets lost and replaced by the maximally mixed state. It therefore models the complete loss of information—both classical and quantum—due to random noise.
- *Dephasing channel.* Let $\mathcal{P}_p\colon \mathrm{Lin}(A) \to \mathrm{Lin}(A)$ be defined by
$$M_A \mapsto (1-p)M_A + p\sum_{a\in\Sigma_A} \langle a|M_A|a\rangle |a\rangle\langle a|.$$
This represents gradual loss of coherence. Mathematically, it surpresses the off-diagonal elements of the density matrix while leaving the diagonal unchanged. It can also be viewed as performing a projective measurement with probability p and forgetting the outcome.
- *Erasure channel.* Let $\mathcal{E}_p\colon \mathrm{Lin}(A) \to \mathrm{Lin}(A')$, where $\mathcal{H}_{A'} = \mathcal{H}_A \oplus \mathrm{span}\{|\perp\rangle\}$, be defined by
$$M_A \mapsto (1-p)M_A + p\,\mathrm{tr}[M_A]\,|\perp\rangle\langle\perp|\,.$$
This channel modes a process in which the state is transmitted without error with probability $1-p$ an replaced by a fixed known orthogonal state $|\perp\rangle$ with probability p. The receiver can detect whether erasure occured by checking whether hte output lies in the original space $\mathcal{H}_A$.
- *Replacement channel.* Let $\mathcal{R}_\rho\colon \mathrm{Lin}(A) \to \mathrm{Lin}(B)$, for $\rho_B \in S(B)$, be defined by
$$M_A \mapsto \mathrm{tr}[M_A]\rho_B.$$
Discard and replace by a fixed state.
- The channlel representing applying the unitary $U_i \in \mathrm{Lin}(A)$ with probability p_i:
$$M_A \mapsto \sum_i p_i U_i M_A U_i^\dagger.$$

4.4 Characterization of quantum channels

We will prove a characterization of CP maps, which will give us several ways to test whether a map is CP. This will lead to characterization of CPTP maps, i.e., quantum channels.

For a positive superoperator $\Phi_{A\to B}$, we need to test the positivity of $\Phi_{A\to B}\otimes\mathcal{I}_R$ on the states that are entangled between A and R. Indeed, if we apply $\Phi_{A\to B}\otimes\mathcal{I}_R$ to a separable state, we get a convex combination of positive operators, which is positive. The following technical tool will be usefull.

Definition 4.14 (Choi operator)**.** Give a superoperator $\Phi_{A\to B}$ and a basis for $\{|a\rangle\}$ for $\mathcal{H}_A$ (typically the standard basis), the *Choi operator* $J^\Phi_{AB} \in \mathrm{Lin}(AB)$ is defined as
$$J^\Phi_{AB} = \sum_{a,a'} |a\rangle\langle a'| \otimes \Phi_{A\to B}[|a\rangle\langle a'|].$$

The *normalized Choi operator* is
$$\omega^\Phi_{AB} = \frac{1}{\dim \mathcal{H}_A} J^\Phi_{AB}.$$

We can interpret the Choi operator is the result of applying $\mathcal{I}_A \otimes \Phi_{A' \to B}$ to the unnormalized maximally entangled state between A and A', where A' is just a copy of A. Also note that

$$\omega_{AB}^{\Phi} = (\mathcal{I}_A \otimes \Phi_{A' \to B})[|\Phi_{AA'}^{+}\rangle\langle\Phi_{AA'}^{+}|].$$

is a quantum state, called *Choi state*, if $\Phi_{A \to B}$ is a quantum channel. In particular, the Choi operator must be positive if $\Phi_{A \to B}$ is a quantum channel.

The Choi operator J_{AB}^{Φ} completely determines the superoperator $\Phi_{A \to B}$. By linearity it suffices to know how it acts on $\{|a\rangle\langle a'|\}$, which forms a basis of $\mathrm{Lin}(A)$, and this information can be obtained from J_{AB}^{Φ}.

Lemma 4.15 (Choi isomorphism)**.** *Let J_{AB}^{Φ} be the Choi operator of a superoperator $\Phi_{A \to B}$. Then for any $M_A \in \mathrm{Lin}(A)$, we have*

$$\Phi_{A \to B}[M_A] = \mathrm{tr}_A[(M_A^T \otimes \mathbb{1}_B) J_{AB}^{\Phi}],$$

where the transpose is computed in the same basis as chose in the definition of Choi operator.

Proof. The Choi operator for Φ is

$$J^{\Phi} = \sum_{a,a'} \phi[|a\rangle\langle a'|] \otimes |a\rangle\langle a| \in \mathrm{Lin}(B \otimes A').$$

We have

$$\begin{aligned}
\mathrm{tr}_{A'}[(\mathbb{1}_B \otimes M_A^T) J^{\Phi}] &= \sum_{a,a'} \mathrm{tr}_{A'}[(\mathbb{1}_B \otimes M_A^T)\phi[|a\rangle\langle a'|] \otimes |a\rangle\langle a'|] \\
&= \sum_{a,a'} \mathrm{tr}_{A'}[\phi[|a\rangle\langle a'|] \otimes M_A^T |a\rangle\langle a'|] \\
&= \sum_{a,a'} \phi[|a\rangle\langle a'|] \,\mathrm{tr}[M_A^T |a\rangle\langle a'|] \\
&= \sum_{a,a'} \langle a'|M_A^T|a\rangle \phi[|a\rangle\langle a'|] \\
&= \sum_{a,a'} \langle a|M_A|a'\rangle \phi[|a\rangle\langle a'|] \\
&= \phi\Big[\sum_{a,a'} \langle a|M_A|a'\rangle |a\rangle\langle a'|\Big] \\
&= \phi[M_A].
\end{aligned}$$

□

Theorem 4.16 (Characterization of CP maps)**.** *For any superoperator $\Phi_{A \to B}$, the following are equivalent:*

(a) $\Phi_{A \to B}$ is completely positive.

(b) The Choi operator is positive, i.e., $J_{AB}^{\Phi} \in \mathrm{PSD}(AB)$

(c) (Kraus reprsentation) There exist operators $K_1, \dots, K_r \in \mathrm{Lin}(A, B)$ *such that*

$$\Phi_{A\to B}[M_A] = \sum_{i=1}^{r} K_i M_A K_i^\dagger.$$

(d) (Stinespring representation) There exsits a quantum system E *and an operator* $V \in \mathrm{Lin}(A, BE)$ *such that*

$$\Phi_{A\to B}[M_A] = \mathrm{tr}_E[V M_A V^\dagger].$$

Proof. (a) $\implies$ (b), we observed before.

We now prove (b) $\implies$ (c). Suppose that J^Φ_{AB} is postiive. There is a decomposition

$$J^\Phi_{AB} = \sum_{i=1}^{r} |v_i\rangle\langle v_i|,$$

for some pairwise orthogonal eigenvectors $|v_i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. We dont take the eigenvectors normalized, for simplicity, we hide the eigenvalues inside. We use the basis $\{|a\rangle\}$. We can write

$$|v_i\rangle = \sum_{a,b} v_{i,ab}|ab\rangle.$$

By Lemma 4.15, we have

$$\begin{aligned}
\Phi_{A\to B}[M_A] &= \sum_i \mathrm{tr}_A[(M_A^T \otimes \mathbb{1}_B)|v_i\rangle\langle v_i|] \\
&= \sum_i \sum_{a,b} \sum_{a'b'} v_{i,ab}\overline{v_{i,a'b'}}\, \mathrm{tr}_A[(M_A^T \otimes \mathbb{1}_B)|a\rangle\langle a'| \otimes |b\rangle\langle b'|] \\
&= \sum_i \sum_{a,b} \sum_{a'b'} v_{i,ab}\overline{v_{i,a'b'}}\, \mathrm{tr}_A[M_A^T|a\rangle\langle a'| \otimes |b\rangle\langle b'|] \\
&= \sum_i \sum_{a,b} \sum_{a'b'} v_{i,ab}\overline{v_{i,a'b'}}\, \mathrm{tr}[M_A^T|a\rangle\langle a'|]|b\rangle\langle b'| \\
&= \sum_i \sum_{a,b} \sum_{a'b'} v_{i,ab}\overline{v_{i,a'b'}}\langle a'|M_A^T|a\rangle|b\rangle\langle b'| \\
&= \sum_i \sum_{a,b} \sum_{a'b'} v_{i,ab}\overline{v_{i,a'b'}}\langle a|M_A|a'\rangle|b\rangle\langle b'| \\
&= \sum_i \sum_{a,b} \sum_{a'b'} v_{i,ab}|b\rangle\langle a|M_A|a'\rangle\langle b'|\overline{v_{i,a'b'}} \\
&= \sum_i K_i M_A K_i^\dagger.
\end{aligned}$$

For (c) $\implies$ (d), we let $E = \mathbb{C}^r$ with basis $\{|i\rangle\}$ and we let

$$V = \sum_{i=1}^{r} K_i \otimes |i\rangle.$$

Then

$$\mathrm{tr}_E[V M_A V^\dagger] = \sum_{i=1}^{r} K_i M_A K_i^\dagger.$$

For (d) $\implies$ (a), we just observe that the partial trace and $M_A \mapsto V M_A V^\dagger$ are completely positive and so is their composition. □

We see that even thought the definition of complete positivity allows for an arbitrary reference system R, which could have arbitrarily large dimension, it suffices to consider the situation when R is a copy of A. If $\Phi_{A\to B} \otimes \mathcal{I}_A$ or equivalently $\mathcal{I}_A \otimes \Phi_{A\to B}$ is positive, then the Choi operator is positive and by Theorem 4.16 this implies that $\Phi_{A\to B}$ is completely positive.

Corollary 4.17. *A superoperator $\Phi_{A\to B}$ is completely positive if and only if $\Phi_{A\to B} \otimes \mathcal{I}_A$ is positive.*

Example 4.18. We compute the Choi operator, Kraus representation, and Stinespring representation for the completely dephasing channel $\mathcal{P} = \mathcal{P}_1$ on a quantum system A. This is defined by

$$M_A \mapsto \sum_a \langle a|M_A|a\rangle |a\rangle\langle a|,$$

for a basis $\{|a\rangle\}$, typically the standard basis of $\mathcal{H}_A = \mathbb{C}^{\dim H_A}$. In other words, the channel sets all off-diagonal terms in the density matrix to zero. The Choi operator is

$$\begin{aligned} J^{\mathcal{P}}_{AA} &= \sum_{a,a'} \mathcal{P}(|a\rangle\langle a'|) \otimes |a\rangle\langle a'| \\ &= \sum_a |a\rangle\langle a| \otimes |a\rangle\langle a|. \end{aligned}$$

This is the unnormalized maximally mixed state.

For the Kraus decompositon, we can just write

$$\mathcal{P}[M_A] \sum_a |a\rangle\langle a|M_A|a\rangle\langle a|,$$

which is a Kraus decomposition.

For the Stinespring representation, we can follow the construction from Theorem 4.16 and we may take E to be a copy of A and

$$V = \sum_a |aa\rangle\langle a| \in \mathrm{Lin}(A, BE).$$

Theorem 4.19 (Characterization of quantum channels)**.** *For any superoperator $\Phi_{A\to B} \in \mathrm{CP}(A,B)$, the following statements are equivalent:*

(a) $\Phi_{A\to B} \in \mathrm{TP}(A,B)$ and hence $\Phi_{A\to B} \in \mathrm{C}(A,B)$.

(b) The Choi operator is such that $\mathrm{tr}_B[J^\Phi_{AB}] = \mathbb{1}_A$

(c) For a Kraus representation $\Phi_{A\to B}[M_A] = \sum_{i=1}^r K_i M_A K_i^\dagger$, it holds that

$$\sum_{i=1}^r K_i^\dagger K_i = \mathbb{1}_A.$$

(d) For a Stinespring representation $\Phi_{A\to B}[M_A] = \mathrm{tr}_E[V M_A V^\dagger]$, it holds that

$$V^\dagger V = \mathbb{1}_A.$$

That is, $V \in \mathrm{Isom}(A, BE)$.

In (c) and (d), if the statements holds for one particular representation, it holds for all.

Proof. From linear algebra we know that $\mathrm{tr}[AX] = \mathrm{tr}[BX]$ for all X and only if $A = B$. As special case of this is that $A = \mathbb{1}$ if and only if for all X, we have $\mathrm{tr}[AX] = \mathrm{tr}[X]$. We will use this several times in this proof.

(a) $\iff$ (b): By Lemma 4.15, we have

$$\begin{aligned}\mathrm{tr}[\Phi_{A\to B}[M_A]] &= \mathrm{tr}[\mathrm{tr}_A[(M_A^T \otimes \mathbb{1}_B)J_{AB}^\Phi]] \\ &= \mathrm{tr}[(M_A^T \otimes \mathbb{1}_B)J_{AB}^\Phi] \\ &= \mathrm{tr}[\mathrm{tr}_B[(M_A^T \otimes \mathbb{1}_B)J_{AB}^\Phi]] \\ &= \mathrm{tr}[(M_A^T \,\mathrm{tr}_B[J_{AB}^\Phi]]\end{aligned}$$

Thus, $\Phi_{A\to B}$ is trace preserving if and only if for all M_A

$$\mathrm{tr}[M_A^T \,\mathrm{tr}_B[J_{AB}^\Phi]] = \mathrm{tr}[M_A] = \mathrm{tr}[M_A^T].$$

(a) $\iff$ (c): If we have a Kraus reresentation $\Phi_{A\to B}[M_A] = \sum_{i=1}^r K_i M_A K_i^\dagger$, then

$$\begin{aligned}\mathrm{tr}[\Phi_{A\to B}[M_A]] &= \sum_{i=1}^r \mathrm{tr}[K_i M_A K_i^\dagger] \\ &= \sum_{i=1}^r \mathrm{tr}[K_i^\dagger K_i M_A] \\ &= \mathrm{tr}[(\sum_{i=1}^r K_i^\dagger K_i) M_A]\end{aligned}$$

Thus $\Phi_{A\to B}$ is trace preserving if and only if for all M_A

$$\mathrm{tr}[M_A] = \mathrm{tr}[(\sum_{i=1}^r K_i^\dagger K_i) M_A].$$

This is equivalent to $\sum_{i=1}^r K_i^\dagger K_i = \mathbb{1}_A$.

(a) $\iff$ (d): If we have Stinespring representation $\Phi_{A\to B}[M_A] = \mathrm{tr}_E[V M_A V^\dagger]$, then it holds that

$$\mathrm{tr}[\Phi_{A\to B}[M_A]] = \mathrm{tr}[\mathrm{tr}_E[V M_A V^\dagger]] = \mathrm{tr}[V^\dagger V M_A].$$

Thus, $\Phi_{A\to B}$ is trace preserving if and only if for all M_A, we have

$$\mathrm{tr}[V^\dagger V M_A] = \mathrm{tr}[M_A].$$

This is equivalent to $V^\dagger V = \mathbb{1}_A$. □

4.5 Physical realization of quantum channels

We will deduce the physical realizability of quantum channels from the Stinespring representation. Suppose that $\Phi_{A\to B}$ is a quantum channel with Stinespring representation given by

$$\Phi_{A\to B}[M_A] = \mathrm{tr}_E[VM_AV^\dagger],$$

for some isometry $V \in \mathrm{Isom}(A, BE)$. We can choose F and E' with $\mathcal{H}_E \subseteq \mathcal{H}_{E'}$ such that $\dim(\mathcal{H}_A \otimes \mathcal{H}_F) = \dim(\mathcal{H}_B \otimes \mathcal{H}_{E'})$ and extend V to a unitary $U \in \mathrm{Lin}(AF, BE')$ such that

$$U|\psi_A\rangle|\phi_F\rangle = V|\psi_A\rangle \in \mathcal{H}_B \otimes \mathcal{H}_E \subseteq \mathcal{H}_B \otimes \mathcal{H}_{E'},$$

for all $|\psi_A\rangle \in \mathcal{H}_A$ and any fixed $|\phi_F\rangle \in \mathcal{H}_F$. Then we have

$$\Phi_{A\to B}[M_A] = \mathrm{tr}_{E'}[U(M_A \otimes |\phi_F\rangle\langle\phi_F|)U^\dagger].$$

So, any quantum channel can be realized by (1) preparing a fixed pure state $|\phi_F\rangle\langle\phi_F|$ in and additional system F, (2) applying a unitary map to AF, discarding the subsystem E'. In physics terms: couple the system to an environment, time-evolve by a global Hamiltionian, and then restrict to a relevant subsystem.

4.6 Measurements as quantum channels

Given measurement $\mu = \mu_A$ on a quantum system A with outcomes in classical system X, we can model this measurement as the channel

$$\mathcal{M}^\mu_{A\to X}[M_A] = \sum_x \mathrm{tr}[\mu(x)M_A]|x\rangle\langle x|.$$

This it is a quantum channel. It maps a state ρ_A to the probability distribution of the measurement outcomes. It is a *quantum-to-classical channel*, meaning that it maps any quantum state to a classical state. Conversely, it can be shown that any quantum-to-classical channel corresponds to a measurement.

Implementing measuremets. We will show that any measurement can be constructed using only a projective measurements.

Theorem 4.20 (Naimark)**.** *Suppose that $\mu = \mu_A$, then there exists system F and a projective measurement $\nu = \nu_{AF}$ such that*

$$\mathcal{M}^\mu_{A\to X}(\rho_A) = \mathcal{M}^\nu_{AF\to X}(\rho_A \otimes |0_F\rangle\langle 0_F|),$$

for some $|0_F\rangle\langle 0_F| \in S(F)$.

Proof. Consider the unitary extension that we had above:

$$\mathcal{M}^\mu_{A\to X}(\rho_A) = \mathrm{tr}_E[U(\rho_A \otimes |0_F\rangle\langle 0_F|)U^\dagger],$$

for a unitary U and some pure state $|0_F\rangle$ on F. The probability of the outcome x is

$$\begin{aligned} p_x(\rho_A) &= \langle x|\mathcal{M}^{\mu}_{A\to X}(\rho_A)|x\rangle \\ &= \langle x|\operatorname{tr}_E[U(\rho_A \otimes |0_F\rangle\langle 0_F|)U^\dagger]|x\rangle \\ &= \operatorname{tr}[(|x\rangle\langle x| \otimes \mathbb{1}_E)U(\rho_A \otimes |0_F\rangle\langle 0_F|)U^\dagger] \\ &= \operatorname{tr}[U^\dagger(|x\rangle\langle x| \otimes \mathbb{1}_E)U(\rho_A \otimes |0_F\rangle\langle 0_F|)] \end{aligned}$$

We define $P_x = U^\dagger(|x\rangle\langle x| \otimes \mathbb{1}_E)U \in \operatorname{Lin}(AF)$, which is a projection. We have

$$\sum_x P_X = \sum_x U^\dagger(|x\rangle\langle x| \otimes \mathbb{1}_E)U = U^\dagger(\sum_x |x\rangle\langle x|)U = U^\dagger U = \mathbb{1}_{AF}.$$

So, P_x defines a projective measurement. □

Non-destructive measurement. So far we considered destructive measurements.

4.7 TODO

- Non-destructive measurements and quantum instrument.
- A few words about continuous time-evolution

Chapter 5

Basic Quantum Information Protocols

After introducing the formalism of quantum theory, we will use this to model *information processing protocols*. We start with two of the most basic protocols, *superdense coding* and *teleportation*. Next, we will see an important principle for determining whether any quantum information is lost under a quantum channel, or whether we can perfectly reconstruct the input.

5.1 Superdense coding and teleportation

We will partially answer two questions: (a) If we can send over a qubit, can we also send classical information, and if so, how much? (b) If we can send over classical bits, can we also send over quantum information, and if so, how much?

For the analysis in this section, we fix some notation. In particular, the *Bell basis* of $\mathcal{H}_A \otimes \mathcal{H}_B = \mathbb{C}^2 \otimes \mathbb{C}^2$ consists of the vectors

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

The basis is set up in such a way that $\Phi_{AB}^{00} = |\Phi_{AB}^{+}\rangle$ and

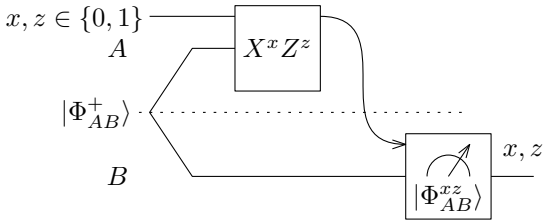
$$\Phi_{AB}^{xz} = (X^x Z^z \otimes \mathbb{1}_B)|\Phi_{AB}^{+}\rangle.$$

Note that

$$\Phi_{AB}^{xz} = (\mathbb{1}_A \otimes (X^x Z^z)^T)|\Phi_{AB}^{+}\rangle = (\mathbb{1}_A \otimes Z^z X^x)|\Phi_{AB}^{+}\rangle.$$

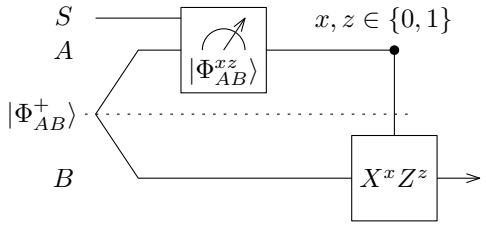
Superdense coding. *Superdense coding* is a quantum channel, which consumes an entangled pair and uses one qubit of communication to send two classical bits.

If Alice wants to send over bits x and z, she applies X^xZ^z to her system. Bob will have Φ_{AB}^{xz} and he can just measure in the Bell basis to find x and z.



Teleportation. *Teleportation* is a quantum channel, which consumes an entangled pair and uses two classical bits of communication to send a qubit.

Alice wants to send quantum information, but she can only send classical information to Bob. They share the maximally entangled state. Alice wants to send the system S to Bob. She perfroms the measurement in the Bell basis on the joint system SA. Then Alice sends the outcome x, z of the measurement to Bob. Finally, Bob applies the operator X^xZ^z to his parf the shared maximally entangled state.



First prepare the maximally entangled state on AB:

$$M_S \mapsto M_S \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|.$$

Then, we measure in the Bell basis on SA:

$$M_S \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+| \mapsto \sum_{x,z\in\{0,1\}} (\langle\Phi_{SA}^{xz}| \otimes \mathbb{1}_B(M_S \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|)|\Phi_{SA}^{xz}\rangle \otimes \mathbb{1}_B) \otimes |zx\rangle\langle zx|.$$

Finally, we send over the classical system with the information about x and z to Bob, and Bob applies Z^zX^x, which gives

$$M_S \mapsto \sum_{x,z\in\{0,1\}} Z^zX^x\big(\langle\Phi_{SA}^{xz}| \otimes \mathbb{1}_B\big)(M_S \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|)\big(|\Phi_{SA}^{xz}\rangle \otimes \mathbb{1}_B\big)X^xZ^z,$$

where we use $(Z^zX^x)^\dagger = X^xZ^z$.

We will denote the last equation by $\Phi_{S\to B}$. Both S and B are quibits, and we would like to prove that $\Phi_{S\to B}$ is in fact the idenitty. To that end, we use $|\Phi_{SA}^{xz}\rangle = (X^xZ^z \otimes \mathbb{1}_A)|\Phi_{SA}^+\rangle$ to write

$$\Phi_{S\to B}(M_S) = \sum_{x,z\in\{0,1\}} Z^zX^x(\langle\Phi_{SA}^{xz}| \otimes \mathbb{1}_B(X^xZ^zM_SZ^zX^x \otimes |\Phi_{AB}^+\rangle\langle\Phi_{AB}^+|)|\Phi_{SA}^{xz}\rangle \otimes \mathbb{1}_B)X^xZ^z.$$

It can be seen that

$$(\mathbb{1}_S \otimes \langle \Phi^+_{AB}|)(|\Phi^+_{SA}\rangle \otimes \mathbb{1}_B) = \frac{1}{2} \sum_{i\in\{0,1\}} |i_S\rangle\langle i_B|.$$

From there, we have

$$\Phi_{S\to B}(M_S) = \frac{1}{4}\sum_{x,z}(Z^z X^x)(X^x Z^z) M_S (Z^z X^x)(X^x Z^z) = M_S.$$

5.2 Decoupling, recovery and error correction

A quantum channel $\Phi_{A\to B}$ is *recoverable* if there is $\mathcal{R}_{B\to A}$ such that $\mathcal{R}_{B\to A} \circ \Phi_{A\to B} = \mathcal{I}_A$.

Definition 5.1. If $\Phi_{A\to B}$ is a quantum channel with Stinespring isometry V, then the *complementary channel* is defined as

$$\Phi^c_{A\to E}[M_A] = \mathrm{tr}_B[V M_A V^\dagger].$$

Theorem 5.2. *Let ρ_{ABE} be the Choi state of the Stinespring extension $\mathcal{V}_{A\to BE}$ of $\Phi_{A\to B}$. The following are equivalent:*

(a) The channel $\Phi_{A\to B}$ is recoverable.

(b) The Choi state ρ_{ABE} satisfies $\rho_{AE} = \rho_A \otimes \rho_E$.

(c) The complementary channel $\Phi^c_{A\to E}$ is constant, that is, $\Phi^c_{A\to E}[M_A] = \mathrm{tr}[M_A]\sigma_E$, for some $\sigma_E \in S(E)$.

For example in the context of quantum computation, these concepts are also used to characterize *error correcting codes*. The idea is that we have a *logical* system A we would like to keep track of without being affected by noise. However, we only have access to physical system B, which is subject to a noise channel Φ_B. An error correcting code is an isometry $V\colon \mathcal{H}_A \to \mathcal{H}_B$, which should be such that we can recover A after the noise channel has been applied to B. In this case the recovery is called *error correction* or *decoding*. The image of V is the code subspace, $VV^\dagger$ is the projection onto the code subspace.

Theorem 5.3. *Suppose Φ_B has Kraus operators $X_0, \ldots, X_{r-1}$ and let $P = VV^\dagger$ be the projection onto the code subspace of $\mathcal{H}_B$. Then the errors due to the noise channel Φ_B can be corrected if and only if*

$$PX_j^\dagger X_i P = s_{ij} P$$

for some complex number s_{ij}.

Chapter 6

Distance Measures

In many situations, when transmitting information, if we allow no probability of error whatsoever, we have no capacity of transmitting information. If we want to send a one bit of information, one thing we can do is to send the same message over n times. If a bit flip occurs with some probability, then the receiver can guess the correct message by taking the most occuring symbol in the string. This leads to an error of more than $n/2$ bit flips occur. By taking sufficiently large n, the probability of error can be made arbitrarily small, but never exactly zero. We introduce various measures of errors for quantum states and channels and derive some of their properties.

6.1 Norms of operators

Since quantum states are operators, we need the notion of distance on the space of operators.

Definition 6.1. If $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, then the *trace norm*, also known as 1-*norm*, is the sum of the singular values of M. We denote it by $\|M\|_1$.

Lemma 6.2. *Let $M \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$.*

(a) $\|M\|_1 = \mathrm{tr}[\sqrt{M^\dagger M}]$.

(b) $\|M\|_1 = \|M^\dagger\|_1 = \|M^T\|_1 = \|\overline{M}\|_1$.

(c) If V and W are isometries, then $\|VMW\|_1 = \|M\|_1$.

(d) If $\mathcal{H} = \mathcal{K}$ and $M = M^\dagger$ has spectrum $\lambda_1, \ldots, \lambda_n$, then $\|M\|_1 = \sum_{i=1}^n |\lambda_i|$.

Note that if follows that $\|\rho\|_1 = 1$, for a quantum state ρ, which makes the norm good for measuring distances between quantum states.

The *Hilbert-Schmidt norm* $\|\cdot\|_{\mathrm{HS}}$, or 2-*norm*, comes from the *Hilbert-Schmidt inner product*, defined by

$$\langle M, N\rangle_{\mathrm{HS}} := \mathrm{tr}[M^\dagger N].$$

So, $\|M\|_2 =: \sqrt{\langle M, M\rangle_{\mathrm{HS}}} = \sqrt{\mathrm{tr}[M^\dagger M]}$. This inner product is equivalent to viewing matrices as vectors and taking the usual inner product. By applying the CS-inequality for this inner product, we get

$$|\,\mathrm{tr}[M^\dagger N]|^2 = |\langle M, N\rangle_{\mathrm{HS}}|^2 \leq \|M\|_2^2 = \mathrm{tr}[M^\dagger M]\,\mathrm{tr}[N^\dagger N].$$

Since $\left\|M^\dagger\right\|_2 = \|M\|_2$, we get

$$|\operatorname{tr}[MN]| = \|M\|_2 \|N\|_2$$

The *operator norm*, or ∞*-norm*, is the largest singular value of the operator, and can be equivalently expressed as

$$\|M\|_\infty = \sup_{\||v\rangle\|=1} \|M|v\rangle\| = \sup_{\||v\rangle\|=1} \sqrt{\langle v|M^\dagger M|v\rangle} = \sup_{\||v\rangle\|=\||w\rangle\|=1} |\langle w|M|v\rangle|.$$

An important property is submultiplicativity:

$$\|MN\|_\infty \leq \|M\|_\infty \|N\|_\infty.$$

Lemma 6.3. *For $M \in \operatorname{Lin}(\mathcal{H})$, we have*

$$\|M\|_1 = \max_U |\operatorname{tr}[MU]|.$$

Moreover, for all $N \in \operatorname{Lin}(\mathcal{H})$

$$|\operatorname{tr}[MN]| \leq \|MN\|_1 \leq \|M\|_1 \|N\|_\infty.$$

6.2 Trace distance

Definition 6.4 (trace distance)**.** Let $\rho, \sigma \in S(\mathcal{H})$. Then the *trace distance* between ρ and σ is

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

Lemma 6.5. *For $\rho, \sigma \in S(\mathcal{H})$, we have*

(a) If $\rho - \sigma$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, then $T(\rho, \sigma) = \sum_{\lambda_i > 0} \lambda_i$.

(b) We have

$$T(\rho, \sigma) = \max_{0 \leq Q \leq \mathbb{1}} \operatorname{tr}[Q(\rho - \sigma)].$$

The maximum is attained by an operator Q which is a projection.

Lemma 6.6. *For $\rho, \sigma \in S(\mathcal{H})$, we have*

(a) $0 \leq T(\rho, \sigma) \leq 1$, and $T(\rho, \sigma) = 0$ if and only if $\rho = \sigma$.

(b) The trace distance is invariant under isometries: if $V \in \operatorname{Isom}(\mathcal{H}, \mathcal{K})$, then

$$T(V\rho V^\dagger, V\sigma V^\dagger) = T(\rho, \sigma).$$

(c) The trace distance is monotonic under quantum channels: if $\Phi_{A\to B} \in C(A, B)$ and $\rho_A, \sigma_A \in S(A)$, then

$$T(\Phi_{A\to B}[\rho_A], \Phi_{A\to B}[\sigma_A]) \leq T(\rho_A, \sigma_A).$$

Theorem 6.7 (Helstrom)**.** *Let $\rho, \sigma \in S(\mathcal{H})$. Suppose that with probability $1/2$ we received the state ρ and with probability $1/2$ we received the state σ. Then the optimal probability of identifying the correct state by a two-outcome measurement is given by*

$$p_{opt} = \frac{1}{2} + \frac{1}{2} T(\rho, \sigma).$$

If p_X and q_X are probability distributions on some classical system X, with deinsity matrices

$$\rho_X = \sum_x p_X(x)|x\rangle\langle x|, \quad \sigma_X = \sum_x q_X(x)|x\rangle\langle x|,$$

then we see that

$$T(p_X, q_X) = \frac{1}{2}\sum_x |p_X(x) - q_X(x)|,$$

which corresponds to the usual *statistical distance* between probability distributions.

If $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$ are pure states, we may also compute their distance to be

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

To see this, we can write $\psi = \alpha|\phi\rangle + \beta|\phi^\perp\rangle$ In the basis $\{|\phi\rangle, |\phi^\perp\rangle\}$, we have

$$\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\sigma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix} = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}.$$

In priciple, the states lie in higher dimension, but all the other matrix elements are zero, so we ignore them. Now, we have

$$\rho - \sigma = \begin{pmatrix} |\beta|^2 & -\alpha\beta^* \\ -\alpha^*\beta & -|\beta|^2 \end{pmatrix}.$$

So, $\mathrm{tr}[\rho - \sigma] = 0$ and

$$\det(\rho - \sigma) = -|\beta|^4 - |\alpha|^2|\beta|^2 = -|\beta|^2(|\alpha|^2 + |\beta|^2) = -|\beta|^2,$$

So the eigenvalues must be $\pm|\beta|$. It follows that $T(\rho, \sigma) = |\beta|$. On the other hand, by definition

$$|\beta| = \sqrt{1 - |\alpha|^2} = \sqrt{1 - |\langle\phi|\psi\rangle|^2}$$

So, we get that

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

6.3 Fidelity and purified distance

From the previous equation, we see that the trace distance between pure states may be computed using their *overlap* $|\langle\phi|\psi\rangle|$. The states are close if their overlap is close to 1. This quantity is also known as *fidelity* and can be naturally extended to mixed states. Observe that if $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$, then

$$|\langle\phi|\psi\rangle| = \sqrt{\langle\psi|\phi\rangle\langle\phi|\psi\rangle} = \sqrt{\mathrm{tr}[\rho\sigma]}.$$

A possible measure would be $\mathrm{tr}[\rho\sigma]$, however, in general $\mathrm{tr}[\rho^2] \neq 1$ and this is not the best choice.

Definition 6.8. Let $\rho, \sigma \in S(\mathcal{H})$. Then the *fidelity* between ρ and σ is defined by

$$F(\rho, \sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1.$$

Note that

$$\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 = \mathrm{tr}\left[\sqrt{(\sqrt{\rho}\sqrt{\sigma})^\dagger\sqrt{\rho}\sqrt{\sigma}}\right] = \mathrm{tr}\left[\sqrt{\sqrt{\sigma}\sqrt{\rho}\sqrt{\rho}\sqrt{\sigma}}\right] = \mathrm{tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right].$$

In general, it is not the case that $\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}} = \sigma^{\frac{1}{4}}\sqrt{\rho}\sigma^{\frac{1}{4}}$. Fidelity is symmetric:

$$\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 = \left\|(\sqrt{\rho}\sqrt{\sigma})^\dagger\right\|_1 = \left\|\sqrt{\sigma}\sqrt{\rho}\right\|_1.$$

However, fidelity is not a metric. This is clear since states are close when they have fidelity close to 1. Nevertheless, it is possible to define a metric based on fidelity as we will see.

If $\rho = |\phi\rangle\langle\phi|$ is pure, then $\sqrt{\rho} = \rho$, and we get

$$\sqrt{\rho}\sigma\sqrt{\rho} = |\phi\rangle\langle\phi|\sigma|\phi\rangle\langle\phi|.$$

This is an operator with rank one, which implies that the square root can be taken outside the trace:

$$\mathrm{tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right] = \sqrt{\mathrm{tr}[|\phi\rangle\langle\phi|\sigma|\phi\rangle\langle\phi|]}.$$

We obtain

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

In particular, if $\sigma = |\psi\rangle\langle\psi|$ is also pure, then $F(\rho,\sigma) = |\langle\phi|\psi\rangle|$.

Theorem 6.9 (Uhlmann)**.** *Let $\rho_A, \sigma_A \in S(A)$. Let R be a reference system such that both ρ_A and ρ_B have purifications on AR. Then*

$$F(\rho_A,\sigma_A) = \max_{|\phi_{AR}\rangle,|\psi_{AR}\rangle} |\langle\phi_{AR}|\psi_{AR}\rangle|,$$

where the maximum is over purifications $|\phi_{AR}\rangle$ and $|\psi_{AR}\rangle$ of ρ_A and σ_A, respectively. Equivalently, if $|\phi_{AR}\rangle$ and $|\psi_{AR}\rangle$ are some fixed purifications of ρ_A and σ_A, then

$$F(\rho_A,\sigma_A) = \max_{U_R} |\langle\phi_{AR}|\mathbb{1}_A \otimes U_R|\psi_{AR}\rangle|,$$

where the maximum is take over all unitaries on R.

Lemma 6.10. *Suppose $\rho, \sigma \in S(\mathcal{H})$.*

(a) $0 \leq F(\rho,\sigma) \leq 1$ and $F(\rho,\sigma) + 1$ if and only if $\rho = 0$.

(b) The fidelity is invariant under isometries: if $V \in \mathrm{Isom}(\mathcal{H},\mathcal{K})$, then

$$F(V\rho V^\dagger, V\sigma V^\dagger) = F(\rho,\sigma).$$

(c) The fidelity is monotonic under quantum channels: if $\Phi_{A\to B} \in C(A,B)$ and $\rho_A, \sigma_A \in S(A)$, then

$$F(\Phi_{A\to B}[\rho_A], \Phi_{A\to B}[\sigma_A]) \geq F(\rho_A,\sigma_A).$$

The inequality in the monotonicity is in the other direction than for the trace distance. The states may get closer to each other as we apply a quantum channel, so their fidelity *increases*.

Let us compute fidelity for classical states. If p_X and q_X are probability distributions on some classical system X with density matrices

$$\rho_X = \sum_x p_X(x)|x\rangle\langle x|, \quad \sigma_X = \sum_x q_X(x)|x\rangle\langle x|,$$

then we can directly see that

$$F(\rho_X, \sigma_X) = \sum_x \sqrt{p_X(x)q_X(x)}.$$

This is not so much used as a similarity measure in probability theory. The main motivation for fidelity stems from its relation to purifications, so it is more natural in the quantum setting.

The fidelity and trace distance can be related to each other by the following.

Theorem 6.11 (Fuchs-van de Graaf inequalities)**.** *For any $\rho, \sigma \in S(\mathcal{H})$ it holds that*

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

These inequalities show that the trace distance and fidelity are very similar distance measures. If ρ and σ are close to each other in the one measure, they are also close in the other measure. Also note that the bounds are independent of the dimension of the Hilbert space. The reason for using two similar measures is more pragmatic. The trace distance has a good operational interpretation (Helstrom's theorem) related to distinguishing states. The fidelity is convenient for pure states as Uhlmann's theorem is a powerful tool.

Definition 6.12 (purified distance)**.** Suppose that $\rho, \sigma \in S(\mathcal{H})$. The the *purified distance* between ρ and σ is

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

For pure states, we have $P(\rho, \sigma) = T(\rho, \sigma)$. From the properties of fidelity, we have $0 \leq P(\rho, \sigma) \leq 1$, invariance under isometries, and monotonicity under quantum channels. Moreover, from Uhlmann's theorem we get

$$P(\rho_A, \sigma_A) = \min_{\rho_{AR}, \sigma_{AR}} P(\rho_{AR}, \sigma_{AR}),$$

where the minimum is over all states ρ_{AR} and σ_{AR} such that $\mathrm{tr}_R[\rho_{AR}] = \rho_A$ and $\mathrm{tr}_R[\sigma_{AR}] = \sigma_A$. We can restrict the minimization to purifications.

Lemma 6.13. *The purified distance defines a metric on $S(\mathcal{H})$.*

6.4 Error measures for quantum channels

The next natural step is to get a good measure for when quantum channels are close. Intuitively, this means if we input the same state in them, they should give nearby states as output. We need to specify which measure we use on states and which states are allowed to serve as input.

For our purposes, it will be enough to measure how close is a channel to the *identity channel*. The reason is that we will be interested in some process and we would like the quantum system

that comes out is a good approximation to the input system. This could bea communication scenario when we want the output of the communication protocil to be a reliable transmission of the input.

Suppose that a quantum system is in a state ρ_A, we apply channel Φ_A and we would like to know if this channel acted similarly to the identity channel $\mathcal{I}_A$. A naive approach could be that we are close to the action of the identity channel if

$$F(\Phi_A[\rho_A], \rho_A) \geq 1 - \varepsilon,$$

for some small $\varepsilon > 0$.

Definition 6.14. Let $\Phi_A \in C(A)$ and $\rho_A \in S(A)$. Then the *entanglement fidelity* of Φ_A with respect to ρ_A is

$$F_E(\Phi_A, \rho_A) = \inf_{\rho_{AR}} F((\Phi \otimes \mathcal{I}_R)(\rho_{AR}), \rho_{AR})$$

where the infimum is over all systems R and state ρ_{AR} which extend ρ_A. The *entanglement purified distance* is

$$P_E(\Phi_A, \rho_A) = \sqrt{1 - F_E(\Phi_A, \rho_A)^2}.$$

The next lemma says that the infimum is attained by an *arbitrary* purification of ρ_A.

Lemma 6.15. *Suppose that $\rho_{AR} = |\phi_{AR}\rangle\langle\phi_{AR}|$ is a purification of ρ_A. Then*

$$F_E(\Phi_A, \rho_A) = F((\Phi \otimes \mathcal{I}_R)(\rho_{AR}), \rho_{AR}) = \sqrt{\langle\phi_{AR}|((\Phi_A \otimes \mathcal{I}_R)(\rho_{AR}))|\phi_{AR}\rangle}.$$

Proof (Sketch). We can always minimize over pure states. We can take purification, fidelity has monotonicity property, the fidelity can only decrees after applying a quantum channel. Additionally all purifications are related by isometries on the references system. If we look at the expression, the channel Φ_A does not act on the reference system. So if we act with isometry on both sides, it does not change the fidelity. □

Lemma 6.16. *Suppose that $\Phi_A[M_A] = \sum_i X_i M_A X_i^\dagger$ is a Kraus representation of a quantum channel. Then*

$$F_E(\Phi_A, \rho_A) = \sqrt{\sum_i |\operatorname{tr}[X_i \rho_A]|^2}.$$

Proof (Sketch). Take the standard purification $\sum_a (\sqrt{\rho_A} \otimes \mathbb{1}_R)|aa\rangle$ and plug it into the previous lemma with Φ_A represented by Kraus representation. □

Chapter 7

Quantum Compression and Entropy

One way to quantify information is by seeing how much a source can be compressed. Suppose that we have source producing symbols x according to a probability distribution p_X. We would like to encode x into r bits in a way that we can recover the original symbol. The minimal r at which we can do this is a measure for the amount of information in the source p_X.

7.1 Classical compression

Let X be a random variable with probability distribution p_X and alphabet Σ_X. The idea of compression is that there is a classical *encoding* channel E to a smaller classical system C, representing *compressed* information, and a classical *decoding* channel D, back from C to X. We say that E and D form a zero-error *r-compression* code for p_X if the alphabet Σ_C of C is such that $\log|\Sigma_C| \leq r$. One option is to demand that we can always recover correctly, that is,

$$D(E(x)) = x \quad \text{if} \quad p_X(x) \neq 0.$$

However, as the following example shows, allowing a small probability of error makes a huge difference in how well once can compress.

Example 7.1. Consider a random variable X taking values $x = 1, \ldots, N+1$, where

$$p_X(x) = 2^{-x}, \quad x = 1, \ldots, N, \quad \text{and} \quad p_X(N+1) = 2^{-N}.$$

If we do not allow any error, we need $\lceil \log N \rceil$ bits. However, if we allow a small probability of error, we may compress as follows: for $x \leq K$ we express x using $\lceil \log K \rceil$ bits and decode accordingly. If $x > K$, we assign an error message (or any random bitstring). Then we see that the probability of error equals

$$\mathbb{P}(D(E(x)) \neq x) = \sum_{x=K+1}^{N+1} p_X(x) = 2^{-K}.$$

If we allow $\varepsilon = 2^{-K}$ error we need $\log K = \log\log(\varepsilon^{-1})$ bits. Thus, we can compress X to

$$r = \left\lceil \log\log \frac{1}{\varepsilon} \right\rceil$$

bits with probability of error at most ε. Note that N can be arbitrarily large, if we allow a small probability of error 2^{-30}, we only need $\lceil \log 30 \rceil = 5$ compressed bits. □

This is the idea behind the *lossy compression*, where a fixed probability of error is allowed in the decoding process. The following definitions captures this formally.

Definition 7.2. We define a *classical* (ε, r)*-compression code* for a distribution $p_X \in \mathcal{P}(X)$ to r bits with error $\varepsilon > 0$ as a system C on an alphabet Σ_C of size $\log|\Sigma_C| \leq r$ and a pair of classical channels

$$E\colon \mathcal{P}(X) \to \mathcal{P}(C), \quad \mathcal{D}\colon \mathcal{P}(C) \to \mathcal{P}(X)$$

which are such that

$$\mathbb{P}(D(E(\mathcal{X})) = \mathcal{X}) = \sum_x p_X(x)\mathbb{P}(D(E(x)) = x) \geq 1 - \varepsilon.$$

Let

$$C^\varepsilon(p_X) = \min\{r : \text{there is a } (\varepsilon, r)\text{-compression code for } p_X\}$$

Definition 7.3. If $p_X \in P(X)$, then the *support* of p_X is given by

$$\operatorname{supp}(p_X) = \{x \in \Sigma_X : p_X(x) > 0\}.$$

Let

$$H_0(p_X) = \log(|\operatorname{supp}(p_X)|)$$

be the *Rényi-0 entropy* of p_X. For $\varepsilon > 0$, we let

$$H_0^\varepsilon(p_X) = \min\{\log|\Omega| : \Omega \text{ such that } \mathbb{P}(x \in \Omega) = \sum_x p_X(x) \geq 1 - \varepsilon\},$$

which is the *smoth Rényi-0 entropy* of p_X.

This is a special case of *Rényi entropy*, which is defined for $0 < \alpha < \infty$ and $\alpha \neq 1$ as

$$H_\alpha(p_X) = \frac{1}{1-\alpha} \log\left(\sum_x p_X(x)^\alpha\right).$$

For $\alpha = 0, 1, \infty$, it is defined as $\lim_{\beta\to\alpha} H_\beta(p_X)$. Clearly H_0 coincides with the above definition. Moreover H_1 is the well-known *Shannon entropy*, which we will see later in this chapter.

Theorem 7.4 (Classical one-shot compression)**.** *Suppose that $p_X \in P(X)$ and $\varepsilon \geq 0$. Then there exists (ε, r)-compression code for p_X if and only if $r \geq H_0^\varepsilon(p_X)$, so*

$$C^\varepsilon(p_X) = H_0^\varepsilon(p_X).$$

Proof (Sketch). From the definition of $H_0^\varepsilon(p_X)$, we can choose Ω such that $\mathbb{P}(x \in \Omega) \geq 1 - \varepsilon$ and $|\Omega| \leq 2^r$. Take the system C which we compress to be the set Ω and define $E(x) = x = D(x)$, for $x \in \Omega$, and $E(x)$ can by anything for $x \notin \Omega$. This is correct compression that has success probability at least $1 - \varepsilon$. We can at least compress to at least the size $H_0^\varepsilon(p_X)$, i.e., $C^\varepsilon(p_X) \geq H_0^\varepsilon(p_X)$.

Suppose that $\mathbb{P}(D(E(p_X)) = p_X) \geq 1 - \varepsilon$ We can assume that D and E are deterministic. We can always thing of applying a deterministic channel with some probability. We can define $\Omega = \{x : D(E(x)) = x\}$. The probability of being in this set must be at least $1 - \varepsilon$. We have $|\Omega| \leq |C|$, i.e., $H_0^\varepsilon(p_X) \leq C^\varepsilon(p_X)$. □

7.2 Quantum compression

Definition 7.5. We define a *quantum* (r, ε)*-compression code* of a state $\rho_A \in S(A)$ to r qubits with error $\varepsilon > 0$ as system C with dimension $|C| = \dim(\mathcal{H}_C)$ such that $\log |C| \leq r$, and a pair of quantum channels

$$\mathcal{E} \in C(A, C), \quad \mathcal{D} \in C(C, A)$$

which are such that

$$F_E(\mathcal{D} \circ \mathcal{E}, \rho_A) \geq 1 - \varepsilon.$$

Let

$$C^{\varepsilon}(\rho_A) = \min\{r : \text{there exists a } (\varepsilon, r)\text{-compression code for } \rho_A\}.$$

Definition 7.6. The *quantum Rényi entropy* of $\rho_A \in S(A)$ is

$$H_0(\rho_A) = \log(\text{rank}(\rho_A)).$$

For $0 < \varepsilon < 1$ the *quantum Rényi entropy* is given by

$$H_0^{\varepsilon}(\rho_A) = \min\{\log(\text{rank}(P_A)) : P_A \in \text{Lin}(A) \text{ projection such that } \text{tr}[P_A \rho_A] \geq 1 - \varepsilon\}.$$

Lemma 7.7. *Let $\rho_A \in S(A)$ and let p_X be the probability distribution given by the spectrum of ρ_A. Then*

$$H_0^{\varepsilon}(\rho_A) = H_0^{\varepsilon}(p_X).$$

Lemma 7.8. *Thre is a $(\varepsilon, H_0^{\varepsilon}(\rho_A))$-compression code for $\rho_A \in S(A)$*

Theorem 7.9 (Quantum one-shot compression)**.** *We have $H_0^{2\varepsilon}(\rho_A) \leq C^{\varepsilon}(\rho_A) \leq H_0^{\varepsilon}(\rho_A)$.*

7.3 Asymptotic compression

Let $X_1, \ldots, X_n$ be a sequence of IID random variables with distribution p_X. We write $X^n = (X_1, \ldots, X_n)$ and $x^n = (x_1, \ldots, x_n) \in \Omega^n$. We write p_{X^n} for the distribution of X^n, which is given by

$$p_{X^n}(x^n) = p_X(x_1) \cdots p_X(x_n).$$

Example 7.10. Consider a binary random variable which takes value 0 with probability 0.1 and value 1 with probability 0.9. Suppose we allow error probability $\varepsilon = 0.05$. If we get a single sample from this source we cannot compress, and therefore, if we have many samples X^n, we cannot do any compression if we only allow codes that compress each X_i separately.

However, if we take for example the distribution of X^3, we get

x^3	$p(x^3)$
000	0.001
001, 010, 100	0.009
011, 101, 110	0.081
111	0.729

Now we see that we can discard the outcomes $000, 001, 010, 100$ to achieve error probability below 0.05 and thus we need only $\log(4) = 2$ bits instead of 3. □

Definition 7.11. The *optimal rate of compression* for a random variable X with probability distribution p_X is

$$r(p_X) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} C^{\varepsilon}(p_{X^n}).$$

It means that $r(p_X)$ is the optimal value such that for any $\varepsilon, \delta > 0$ we can find n_0 such that are block codes with error probability at most ε for $n > n_0$ which need at most $r(p_X) + \delta$ bits per symbol, i.e., $\frac{1}{n} C^{\varepsilon}(p_{X^n}) \leq r + \delta$.

Definition 7.12. The *optimal rate of compression* for a quantum state $\rho_A \in S(A)$ is

$$r(\rho_A) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} C^{\varepsilon}(\rho_A^{\otimes n}).$$

By previous theorems, we have

$$r(p_X) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_0^{\varepsilon}(p_{X^n})$$

and

$$r(\rho_A) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_0^{\varepsilon}(\rho_A^{\otimes n}).$$

In the remaining, we will show that

$$r(p_X) = H(p_X) \quad \text{and} \quad r(\rho_A) = H(\rho_A),$$

where $H(p_X)$ is the *Shannon entropy* and $H(\rho_A)$ is the *von Neumann entropy.*

7.4 Classical and quantum entropy

It is reasonable to suppose that the amount of surprise generating by an event E should depend on the probability of E. For example, we would be less surprised to learn that event E occurred after rolling a pair of dice when E represents the event that the sum of the dice is even than when E represents the event that the sum of the dice is 12.

To quanitfy the concept of "surprise", we shall assume thatthe amount of surprise one feels upon learning that an event E occured only depends on the probability of E. Let $S(p)$ denote the the surprise generated by an event having probability p. What are some reasonable assumptions? First, let us assume that $S(p)$ is defined for $0 < p \leq 1$, but not for events having probability 0. These are some reasonable assumptions:

1. $S(1) = 0$; if an event is certain, then there is no surprise.
2. $S(p)$ is a strictly decreasing function; the the more unlikely an event is to occur, the greater is the surprise.
3. $S(p)$ is continuous.
4. $S(pq) = S(p) + S(q)$, for $0 < p, q \leq 1$. Consider two *independent* events E and F with $\mathbb{P}(E) = p$ and $\mathbb{P}(F) = q$. Since $\mathbb{P}(EF) = pq$, the surprise generated by the occurence of both E and F is $S(pq)$. Suppose now that we are told first about the occurence of E and later of F. Since $S(p)$ is the surprise evoked by E, if follows that $S(pq) - S(p)$ is the additional surprise evoked when we are informed that F also occured. Since F is independent of E, the knowledge that E occured does not change the probability of F. So, the additional surprise should be just $S(q)$.

Proposition 7.13. *If $S(\cdot)$ satisfies the items 1.–4., then $S(p) = -C \log_2 p$, where C is an arbitrary positive integer.*

Proof. From item 4., we have

$$S(p^m) = mS(p).$$

For an integer n, we have

$$S(p) = S(p^{1/n} \cdots p^{1/n}) = nS(p^{1/n}),$$

which implies

$$S(p^{1/n}) = \frac{1}{n} S(p).$$

So, from the previous two equations, we get

$$S(p^{m/n}) = mS(p^{1/n}) = \frac{m}{n} S(p).$$

Thus,

$$S(p^x) = xS(p),$$

for every positive rational number x. By item 3 (continuity), this equation if valid for all real numbers $x \geq 0$.

For any $0 < p \leq 1$, let $x = -\log_2 p$. Then $p = (\frac{1}{2})^x$ and from the last equation we get

$$S(p) = S\left(\left(\frac{1}{2}\right)^x\right) = xS\left(\frac{1}{2}\right) = -C \log_2 p,$$

where $C = S(\frac{1}{2}) > S(1) = 0$ by item 1. and item 2. □

For our purposes we will set $C = 1$. All logarithms are to base 2. We also make the convention that $0 \log 0 = 0$

Definition 7.14. Let p_X be a probability distribution. Then the *Shannon entropy* of p_X is given by

$$H(p_X) = -\sum_x p_X(x) \log(p_X(x)).$$

We also write $H(X)$ for $H(p_X)$. We can also write $H(X) = -\mathbb{E} \log(X)$.

Lemma 7.15. *Let p_X be a probability distribution.*

(a) $H(X) \geq 0$ and $H(X) = 0$ if and only if p_X is deterministic.

(b) $H(X) \leq H_0(p_X) \leq \log(|X|)$ and the equality occurs if and only if p_X is the uniform on all outcomes.

Definition 7.16. The *von Neumann entropy* of a state $\rho_A \in S(A)$ is given by

$$H(\rho_A) = -\operatorname{tr}[\rho_A \log \rho_A].$$

Lemma 7.17. *Let $\rho_A \in S(A)$.*

(a) $H(\rho_A) \geq 0$ and $H(\rho_A) = 0$ if and only if ρ_A is pure.

(b) $H(\rho_A) \leq H_0(\rho_A) \leq \log(\dim(\mathcal{H}_A))$, and the equality occurs if and only if ρ_A is the maximally mixed state.

7.5 Source coding theorems

The main technical tool are *typical sets*.

Definition 7.18. If X^n is an IID sequence of random variables with distribution p_X and $\varepsilon > 0$, the the *typical set* $T_{n,\varepsilon}(p_X)$ is

$$T_{n,\varepsilon} = \left\{ x^n \in \Omega_X^n : \left| \frac{1}{n} \sum_{i=1}^{n} -\log(p_X(x_i)) - H(X) \right| \leq \varepsilon \right\}.$$

The quantity $-\frac{1}{n}\log(p_{X^n}(x^n))$ represents "surprisal per symbol". This is equal to the sum in the definition of the typical set.

Lemma 7.19. *The typical set $T_{n,\varepsilon}(p_X)$ has the following properties.*

(a) For $x^n \in T_{n,\varepsilon}(p_X)$, we have

$$2^{-n(H(X)+\varepsilon)} \leq p(x^n) \leq 2^{-n(H(X)-\varepsilon)}.$$

(b) We have

$$|T_{n,\varepsilon}(p_X)| \leq 2^{n(H(X)+\varepsilon)}$$

(c) We have

$$\lim_{n\to\infty} \mathbb{P}(x^n \in T_{n,\varepsilon}(p_X)) = 1.$$

Proof. (a) If $x^n \in T_{n,\varepsilon}(p_X)$, then by definition, we have

$$\frac{1}{2^{n(H(X)+\varepsilon)}} \leq p(x^n) \leq \frac{1}{2^{n(H(X)-\varepsilon)}}.$$

(b) We have

$$1 \leq \mathbb{P}(x^n \in T_{n,\varepsilon}(p_X)) = \sum_{x^n \in T_{n,\varepsilon}(p_X)} p(x^n) \geq |T_{n,\varepsilon}(p_X)| \frac{1}{2^{n(H(X)+\varepsilon)}}.$$

Define $\mathbf{Y}_i = -\log(p(\mathbf{X}_i))$. By construction $\mathbb{E}\mathbf{Y}_i = H(X)$. The weak law of large numbers implies that

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i - H(X) \right| \geq \varepsilon \right) = 0.$$

This implies the result. □

Theorem 7.20 (Shannon)**.** *Let p_{X^n} be the distribution of X^n. The the optimal asymptotic rate of compression is given by $r(p_X) = H(X)$.*

Proof. Recall that we have

$$r(p_X) = \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} H_0^\varepsilon(p_{X^n}).$$

Thus, we need to estimate $\frac{1}{n} H_0^\varepsilon(p_{X^n})$.

For the upper bound, let $\delta > 0$. Let $\varepsilon(n) = \mathbb{P}(x^n \notin T_{n,\varepsilon}(p_X))$. By the previous lemma, $\varepsilon(n) \to 0$ as $n \to \infty$. We have

$$\frac{1}{n} H_0^{\varepsilon(n)}(p_{X^n}) \leq \frac{1}{n} \log(|T_{n,\delta}|) \leq H(p_X) + \delta.$$

The second inequality is from the definition of H_0^ε and the third is from the previous lemma.

For the lower bound, let $\delta > 0$. Further, let $\varepsilon(n)$ be any sequence such that $\varepsilon(n) \to 0$ as $n \to \infty$. By definition of H_0^ε, there is a set Ω_n such that $H_0^{\varepsilon(n)}(p_{X^n}) = \log(|\Omega_n|)$. We have

$$1 - \varepsilon(n) \leq \mathbb{P}(x^n \in \Omega_n) \leq \mathbb{P}(x^n \in \Omega_n \cap T_{n,\delta}(p_X)) + \mathbb{P}(x^n \notin T_{n,\delta}(p_X)).$$

The first inequality is just from the definition of H_0^ε and the second inequality is the union bound. Using the previous lemma, the first term on the right hand side can be bounded as

$$\mathbb{P}(x^n \in \Omega_n \cap T_{n,\delta}(p_X)) = \sum_{x^n \in \Omega_n \cap T_{n,\delta}} p_{X^n}(x^n) \leq |\Omega_n \cap T_{n,\delta}| \frac{1}{2^{n(H(X)-\delta)}} \leq |\Omega_n| \frac{1}{2^{n(H(X)-\delta)}}.$$

This implies that

$$\begin{aligned} |\Omega_n| &\geq (1 - \varepsilon(n) - \mathbb{P}(x^n \notin T_{n,\delta}(p_X)))2^{n(H(X)-\delta)} \\ \log(|\Omega_n|) &\geq \log(1 - \varepsilon(n) - \mathbb{P}(x^n \notin T_{n,\delta}(p_X))) + n(H(X) - \delta). \end{aligned}$$

Thus,

$$\frac{1}{n} H_0^{\varepsilon(n)}(p_X) = \frac{1}{n} \log(|\Omega_n|) \geq H(X) - \delta + f(n),$$

where $f(n) \to 0$ as $n \to \infty$, by the previous lemma. □

If $\rho_A \in S(A)$ has spectral decomposition

$$\rho_A = \sum_x p(x) |\psi_x\rangle\langle\psi_x|,$$

then the state $\rho_A^{\otimes n}$ has spectral decomposition

$$\rho_A^{\otimes n} = \sum_{x_1,\ldots,x_n} p(x_1) \cdots p(x_n) |\psi_{x_1}\rangle\langle\psi_{x_1}| \otimes \cdots \otimes |\psi_{x_n}\rangle\langle\psi_{x_n}|.$$

We define the *typical subspace* $S_{n,\varepsilon}(\rho_A)$ for n copies of ρ_A and $\varepsilon > 0$, by restricting to the subspace spanned by all $|\psi_{x_1}\rangle \cdots |\psi_{x_n}\rangle$ such that $x^n \in T_{n,\varepsilon}(p)$:

$$S_{n,\varepsilon}(\rho_A) = \text{span}\{|\psi_{x_1}\rangle \cdots |\psi_{x_n}\rangle \in \mathcal{H}_A^{\otimes n} : x^n \in T_{n,\varepsilon}(p)\}.$$

The *typical projector* is the projection $\Pi_{n,\varepsilon}$ onto $S_{n,\varepsilon}(\rho_A)$.

Lemma 7.21. *Let $\rho_A \in S(A)$ and $\varepsilon > 0$.*

(a) The non-zero eigenvalues of $\Pi_{n,\varepsilon}\rho_A^{\otimes n} = \Pi_{n,\varepsilon}\rho_A^{\otimes n}\Pi_{n,\varepsilon}$ are all in the interval

$$[2^{-n(H(\rho_A)+\varepsilon)}, 2^{-n(H(\rho_A)-\varepsilon)}].$$

(b) The dimension of the typical subspace is bounded by

$$\dim(S_{n,\varepsilon}(\rho_A)) \leq 2^{n(H(\rho_A)+\varepsilon)}.$$

(c) We have

$$\lim_{n\to\infty} \operatorname{tr}[\Pi_{n,\varepsilon}\rho_A^{\otimes n}] = 1,$$

that is, as $n \to \infty$, if we measure whether we are in the typical subspace (corresponding to the measurement $\{\Pi_{n,\varepsilon}, \mathbb{1}_{A^n} - \Pi_{n,\varepsilon}\}$), the probability of being in the typical subspace goes to 1.

Theorem 7.22 (Schumacher)**.** *The optimal asymptotic rate of compressing a quantum state $\rho = \rho_A \in S(A)$ is given by $r(\rho) = H(\rho)$.*

Proof. Similarly as in the classical case, we start with

$$r(\rho_A) = \lim_{\varepsilon\to 0}\lim_{n\to\infty} \frac{1}{n} H_0^{\varepsilon}(\rho_A^{\otimes n})$$

and we would like to derive and upper and lower bound.

We start with the upper bound. Let $\delta > 0$ and let $\Pi := \Pi_{n,\delta}$ be the typical projector onto $S_{n,\delta}(\rho)$. Let $\varepsilon(n) := \mathbb{P}(\text{not typical}) = 1 - \operatorname{tr}[\Pi\rho^{\otimes n}]$. By the previous lemma $\varepsilon(n) \to 0$ as $n \to \infty$. Clearly Π is a valid projector in the definition of $H_0^{\varepsilon(n)}(\rho^{\otimes n})$ since $\operatorname{tr}[\Pi\rho^{\otimes n}] \geq 1 - \varepsilon(n)$. Therefore,

$$H_0^{\varepsilon(n)} \leq \log(\operatorname{rank}(\Pi)) = \log(\dim(S_{n,\delta}(\rho))) \leq n(H(\rho) + \delta).$$

Hence,

$$\frac{1}{n} H_0^{\varepsilon(n)}(\rho^{\otimes n}) \leq H(\rho) + \delta.$$

Now, we proceed with the lower bound. Let $\delta > 0$. Let $\varepsilon(n) \to 0$ be any sequence. Let P_n be a projector achieveing the minimum in the definition of $H_0^{\varepsilon(n)}(\rho^{\otimes n}$, so

$$H_0^{\varepsilon(n)}(\rho^{\otimes n}) = \log(\operatorname{rank}(P_n)), \quad \operatorname{tr}[P_n\rho^{\otimes n}] \geq 1 - \varepsilon(n).$$

Now

$$\begin{aligned} 1 - \varepsilon(n) \leq \operatorname{tr}[P_n\rho^{\otimes n}] &= \operatorname{tr}[P_n(\Pi + (\mathbb{1} - \Pi))\rho^{\otimes n}] \\ &= \operatorname{tr}[P_n\Pi\rho^{\otimes n}] + \operatorname{tr}[P_n(\mathbb{1} - \Pi)\rho^{\otimes n}] \\ &\leq \operatorname{tr}[P_n\Pi\rho^{\otimes n}] + \operatorname{tr}[(\mathbb{1} - \Pi)\rho^{\otimes n}], \end{aligned}$$

where the last inequality frollows from the fact that $\mathbb{1} - \Pi$ is a projection that commutes with $\rho^{\otimes n}$, so $(\mathbb{1} - \Pi)\rho^{\otimes n}(\mathbb{1} - \Pi)$ is positive semidefinite. For a positive semidefinite matrix X and a projection P, we have $\operatorname{tr}[PX] \leq \operatorname{tr}[X]$. Further, we bound

$$\operatorname{tr}[P_n\Pi\rho^{\otimes n}] \leq \|P_n\|_1 \left\|\Pi\rho^{\otimes n}\right\|_\infty \leq \operatorname{rank}(P_n)2^{-n(H(\rho)-\delta}},$$

we are using the previous lemma.

$$\begin{aligned} \operatorname{rank}(P_n) &\geq (1 - \varepsilon(n) - \operatorname{tr}[(\mathbb{1} - \Pi)\rho^{\otimes n}])2^{n(H(\rho)-\delta)} \\ \log(\operatorname{rank}(P_n)) &\geq \log(1 - \varepsilon(n) - \operatorname{tr}[(\mathbb{1} - \Pi)\rho^{\otimes n}]) + n(H(\rho) - \delta). \end{aligned}$$

Thus,

$$\frac{1}{n} H_0^{\varepsilon(n)}(\rho) = \frac{1}{n}\log(\operatorname{rank}(P_n)) \geq H(\rho) - \delta + f(n),$$

where $f(n) \to 0$ as $n \to \infty$, by the previous lemma. □

Chapter 8

Bounds on Information Processing

8.1 Entropy inequalities

If we have two systems A and B with some state $\rho_{AB} \in S(AB)$, then we have entropies $H(A)$, $H(B)$, and $H(AB)$.

Lemma 8.1. *If ρ_{AB} is pure, then $H(AB) = 0$ and $H(A) = H(B)$. If ρ_{AB} is a product state, then $H(AB) = H(A) + H(B)$.*

Proof. The first part follows from previous section and Schmidt decomposition.

For the second part, suppose that $\rho_{AB} = \rho_A \otimes \rho_B$ and

$$\rho_A = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad \sum_j q_j |\phi_j\rangle\langle\phi_j|$$

are corresnponding spectral decompositions. Then

$$\rho_{AB} = \sum_{i,j} p_i q_j |\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j|$$

is the spectral decomposition of $\rho_A \otimes \rho_B$, with eigenvalues $p_i q_j$ corresnponding to eigenvectors $|\psi_i\rangle \otimes |\phi_j\rangle$. The function $x \log x$ is defined for all $x \geq 0$, after definiing $0 \log 0 := 0$. We have

$$\begin{aligned}
\rho_A \otimes \rho_B \log(\rho_A \otimes \rho_B) &= \sum_{i,j} p_i q_j \log(p_i q_j) |\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j| \\
&= \sum_i p_i \log(p_i) |\psi_i\rangle\langle\psi_i| \otimes \sum_j q_j |\phi_j\rangle\langle\phi_j| + \sum_i p_i |\psi_i\rangle\langle\psi_i| \otimes \sum_j q_j \log(q_j) |\phi_j\rangle\langle\phi_j| \\
&= \rho_A \log(\rho_A) \otimes \rho_B + \rho_A \otimes \rho_B \log(\rho_B).
\end{aligned}$$

Taking traces of boths sides, we get the second part of the statement. □

Lemma 8.2. *Let $p_{XY} \in \mathcal{P}(XY)$. The Shannon entropy satisfies:*

- (monotonicity) $H(XY) \geq H(Y)$.
- (subadditivity) $H(XY) \leq H(X) + H(Y)$.

Subadditivity is also true for the von Neumann entropy, that is, if $\rho_{AB} \in S(AB)$, then $H(AB) \leq H(A) + H(B)$. A possible proof is based on the operational interpretation of entropy in terms of compression from the previous chapter. Intuitively, we can either separately compress A and B at the rate $H(A) + H(B)$ or we can compress the joni system which may lead to a more eefficient compression at rate $H(AB)$.

Monotonicity is not true for von Neumann entropy. For instance if ρ_{AB} is pure and entnangled then $H(AB) = 0$ and $H(A) > 0$ since the reduced state is maximally mixed.

Instead, we have *strong subadditivity (SSA)* and equivalently *weak monotonicity (WM).*

Theorem 8.3 (Strong subadditivity)**.** *If $\rho_{ABC} \in S(ABC)$, then*

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

Clearly if B is an empty system, then strong subadditivity is exactly subadditivity. The proofs of strong subadditivity is quite involved and we postpone it until the end of the chapter.

Theorem 8.4 (Weak monotonicity)**.** *If $\rho_{ABC} \in S(ABC)$, then*

$$H(A) + H(C) \leq H(AB) + H(BC).$$

Proof. Let ρ_{ABCD} be a purfication of ρ_{ABC}. By strong subadditivity

$$H(BCD) + H(C) \leq H(BC) + H(CD).$$

Since ρ_{ABCD} is pure, $H(BCD) = H(A)$ and $H(CD) = H(AB)$. □

Strong subadditivity is usefull for showing optimality; it is the main constraint. Weak monotonicity is related to *monogamy of entanglement* – it is not possible for Bob to be maximally entangled with Alice and Charlie. If state is pure on AB and entangled, then $H(AB) < H(A)$. Weak monotonicity tells us that we cannot have both $H(AB) < H(A)$ and $H(BC) < H(C)$.

If one of the systems is classical we do still have monotonicity. If we have an ensemble of states $\{p_X(x), \rho_{A,x}\}$ where we have state $\rho_{A,x} \in S(A)$ with probability $p_X(x)$, then we may model this by a classical–quantum system XA and a classical–quantum state

$$\rho_{XA} = \sum_x p_X(x) \, |x\rangle\langle x| \otimes \rho_{A,x}.$$

Lemma 8.5 (Lemma 9.6)**.** *Let ρ_{XA} be a classical–quantum state as in above, then*

$$H(XA) = \sum_x p_X(x) \, H(\rho_{A,x}) + H(p_X).$$

Moreover,

$$H(XA) \geq H(X) \qquad \textit{and} \qquad H(XA) \geq H(A).$$

Proof. We are given the classical–quantum state

$$\rho_{XA} = \sum_x p_X(x) \, |x\rangle\langle x| \otimes \rho_{A,x}.$$

Notice that, since the $\{|x\rangle\}$ form a basis of the X system, ρ_{XA} has a matrix block-diagonal form:

$$\rho_{XA} = \begin{pmatrix} p_X(1)\rho_{A,1} & 0 & \cdots \\ 0 & p_X(2)\rho_{A,2} & \cdots \\ \vdots & & \ddots \end{pmatrix},$$

hence to compute the logarithm of ρ_{XA} we can consider its action on the individual diagonal blocks. That is,

$$\log \rho_{XA} = \begin{pmatrix} \log\big(p_X(1)\rho_{A,1}\big) & 0 & \cdots \\ 0 & \log\big(p_X(2)\rho_{A,2}\big) & \cdots \\ \vdots & & \ddots \end{pmatrix} = \sum_x |x\rangle\langle x| \otimes \log\big(p_X(x)\rho_{A,x}\big).$$

Note also that

$$\log\big(p_X(x)\rho_{A,x}\big) = \log p_X(x)\, \mathbb{I}_A + \log \rho_{A,x}.$$

We now compute

$$\begin{aligned} H(XA)_\rho &= -\operatorname{tr}\big[\rho_{XA} \log \rho_{XA}\big] \\ &= -\operatorname{tr}\left[\left(\sum_x p_X(x)|x\rangle\langle x| \otimes \rho_{A,x}\right)\left(\sum_y |y\rangle\langle y| \otimes \big(\log p_X(y)\mathbb{1}_A + \log \rho_{A,y}\big)\right)\right] \\ &= -\sum_{x,y} p_X(x)\ \operatorname{tr}\big[\big(|x\rangle\langle x||y\rangle\langle y|\big) \otimes \big(\rho_{A,x}\log p_X(y) + \rho_{A,x}\log \rho_{A,y}\big)\big] \\ &= -\sum_x p_X(x)\,(\log p_X(x)\ \operatorname{tr}[\rho_{A,x}] + \operatorname{tr}[\rho_{A,x}\log \rho_{A,x}]) \\ &= -\sum_x p_X(x)\log p_X(x) - \sum_x p_X(x)\operatorname{tr}[\rho_{A,x}\log \rho_{A,x}] \\ &= H(p_X) + \sum_x p_X(x) H(\rho_{A,x}). \end{aligned}$$

Now we compute

$$\rho_X = \sum_x p_X(x)|x\rangle\langle x|\ \operatorname{tr}[\rho_{A,x}] = \sum_x p_X(x)|x\rangle\langle x|,$$

so $H(X)_\rho = H(p_X)$.

Comparing with the expression for $H(XA)_\rho$, and noticing that $H(\rho_{A,x}) \geq 0$ for all x, we see that

$$H(XA)_\rho \geq H(X)_\rho.$$

□

8.2 The conditional entropy

For a joint probability distribution p_{XY} we have *conditional probabilities* $p_{X|Y}$, which is the probability of x given y and which is defined by the relation $p_{XY}(x,y) = p_{X|Y}(x)\, p_Y(y)$. In particular, we may consider the *entropy of X given y*:

$$H(X \mid Y = y) := -\sum_x p_{X|y}(x)\ \log\big(p_{X|y}(x)\big).$$

The *conditional Shannon entropy* of X given Y is the expected value of $H(X \mid Y = y)$:

$$H(X \mid Y) = \sum_y p_Y(y)\, H(X \mid Y = y).$$

The intuitive interpretation of this expression is that it is the expected amount of information in X once we learn the outcome of Y. An easy calculations shows

$$H(X \mid Y) = H(XY) - H(Y).$$

In words, the information we have about X when we know Y is the total information on XY minus the amount of information in Y.

i

Definition 8.6 (Conditional entropy)**.** If $\rho_{AB} \in \mathcal{S}(AB)$, the *conditional entropy of A conditioned on B* is defined as

$$H(A \mid B)_\rho = H(AB)_\rho - H(B)_\rho.$$

We omit the dependence on ρ_{AB} and write $H(A \mid B)$ if the state is clear from the context.

It has the perhaps surprising property that it is possible that $H(A \mid B) < 0$ (since the von Neumann entropy is not monotonic). We will see a nice operational interpretation of this fact later.

Example 8.7. We compute the conditional entropy for three important examples of states on two qubits A and B.

- If $\rho_{AB} = \frac{1}{4}\,\mathbb{I}_{AB}$ is the maximally mixed state, we have

  $$H(AB) = 4 \times \tfrac{1}{4}\log(4) = 2, \qquad H(B) = 2 \times \tfrac{1}{2}\log(2) = 1,$$

  so $H(A \mid B) = 1 = H(A)$. We see that $H(A \mid B) = H(A)$, so when we get B we learn nothing about A and the amount of information in A stays the same.
- If $\rho_{AB} = \frac{1}{2}\big(|00\rangle\langle 00| + |11\rangle\langle 11|\big)$ is the maximally correlated state, we see that ρ_{AB} has nonzero eigenvalues $\frac{1}{2}, \frac{1}{2}$, while the reduced density matrices are maximally mixed, so

  $$H(AB) = \tfrac{1}{2}\log(2) + \tfrac{1}{2}\log(2) = 1, \qquad H(B) = \tfrac{1}{2}\log(2) + \tfrac{1}{2}\log(2) = 1.$$

  Hence $H(A \mid B) = 0$. This makes sense with our (classical) interpretation: if we learn the outcome of B we know exactly what A is, and hence there is no information left in A.
- If $\rho_{AB} = |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|$ is a maximally entangled state, then $H(AB) = 0$ since the state is pure, and as the reduced states are maximally mixed we have $H(B) = 1$, and hence

  $$H(A \mid B) = -1.$$

Lemma 8.8. *Let $\rho_{AB} \in \mathcal{S}(AB)$, then $H(A \mid B) = H(A \mid B)_\rho$ has the following properties.*

(a) If ρ_{AB} is pure, then $H(A|B) = -H(A) = -H(B)$. If ρ_{ABC} is pure, $H(A|B) = -H(A|C)$.

(b) We have the lower bound

$$H(A|B) \geq -\min\big(H(A), H(B)\big) \geq -\log(|A|).$$

If the system X is classical, we have $H(A \mid X) \geq 0$ and $H(X|A) \geq 0$.

(c) We have the upper bound

$$H(A|B) \leq H(A) \leq \log(|A|).$$

The first inequality is an equality if and only if $\rho_{AB} = \rho_A \otimes \rho_B$ is a product state.

(d) The conditional entropy is invariant under isometries on the subsystems. That is, if $V \in \mathrm{Isom}(A, A')$ and $W \in \mathrm{Isom}(B, B')$, and $\sigma_{A'B'} = (V \otimes W)\rho_{AB}(V^\dagger \otimes W^\dagger)$, then

$$H(A|B)_\rho = H(A' \mid B')_\sigma.$$

Proof. If ρ_{ABC} is pure, then

$$H(A|B) = H(AB) - H(B) = H(C) - H(AC) = -H(A|C).$$

If ρ_{AB} is pure, $H(AB) = 0$ and $H(A) = H(B)$, proving (a). Next, for (b), it is clear from the definition that $H(A \mid B) = H(AB) - H(B) \geq -H(B)$. To prove $H(A \mid B) \geq -H(A)$, let ρ_{ABC} be a purification of ρ_{AB}. Then strong subadditivity gives

$$H(ABC) + H(B) \leq H(AB) + H(BC),$$

where $H(ABC) = 0$ and $H(BC) = H(A)$ using that ρ_{ABC} is pure. Rearranging the terms gives $H(A \mid B) = H(AB) - H(B) \geq -H(A)$. If X is classical, the previous lemma implies that $H(A \mid X)$ and $H(X \mid A)$ are non-negative. Subadditivity of the von Neumann entropy is equivalent to (c). Finally, (d) is a direct consequence of the invariance of the von Neumann entropy under isometries. □

Theorem 8.9 (Data processing conditional entropy)**.** *If $\Phi_{B\to C} \in C(B, C)$ and we have $\rho_{AB} \in S(AB)$, $\sigma_{AC} = (\mathbb{1}_A \otimes \Phi_{B\to C})(\rho_{AB})$, then*

$$H(A \mid C)_\sigma \geq H(A \mid B)_\rho.$$

Proof. Let $V \in \mathrm{Isom}(B, CE)$ be a Stinespring extension of $\Phi_{B\to C}$, and let $\omega_{ACE} = (\mathbb{I}_A \otimes V)\rho_{AB}(\mathbb{I}_A \otimes V^\dagger)$. By invariance of the von Neumann entropy under isometries we have

$$H(A \mid B)_\rho = H(A \mid CE)_\omega.$$

We now use strong subadditivity, which states that $H(ABC) + H(B) \leq H(AB) + H(BC)$. Rewriting this inequality gives

$$H(ABC) - H(BC) \leq H(AB) - H(B),$$

and hence

$$H(A \mid BC) \leq H(A \mid B).$$

Applying this to the state ω_{ACE} yields $H(A \mid CE)_\omega \leq H(A \mid C)_\omega$. Since $\omega_{AC} = \sigma_{AC}$, this implies

$$H(A \mid B)_\rho = H(A \mid CE)_\omega \leq H(A \mid C)_\sigma,$$

which proves the claim. □

The intuition for this is that it is a *monogamy of entanglement* relation. The conditional entropy is maximally negative if the state is a pure maximally entangled state. Weak monotonicity says that we cannot have the situation where ρ_{AB} and ρ_{BC} are both very entangled, since the sum of the conditional entropies must be positive.

8.3 The mutual information

Another natural entropic quantity is the *mutual information*.

Definition 8.10 (Mutual information)**.** Given classical X and Y we define the mutual information as

$$I(X:Y) = H(X) + H(Y) - H(XY),$$

and similarly for a quantum state we let $\rho_{AB} \in \mathcal{S}(AB)$

$$I(A:B)_\rho = H(A)_\rho + H(B)_\rho - H(AB)_\rho.$$

We write $I(A:B)$ if the state is clear from the context.

The idea is that it is a measure for the correlation between A and B. You can think of it as 'the amount of information you can learn about A from B'. The mutual information is related to the conditional entropy as follows (as seen directly from the definition):

$$I(A:B) = H(A) - H(A \mid B) = H(B) - H(B \mid A).$$

Example 8.11. We compute the mutual information entropy for the same qubit states as for conditional entropy.

- If $\rho_{AB} = \frac{1}{4}\mathbb{I}_{AB}$ is the maximally mixed state we have

  $$H(AB) = 2 \qquad H(A) = H(B) = 1$$

  so $I(A:B) = 1 + 1 - 2 = 0$. Indeed, A and B are independent, so we learn nothing about A from B.
- If $\rho_{AB} = \frac{1}{2}\big(|00\rangle\langle 00| + |11\rangle\langle 11|\big)$ is the maximally correlated state

  $$H(AB) = 1 \qquad H(A) = H(B) = 1$$

  and hence $I(A:B) = 1 + 1 - 1 = 1$. The maximally correlated state indeed represents one bit of correlation.
- If $\rho_{AB} = |\Phi^+_{AB}\rangle\langle\Phi^+_{AB}|$ is a maximally entangled state

  $$H(AB) = 0 \qquad H(A) = H(B) = 1$$

  and therefore $I(A:B) = 1 + 1 - 0 = 2$, so this is a 'stronger' correlation than for the maximally correlated state.

The mutual information has the following basic properties.

Lemma 8.12. *Let $\rho_{AB} \in \mathcal{S}(AB)$, then the mutual information $I(A:B) = I(A:B)_\rho$ has the following properties:*

(a) If ρ_{AB} is pure, then $I(A:B) = 2H(A) = 2H(B)$.

(b) $I(A:B) \geq 0$ with equality if and only if $\rho_{AB} = \rho_A \otimes \rho_B$.

(c) We have the upper bound

$$I(A:B) \leq 2\min\big(H(A), H(B)\big) \leq 2\min\big(\log(|A|), \log(|B|)\big).$$

(d) If the system X is classical, then

$$I(X:B) \leq \min\big(H(X), H(B)\big) \leq \min\big(\log(|X|), \log(|B|)\big).$$

(e) The mutual information is invariant under isometries on the subsystems. That is, if $V \in \mathrm{Isom}(A, A')$ and $W \in \mathrm{Isom}(B, B')$, and $\sigma_{A'B'} = (V \otimes W)\rho_{AB}(V^\dagger \otimes W^\dagger)$, then

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

Proof. This follows directly from the properties of the conditional entropy. □

Finally, we have a data processing inequality for the mutual information:

Theorem 8.13 (Data processing mutual information)**.** *If $\Phi_{B\to C} \in \mathcal{C}(B, C)$, then for $\rho_{AB} \in \mathcal{S}(AB)$, $\sigma_{AC} = (\mathbb{I}_A \otimes \Phi_{B\to C})(\rho_{AB})$ we have*

$$I(A:B)_\rho \geq I(A:C)_\sigma.$$

Proof. This is a direct consequence of the data processing inequality for the conditional entropy. □

As before, it has the intuitive interpretation that by only acting on one of the subsystems we can never get more information about the other subsystem. This statement is easily seen to be equivalent to strong subadditivity, and assigns a nice operational meaning to strong subadditivity.

8.4 The Holevo bound

We will now investigate the question of how much classical information we can encode in a quantum state. We already know, from the superdense coding protocol, that if we have entanglement available we can send over two classical bits using one qubit. Now we will look at the situation where we try to encode some classical register X into an ensemble of quantum states where we have state $\rho_{A,x} \in \mathcal{S}(A)$ with probability p_x. This gives rise to an associated classical–quantum state

$$\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_{A,x}.$$

Definition 8.14. We define the *Holevo χ-quantity* of an ensemble $\{p_x, \rho_{A,x}\}$ as

$$\chi(\{p_x, \rho_{A,x}\}) = I(X:A)_\rho.$$

Writing out the definition, using Lemma 9.6 we see that

$$\chi(\{p_x, \rho_{A,x}\}) = H\left(\sum_x p_x \rho_{A,x}\right) - \sum_x p_x H(\rho_{A,x}).$$

Moreover, by Lemma 9.13 we have

$$0 \leq \chi(\{p_x, \rho_{A,x}\}) \leq \min\big(H(p), H(\rho_A)\big). \tag{8.4.1}$$

Note that the upper bound, which you can show in Exercise 9.5, relies on strong subadditivity again!

Now we think of the following set-up: Alice has a classical source X and chooses to encode this using an ensemble of quantum states (i.e. if she has classical x she encodes this into $\rho_{A,x}$). She then sends over the state to Bob, who will do a measurement, to a classical register Y. The question is how much Bob can learn about X. An upper bound is given by

Theorem 8.15 (Holevo bound)**.** *The mutual information between X and Y is upper bounded by*

$$I(X:Y) \leq \chi(\{p_x, \rho_{A,x}\}).$$

Proof. The final state is obtained by taking the classical–quantum state ρ_{XA} and applying a measurement channel $\Phi_{A\to Y}$ to the A-system, so the classical state between X and Y is given by $\sigma_{XY} = (\mathbb{I}_X \otimes \Phi_{A\to Y})(\rho_{XA})$ and we have

$$I(X:Y)_\sigma \leq I(X:A)_\rho = \chi(\{p_x, \rho_{A,x}\}),$$

so the result follows directly from the data processing inequality in Theorem 9.14. □

This shows that if we try to encode into an n-qubit system A, the Holevo quantity is upper bounded by $H(A) \leq \log(d_A) = n$, and we can not achieve more than n bits of mutual information between Alice and Bob by sending over one qubit. In other words, this proves that the use of shared entanglement in superdense coding is necessary!

Remark 8.16. Why does $I(X:Y) \leq n$ mean that we can not communicate more than n bits? Note first that if there exists a (classical) channel from Y to X' which is such that it exactly recovers the original message, and if the source X has information content $H(X)$, then $I(X:X') = H(X)$, so for exact transfer of the source we have $H(X) \leq n$.

8.5 TODO

- Proof of SSA.

Chapter 9

Quantum Key Distribution

Appendix A

Tensor Product