Quantum Information Theory

quinten tupker

January 22 2021 - March 5, 2021

Introduction

These notes are based on the course lectured by Professor S Strelchuk in Lent 2020. This was lectured online due to measures taken to counter the spread of Covid-19 in the UK. These are not necessarily an accurate representation of what was lectures, and represent solely my personal notes on the content of the course, combinged with probably, very very many personal notes and digressions... Of course, any corrections/comments would be appreciated.

Information theory is the theory of information storage and transmission. It provides the theoretical limits on what is possible with information technologies in much of our world, and a framework to study many other fields, such as animal communication as well. This is the quantum qersion of that theory.

1 Classical Information Theory

We begin by observing that information is closely related to uncertainty. In particular, what might be able to say it is the opposite of uncertainty, and so then, it is no surprise that we build our theory of information using concepts from probability theory. As such we define

Definition 1. The **surprisal** of random variable X taking values in discrete finite **alphabet** J according to distribution $p(x)$ is

$$\mathcal{I}(x) = -\log(p(x))$$

Definition 2. The **Shannon entropy** of a random variable X is

$$H(X) = -\sum_{x\in J} p(x)\log(p(x))$$

(the logarithm is base 2)

This may appear to be a somewhat arbitrary definition, but it has a strong theoretical basis given that

Theorem 1. *The **Shannon Source Coding Theorem** states (informally) that the limit that information can be compressed so that it can be reliably retrieved is the Shannon entropy of the source.*

Here a basic example of a source is a **memoryless source** which is an object producing a sequence of signals, but since it is memoryless, each signal is completely independent from any other, so $\mathbb{P}(u_1, \ldots, u_n) = \mathbb{P}(u_1) \ldots \mathbb{P}(u_n)$. It is also known as an **i.i.d. information source**.

But how do we actually compress information? Conceptually there are two ways

- a **variable length encoding** stores higher probability signals in shorter codes, and lower probability signals in longer codes.
- a **fixed length encoding** stores higher probability signals in unique fixed length codes, and lower probability signals in the same fixed code.

Example 1. *An example of a fixed length code for the numbers* $1, \ldots, 8$ *are their binary representations, but if we also know that* $p(1, \ldots, 8) = 1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64$ *then the code* $C(1, \ldots, 8) = 0, 10, 110, 1110, 111100, 111101, 111110, 111111$ *has an average length 2 compared to the fixed length of 3. Furthermore, the Shannon entropy of this source is 2 as well, so this is maximally efficient.*

1.1 Classical Data Compression

Let's start making the definitions necessary to formalise compression.

Definition 3. A **compression map** is a map $C^n : u^{(n)} = (u_1, \ldots, u_n) \mapsto x(x_1, \ldots, x_{nR})$ sending a **message** u to a **code** x.

Definition 4. A **decompression map** D^n sends $D^n : x \in \{0,1\}^{\lceil nR \rceil} \mapsto u'^{(n)}$ with probability $\mathbb{P}(u^{(n)}|x)$.

Definition 5. A **code** of rate R and blocklength n is the triple (C^n, D^n, R).

Here we can compute the **probability of error** as

$$P_{av}^{(n)}(C_n) = \sum_{u^{(n)} \in J^n} \mathbb{P}(u^{(n)}) \mathbb{P}(D^n(C^n(u^{(n)})) \neq u^{(n)}). \tag{1}$$

[End of lecture 1]

Definition 6. A compression/decompression scheme is **reliable** iff $\forall \epsilon > 0 \exists$ sequence of codes C_n such that $\lim_{n\to\infty} P_{av}^{(n)}(C_n) = 0$

(does this require the code to i.i.d?) As such one could define the data compression rate as

$$\inf\{R : \exists C_n = \{C^n, D^n, R\} : \lim_{n\to\infty} P_{av}^{(n)}(C_n) = 0\} \tag{2}$$

1.2 Typical Sequences

It is honestly surprising how easy it is to prove Shannon's source coding theorem, since we find we in fact only need one simple, and honestly rather crude tool to do so. That tool is

Definition 7. a **typical set**, denoted $T_\epsilon^{(n)}$, which is the set of sequences $u = (u_1, \ldots, u_n)$ satisfying

$$2^{-n(H(u)+\epsilon)} \leq \mathbb{P}(u) \leq 2^{-n(H(u)-\epsilon)} \tag{3}$$

Why "typical"? Because if we have a memoryless source generating a sequence of length n, then a "typical" sequence occurs with probability

$$\prod_{u \in J} \mathbb{P}(u)^{n\mathbb{P}(u)} = 2^{-nH(u)} \tag{4}$$

(a typical sequence would have the expected value as the number of occurences of each letter in the alphabet). We consequently find the following theorem (not proven)

Theorem 2. $\forall \epsilon, \delta > 0, \exists n$ *such that*

1. $u \in T_\epsilon^{(n)} \implies H(u) - \epsilon \leq \frac{-1}{n} \log(\mathbb{P}(u)) \leq H(u) + \epsilon$
2. $\mathbb{P}(T_\epsilon^{(n)}) > 1 - \delta$
3. $|T_\epsilon^{(n)}| \leq 2^{n(H(u)+\epsilon)}$
4. $|T_\epsilon^{(n)} > (1-\delta)2^{2(H(u)-\epsilon)}$

and so

Corollary 1. $\forall \epsilon, \delta > 0, \exists n_0, \forall n > n_0, J^n$ *decomposes into the disjoint* ***atypical and typical sets*** $A_\epsilon^{(n)}, T_\epsilon^{(n)}$ *satisfying*

1. $\mathbb{P}(A_\epsilon^{(n)}) < \delta$
2. $2^{-n(H(u)+\epsilon)} \leq \mathbb{P}(u) \leq 2^{-n(H(n)-\epsilon)}$ *(on the typical set?)*

That allows us to formally state

Theorem 3. ***Shannon's source coding theorem****, which claims that for a i.i.d. source U, if $R > H(U)$, then we can find a reliable compression scheme and if $R < H(U)$ there are no reliable compression schemes.*

Proof. If $R > H(U)$ then pick $\epsilon > 0$ such that $H(U) + \epsilon < R$ and n such that $T_\epsilon^{(n)}$ satisfies our typical set theorem condiions. Then for $\delta > 0$ there are at most $2^{n(H(U)+\epsilon)} < 2^{nR}$ ϵ-typical sequences. Our compression scheme then works according to

1. Split J^n into typical and atypical sequences
2. order typical sequences somehow (say lexicographically), assigning each sequence a binary index.
3. typical sequences are sent to their binary code, prefixed with a 1, leading to a total length of $\lceil nR \rceil + 1$
4. atypical sequences are all sent to the fixed string $00\ldots0$ of length $\lceil nR \rceil + 1$

If $R < H(U)$ there does not exist a reliable compression scheme (see lemma below). □

Lemma 1. *For a collection of strings of length n, $S(n)$, with $|S(n)| \leq 2^{nR}, R < H(U)$. Then, $\forall \delta > 0, \exists n, \sum_{u \in S(n)} \mathbb{P}(u) < \delta$*

Basically, when $R < H(U)$ we will always get some typical sets that are indistinguishable after compression. Assume $S(n)$ is a set of such indistinguishables. This lemma states that these become less significant?

Proof. Again, we split $S(n)$ into its typical and atypical part and then observe

$$\begin{aligned}
\mathbb{P}(S(n)) &= \sum_{u \in S(n)} \mathbb{P}(u) \\
&= \sum_{u \in S(n) \cap T_\epsilon^{(n)}} \mathbb{P}(u) + \sum_{u \in S(n) \cap A_\epsilon^{(n)}} \mathbb{P}(u) \\
&\leq |S(n)| 2^{-n(H(n)-\epsilon)} + \mathbb{P}(A_\epsilon^{(n)}) \\
&\leq 2^{-n(H(n)-R)} + \mathbb{A}_\epsilon^{(\kappa)}
\end{aligned}$$

□

which completes the proof of Shannon's source coding theorem. Finally, we note that using a variable length encoding scheme or something like that does not fundamentally change this result. [End of lecture 2]

Today we discuss the types of relationships one can have between various types of entropy. As such,

Definition 8. the **joint entropy of** X, Y is

$$H(X,Y) = -\sum_{x \in J_X} \sum_{y \in J_Y} p(x,y) \log(p(x,y)) \tag{5}$$

Definition 9. the **conditional entropy** is

$$H(Y|X) = -\sum_{x \in J_X} \sum_{y \in J_Y} p(x,y) \log(p(y|x)) = -\sum_{x \in J_X} p(x) H(Y|X=x) \tag{6}$$

where we note the "chain rule"

$$H(X,Y) = H(Y|X) + H(X) \tag{7}$$

which generalises nicely as

$$H(X,Y,Z) = H(X) + H(Y|X) + H(Z|Y,X) \tag{8}$$

and similar for higher numbers of variables.

Definition 10. We further define p to be **absolutely continuous** wrt q iff $q(x) = 0 \implies p(x) = 0$ or equivalently $\text{supp}(p) \subseteq \text{supp}(q)$ which we can denote as $p << q$.

and

Definition 11. the **relative entropy** or **Kullback-Leibler divergence** of $p << q$ to be

$$D(p||q) = \sum_{x \in J} p(x) \log(p(x)/q(x)) \tag{9}$$

Note that if $q(x) = 1 \forall x \in J$ then $D(p||q) = -H(X)$ so this is stronger than the Shannon entropy.

This measures how different two distributions are in a certain sense, however, it certainly is not a metric (not symmetric, and no triangle inequality).

Definition 12. We similarly define the **mutual information** between X, Y to be

$$I(X:Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) \tag{10}$$

Definition 13. and the **conditional mutual information** (CMI)

$$I(X:Y|Z) = H(X|Z) - H(X|Y,Z) \tag{11}$$

Intuition wise, one can think of $H(X,Y)$ as adding entropy, $I(X:Y)$ as taking the intersection between the two, $D(p||q)$ as measuring the difference between one distribution contained in another, and $H(Y|X)$ as removing the overlap of one distribution into another ($H(Y|X) = H(Y) - I(Y:X)$). This also explains the commutativeness of these various operations. I'm not sure about the CMI yet...

From here we can state

Theorem 4. *the data processing inequality for Markov Chain $X \to Y \to Z$:*

$$I(X:Y) \geq I(X:Z). \tag{12}$$

If we imagine X as a perfect source, Y as an observed, noisy signal, and Z as a "cleaned up" version of Y after some data processing, then this states that no matter the data processing used to clean up Y, Z can never contain more information about X than Y.

We finally have the following theorem.

Theorem 5. *1.* $p << q \implies D(p||q) \geq 0$ *with equality iff* $p = q$

2. $H(x) \geq 0$ *with equality iff* X *deterministic*

3. $H(X|Y) \geq 0$ *or equivalently* $H(X,Y) \geq H(Y)$

4. $H(X) \leq \log(|J|)$

5. $H(X,Y) \leq H(X) + H(Y)$ *or equivalently* $H(Y) \geq H(Y|X)$ *with equality iff* $X \perp Y$

6. H *is concave meaning* $H(\lambda p_x + (1-\lambda)p_y) \geq \lambda H(p_x) + (1-\lambda)H(p_Y)$

7. $I(X:Y) \geq 0$ *and equal iff* $X \perp Y$

All of these are exercises on the example sheet. [End of lecture 4]

1.3 Classical Information Transmission

Let's build the infrastructure for a bound on reliable information transmission.

Definition 14. A **discrete channel** is a combination of

- discrete alphabets J_X, J_Y
- a set of conditional probabilities $\mathbb{P}(y_1, \ldots, y_n | x_1, \ldots x_m)$

Definition 15. A memoryless channel is one satisfying

$$\mathbb{P}(y_1, \ldots, y_n | x_1, \ldots, x_n) = \prod_{i=1}^{n} \mathbb{P}(y_i | x_i) \tag{13}$$

These are completely characterised by the so-called **channel matrix** $P = \mathbb{P}(y|x)$.

Example 2. *For example we can consider a* ***binary symmetric channel*** *for* $\mathbb{P}(0|0) = \mathbb{P}(1|1) = 1-p$ *where* $p_{err} = 3p^2(1-p) + p^3 = 3p^2 - 2p^3$ *if we use the redundant codingg* $0 \mapsto 000, 1 \mapsto 111$.

Definition 16. Now for messages $m = x_1, \ldots, x_n$ and $m' = y_1, \ldots, y_m$ and $m \in [M] = \{1, \ldots, M\}$ we **encoding** $E_n : [M] \to J_X^n$, **decoder** $D_n : [M] \to J_X^n$, and rate of encoding R such that $M = \lfloor 2^{nR} \rfloor$. The triple $C_n = (E_n, D_n, R)$ is then called an **error correcting code**.

Definition 17. and unlike the average error probabilities consider in information storage, here the error probability is the maximum

$$P_{err}(C_n) = \max_{m \in [M]} \mathbb{P}(D_n(Y^{(n)}) \neq m | X^{(n)} = E_n(m)) \tag{14}$$

Definition 18. and rate R is called **achievable** if there exists an error correcting code such that

$$\lim_{n \to \infty} P_{err}(C_n) = 0 \tag{15}$$

and so quite naturally

Definition 19. the **capacity** of a discrete memoryless channel is $C(N) = \sup\{R : R \text{ is achievable}\}$.

for which

Theorem 6. ***Shannon's noisy channel coding theorem*** *states that*

$$C(N) = \max_{p(x)} I(X : Y) \tag{16}$$

(note that the maximum is only taken over the input, not the output distribution)

Theorem 7. *which has the properties*

- $C(N) \geq 0$
- $C(X) \leq \log |J_X|, \log |J_Y|$

rigorous proof of which is found in the 1991 book *Elements of Information Theory.*

Example 3. *Finally, in the binary symmetric channel from before we find that*

$$I(X : Y) = H(Y) - H(Y|X) = H(Y) - h(p) \implies C(N) = 1 - h(p) \tag{17}$$

[End of lecture 4]

2 Quantum Information Theory

2.1 Quantum States

The lecturer reviews quantum states and how they are represented using linear algebra. Here we note that the **Hemming space** is the space $\mathbb{C}^{2^n}$, and use $B(H)$ to represent the space of bounded linear operators on H. We also mention the **Pauli Matrices** and their properties

- $\sigma_\alpha^2 = I$
- $\sigma_\alpha \sigma_\beta = i\epsilon_{\alpha\beta\gamma}\sigma_\gamma$
- $\{\sigma_\alpha, \sigma_\beta\} = 0$ for $\alpha \neq \beta$

2.2 Open Quantum Systems

We generalise our formalism a bit to account for external noise. As such, the lecturer reviews the postulates of quantum mechanics. Then we define

Definition 20. a **density matrix** to be a matrix representing a combination states

$$\rho = \sum p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{18}$$

which can be equivalently be defined by the properties that

- $\rho \geq 0$ (positive semi-definite) which implies ρ Hermitian
- $\operatorname{tr} \rho = 1$

which serves as a more abstract definition density matrices. If we let $\mathcal{D}(H)$ represent the **set of density matrices** over space H then we notice the property that $\sigma_i \in \mathcal{D}(H), \sum p_i = 1 \implies \sum p_i \sigma_i \in \mathcal{D}(H)$. [End of lecture 5]

Introducing some vocabulary, we say

Definition 21. a state ρ is a **pure state** if $\rho = |\psi\rangle \langle psi|$ or equivalently that $\rho^2 = \rho$ or $\operatorname{tr} \rho^2 = 1$. If not, we call ρ a **mixed state**

Definition 22. and we can define the **purity** of ρ to be $\operatorname{tr} \rho^2$ which is 1 for a pure state and strictly less than 1 for a mixed state. It is minimised for I/d.

Definition 23. For a product space $H_A \otimes H_B$ we find it useful to define the **partial trace** tr_B that sends an operator

$$X_{AB} \mapsto X_A = \operatorname{tr}_B X_{AB} = \sum \left\langle e_i^B \right| X_{AB} \left| e_i^B \right\rangle \tag{19}$$

for an orthonormal basis e_i^B.

We can then define $\operatorname{tr}_{AB} = \operatorname{tr}_A \circ \operatorname{tr}_B$. For observable $M_{AB} = M_A \otimes I_B$ we then find that the expectation value

$$\langle M_{AB} \rangle = \operatorname{tr}(M_{AB} \rho_{AB}) = \operatorname{tr}(M_A \rho_A) \tag{20}$$

where

$$\rho_{AB} = \sum i, j, \alpha, \beta a_{i\alpha, j\beta} |i\rangle \langle j|_A \otimes |\alpha\rangle \langle \beta|_B \implies \rho_A = \sum_{i,j,\alpha} a_{i\alpha, j\alpha} |i\rangle \langle j| \tag{21}$$

We then see that $\operatorname{tr}_B(\rho_A \otimes \rho_B) = \rho_A$ for states ρ_A, ρ_B (called a **bipartite system**), while for maximally entangled states such as the **Bell states** or **EPR pairs**

$$\left|\phi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle) \tag{22}$$

$$\left|\chi^{\pm}\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle) \tag{23}$$

have $\operatorname{tr}_B \rho_{AB} = I_A/2$ meaning that the qubits are maximally mixed, even though the global state is known! For both density matrices and state vectors, we define them to be **separable** if they can be written as tensor products $|\psi\rangle = |\phi\rangle_A \otimes |\chi\rangle_B$ or $\rho_{AB} = \sum_i p_i \omega_i^A \otimes \sigma_i^B$ and if not we call them **entangled**. [End of lecture 6]

2.3 Schmidt Decomposition

The lecturer gives the following special case of the Schmidt decomposition theorem:

Theorem 8. *The **Schmidt decomposition theorem** states that a pure state $|\Psi_{AB}\rangle$ with $d_A = dim(\mathcal{H}_A), d_B = dim(\mathcal{H}_B)$ where the spaces $\mathcal{H}_A, \mathcal{A}_B$ has Schmidt bases $|i_A\rangle, |i_B\rangle$ and nonnegative real **Schmidt coefficients** λ_i such that*

$$\sum_{i=1}^{\min(d_A,d_B)} \lambda_B^2 = 1 \tag{24}$$

and note here that $\rho_A = \sum_i \lambda_i^2 |i_A\rangle \langle i_A|$ and the same for ρ_B.

Proof. The lecturer provides a proof using the Singular Value Decomposition (SVD), and uses the fact that $\operatorname{tr} \rho_A = 1 = \operatorname{tr} \rho_B$ to get $\sum \lambda_i^2 = 1$. □

From here we note that if ρ_A, ρ_B have no degenereate eigenvalues other than 0, the Ψ_{AB} is determined uniquely by ρ_A, ρ_B since one can

1. diagonalise ρ_A, ρ_B
2. and then match up eigenvectors corresponding to the same eigenvalues

We also define

Definition 24. the texbfSchmidt rank of pure bipartite state $|\Psi_{AB}\rangle$ to be the non-zero coefficients in its Schmidt decomposition. This is denoted by $n(\Psi_{AB})$.

This provides a simple way of checking whether or not $|\Psi_{AB}\rangle$ is entangled since it is entangled iff $n(\Psi_{AB}) > 1$.

2.4 Purification

A useful mathematical trick is to represent mixed states as pure states. In particular, for density ρ_A on system A we can introduce a **reference system** R and pure state Ψ_{AR} such that

$$\rho_A = \operatorname{tr}_R(|\Psi_{AR}\rangle \langle \Psi_{AR}|) \tag{25}$$

In particular, we can do so for $\mathcal{H}_R \simeq \mathcal{H}_A$ by considering for

$$\rho_A = \sum_i p_i |i_A\rangle \langle i_A| \tag{26}$$

the pure state

$$|\Psi_{AB}\rangle = \sum_i \sqrt{p_i} \langle i_A| \otimes \langle i_R| \tag{27}$$

(which is just the Schmidt decomposition of Ψ_{AR}. More generally, we can use the **canonical purification**

$$|\Psi_{AR}\rangle = \sqrt{d}(\sqrt{\rho_A} \otimes I_R) |\Omega\rangle \tag{28}$$

where

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle |i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R \tag{29}$$

Also, we not that any pure state $|\Psi_{AR}\rangle$ with reduced state ρ_A can be written as

$$|\Psi_{AR}\rangle = \sqrt{d}(\sqrt{\rho_A} \otimes V) |\Omega\rangle \tag{30}$$

where V is an isometry rotating $|i\rangle$ into the Schmidt basis of $|\Psi_{AR}\rangle$. [End of lecture 7]

2.5 The no cloning theorem

Although it shows up in part II, we review the no cloning theorem.

Theorem 9. *The **no cloning theorem** states that there is no quantum process that universally clones quantum states. More precisely, one can only design a quantum process that clones orthogonal states.*

Proof. Suppose that for two states $|\psi\rangle, |\phi\rangle$ we have that for any state $|s\rangle$

$$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle$$
$$U(|\phi\rangle \otimes |s\rangle) = |\phi\rangle \otimes |\phi\rangle$$

meaning that by taking the inner product we have $\langle\phi|\psi\rangle = \langle\phi|\psi\rangle^2$ meaning that ϕ, ψ are either identical or orthogonal. Now, one might argue that this assumes that we are in a pure state, and also assumes that any operation we use is unitary. However, by considering the purification of a non-pure state this generalises to non-pure states. Also, by adding an ancilla we can make (any?) state unitary so we this generalises to copying any quantum state. □

2.6 Time evaluation of open systems

2.7 Quantum operations

To describe changes in open quantum systems, we use **quantum operations** which we define to be **linear completely positive trace preserving** maps or CPTP maps. These describe discrete changes in a system, and have the advantage that they do no explicitly introduce time. Specifically, we consider these maps to be maps of densities

$$\Lambda : \mathcal{D}(\mathcal{H}) \to \mathcal{D}\mathcal{H}; \rho \mapsto \rho' \tag{31}$$

satisfying the properties that Λ is

- linear (allowing us to interpret mixed states probabilistically)
- trace-preserving (we need to preserve total probability)

- positive (meaning that $\rho \geq 0 \implies \lambda(\rho) \geq 0$. Here $\rho \geq 0$ means ρ is positive semi-definite.
- **complete positivity** which means that $\forall$ ancillas B the map $\Lambda \otimes \mathrm{id}_B$ is positive too. This is a stronger statement than positivity, since in particular we note that $\rho \mapsto \rho^T$ is not completely positive.

Theorem 10. *A map $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ for $\mathcal{H} \simeq \mathbb{C}^d$ is completely positive if and only if*

$$(\Lambda \otimes id_d) |\Omega\rangle \langle\Omega| \geq 0$$

where $|\Omega\rangle = \frac{1}{\sqrt{d}} \sum |ii\rangle \in \mathbb{C}^d \times \mathbb{C}^d$ is the maximally entngled state of Schmidt rank d.

Proof. Necessity of this condition is clear. To prove sufficiency we observe that $\forall k \geq 1$ we have that $\rho \in \mathcal{D}(\mathcal{H} \otimes \mathbb{C}^d)$ so we that if $\rho = \sum p_i |\phi_i\rangle \langle\phi_i|$ then

$$\forall (\Lambda \otimes \mathrm{id}_k) |\phi_i\rangle \langle\phi_i| \geq 0 \implies (\Lambda \otimes \mathrm{id}_k)\rho \geq 0$$

From the last lecture we know that nay pure state can be written in the form $(I_d \otimes R) |\Omega\rangle$ for $R \in \mathcal{B}(\mathbb{C}^d, \mathbb{C}^k)$ so in particular we see that $(\Lambda \otimes \mathrm{id}_k) |\phi_i\rangle \langle\phi_i| \geq 0$ can be written in the forms

$$(\Lambda \otimes \mathrm{id}_k)(I_d \otimes R_i) |\Omega\rangle \langle\Omega| (I_d \otimes R_i^\dagger) \geq 0$$
$$(I_{d'} \otimes R_i)(\Lambda \otimes \mathrm{id}_d) |\Omega\rangle \langle\Omega| (I_{d'} \otimes R_i^\dagger) \geq 0$$

meaning that since the outside conjugation is invertible, we are positive iff

$$(\Lambda \otimes \mathrm{id}_d) |\Omega\rangle \langle\Omega| \geq 0$$

as required □

Note that here we call $J(\Lambda) = J = (\Lambda \otimes \mathrm{id}_d) |\Omega\rangle \langle\Omega|$ the **Choi matrix** or the **Choi state** of Λ. [End of lecture 8]

We mentioned earlier that any quantum operation can be represented as a unitary operation on a larger space. Formally, this is described by

Theorem 11. ***Stinespring's Dilation Theory*** *which states that for any CPTP operator $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ there exists a Hilbert space $\mathcal{H}'$ and a unitary operator $U \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}')$ such that for any $\rho \in \mathcal{D}(\mathcal{H})$*

$$\Lambda(\rho) = \mathrm{tr}_{\mathcal{H}'}(U(\rho \otimes \varphi)U^\dagger) \tag{32}$$

where φ is fixed in $\mathcal{D}(\mathcal{H}')$ (and can be chosen to be pure).

In essence this states that any quantum operation can be described by

1. Adding an ancilla $\mathcal{H}'$
2. Converting it to a unitary operator on the extended space (U)
3. Removing the ancilla $\mathrm{tr}_{\mathcal{H}'}$

2.8 Krauss representation or operator sum representation

Just a side note, $T : \rho \mapsto \rho^T$ is not CP since $T \otimes \mathrm{id}$ acts as a swap operator which has -1 as an eigenvalue.

Theorem 12. *The **Kraus Representation Theorem** states that $\Lambda : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is CPTP iff*

$$\Lambda(\rho) = \sum_k A_k \rho A_k^\dagger \tag{33}$$

*for some linear **Krauss operators** $A_k \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ satisfying $\sum A_k^\dagger A_k = I$.*

In fact, this can be seen as a restatement/special case of Stinespring's Dilation theorem since wlog we can choose the initial ancilla to be in the state $\phi = |\varphi\rangle\langle\varphi|$ so that if $|e_k\rangle$ is an orthonormal basis of the ancilla then

$$A(\rho) = \sum_k \langle e_k| U(\rho \otimes \varphi)U^\dagger |e_k\rangle = \sum_k A_k \rho A_k^\dagger \tag{34}$$

where $A_k = \langle e_k| U |\phi\rangle$ is an operator on the system state (the ancilla part has already been evaluated). Then we have $\sum_k A_k^\dagger A_k = I$ by the completeness of $|e_k\rangle$ and U is unitary since $|\varphi\rangle$ is normalised. Finally, we see that $\Lambda(\rho)$ is certainly linear, and it is completely positive since

$$(\Lambda \otimes \mathrm{id}_d)|\Omega\rangle\langle\Omega| = \sum_k \langle\psi|(A_k \otimes I)|\Omega\rangle\langle\Omega|(A_k^\dagger \otimes I)|\psi\rangle = \sum_k \langle\varphi_k|\rangle \Omega \langle\Omega|\varphi_k\rangle \geq 0 \tag{35}$$

for $|\varphi_k\rangle = (A_k^\dagger \otimes I)|\psi\rangle$ since $\Omega = |\Omega\rangle\langle\Omega|$ is positive (semi-definite).

Basically, the Kraus representation is a way of achieving what Stinespring's Dilation does without using an ancilla. Note that the Kraus representation is not unique. [End of lecture 9]

2.9 Choi-Jamilkowski Isomorphism

Theorem 13. *The **Choi-Jamilkowski Isomorphism** states that there is an isomorphism between $\Lambda : \mathcal{M}_d \to \mathcal{M}_{d'}$ and $J \in \mathcal{B}(\mathbb{C}^{d'} \otimes \mathbb{C}^d)$ given by*

$$J = (\Lambda \otimes id_d)|\Omega\rangle\langle\Omega| \tag{36}$$

$$\mathrm{tr}(A\Lambda(B)) = d\,\mathrm{tr}\big(J(A \otimes B^T)\big) \tag{37}$$

where the last holds for all $A \in \mathcal{M}_{d'}, B \in \mathcal{M}_d$ and we have the properties

1. *Λ is completely positive iff $J \geq 0$*
2. *Λ is trace-preserving iff $\mathrm{tr}_A J = I/d$ where tr_A is the partial trace on $\mathbb{C}^{d'}$*

To justify this, we first define the **adjoint** of $\Lambda : \mathcal{M}_d \to \mathcal{M}_{d'}$ to be $*^*$ satisfying $\mathrm{tr}(A\Lambda(B)) = \mathrm{tr}(\Lambda^*(A)B)$. Also, we note that complete positivity and trace-preserving follow from our earlier work. To establish that these are truly mutual inverses, let's start by reviewing the following identities (proof left as an exercise) for $A, B \in \mathcal{B}(\mathbb{C}^d)$:

- $(A \otimes I)\,|\Omega\rangle = (I \otimes A^T)\,|\Omega\rangle$
- if $F\,|ij\rangle = |ji\rangle$, then $d\,|\Omega\rangle\langle\Omega|^{T_B} = F$ where T_B is the partial transpose defined such that for $C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ has $\langle ij|\,C^{T_A}\,|kl\rangle = \langle kj|\,C\,|il\rangle$
- $\operatorname{tr}((A \otimes B)F) = \operatorname{tr}(AB)$
- $(A \otimes I)F = F(I \otimes A)$

As such, we argue that $\operatorname{tr}(A\Lambda(B)) = d\operatorname{tr}\big(J(A \otimes B^T)\big)$ since

$$\begin{aligned}
d\operatorname{tr}\big(J(A \otimes B^T)\big) &= d\operatorname{tr}\big((\Lambda \otimes \mathrm{id}_B)\,|\Omega\rangle\langle\Omega|\,(A \otimes B^T)\big) \\
&= d\operatorname{tr}\left((\Lambda \otimes \mathrm{id}_B)\frac{1}{d}F^{T_B}(A \otimes B^T)\right) \\
&= \operatorname{tr}\big(F^{T_B}(\Lambda^* \otimes \mathrm{id}_B)(A \otimes B^T)\big) \\
&= \operatorname{tr}\big(d\,|\Omega\rangle\langle\Omega|\,(\Lambda^*(A) \otimes B^T)\big) \\
&= d\operatorname{tr}(B\Lambda^*(A)) \\
&= d\operatorname{tr}(A\Lambda(B))
\end{aligned}$$

This shows we are an inverse in one direction, giving us injectivity in a certain direction. As such it suffices to show that the map $\Lambda \to J$ is surjective, but we can justify this by breaking J into rank one $|\psi\rangle\langle\psi|$ bits and using $|\psi\rangle = (R \otimes I)\,|\Omega\rangle$ from before.

2.10 From Choi-Jamilkowski to Kraus Representation to Stinespring's Dilation

We note that it is possible to use the Choi-Jamilkowski Isomorphism to prove the Kraus representation to prove Stinespring's Dilation theorem. We begin by proving the Kraus representation from Choi-Jamilkowski

Proof.

Note that $J(\Lambda) = \sum p_i\,|\psi_i\rangle\langle\psi_i| \geq 0$ for $|\psi_i\rangle \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$, but we also know that we can write $|\psi_i\rangle = (R_i \otimes I)\,|\Omega\rangle$ for $R_i \in \mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ so $J(\Lambda) = \sum_i (A_i \otimes I)\,|\Omega\rangle\langle\Omega|\,(A_i^\dagger \otimes I)$ where $A_i = \sqrt{p_i}R_i$, and since $J \leftrightarrow \Lambda$ is an isomorphism so we can describe Λ entirely by its action on $\rho \in \mathcal{D}(\mathcal{H}_A)$ as $\Lambda(\rho) = \sum A_i \rho A_i^\dagger$ as required. Furthermore, Λ is trace preserving since $\forall \rho \in \mathcal{B}(\mathcal{H}_A), \operatorname{tr}\Lambda(\rho) = \operatorname{tr}\Big((\sum_i A_i^\dagger A_i)\rho\Big) = \operatorname{tr}(\rho)$ meaning that since this holds for any ρ, we have $\sum A_i^\dagger A_i = I$. □

For the proof of Stinespring's Dilation theorem from the Kraus representation we see that

Proof. For a Hilbert space $\mathcal{H}_E$ with dimension $r \geq \dim(J(\Lambda))$ (why?) we can define U by $|\psi_A\rangle \otimes |\varphi\rangle$ forall $|\psi_A\rangle \in \mathcal{H}_A$ such that if $\varphi = |\varphi\rangle\langle\varphi|$ then

$U\left|\psi_A\right\rangle \otimes \left|\varphi\right\rangle = \sum_i A_i \left|\psi_A\right\rangle \otimes \left|i\right\rangle$ for an orthonormal basis $\left|i\right\rangle$ and Kraus operators A_i. Then U is an isomerty, and we see that for $\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$ then we have that $\mathrm{tr}_{\mathcal{H}_E}(U(\rho \otimes \phi)U^\dagger) = \sum_i A_i \rho A_i^\dagger$ which is the same as $\Lambda(\rho)$ by Kraus representation. □

[End of lecture 11]

2.11 Generalised Measurements

Traditionally we characterise measurements entirely as projection operators, but we find we need a more powerful notion of measurement in certain situations.

Example 4. *Consider the state $\left|\psi\right\rangle$ satisfying $\sigma \cdot \hat{n} \left|\psi\right\rangle = \left|\psi\right\rangle$. If we are given ψ and want to obtain $\hat{n}$ using only one copy of ψ then this is impossible using only projective measurements.*

As such we develop the

Definition 25. generalised measurement postulate which states that a quantum measurement is a collection of **measurement operators** on the Hilbert space such that

$$\sum H_a^\dagger H_a = I \qquad\qquad \mathbb{P}(a) = \left\langle\psi\right| H_a^\dagger H_a \left|\psi\right\rangle \tag{38}$$

and the final measured state collapses to

$$\frac{H_a \left|\psi\right\rangle}{\sqrt{\left\langle\psi\right| H_a^\dagger H_a \left|\psi\right\rangle}} \tag{39}$$

Note that the projective measurement postulate is the special case when

$$P_a^\dagger = P_a \qquad\qquad P_a P_b^\dagger = \delta_{ab} P_a \tag{40}$$

This is equivalent to requiring that the H_a are orthogonal projections.

2.11.1 Postive Operator Value Measurements (POVMs)

We notice that $E_a = H_a^\dagger H_a$ satisfies

- $E_a^\dagger = E_a$
- $E_a \geq 0$
- $\sum_a E_a = I$
- $\mathbb{P}(a) = \left\langle\psi\right| E_a \left|\psi\right\rangle = \mathrm{tr}(E_a \rho)$ for $\rho = \left|\psi\right\rangle \left\langle\psi\right|$

As such, we notice that if we do not care about the final state that the measurement leaves are system in, then we can define a

Definition 26. positive operator value measurement (POVM) to be a partition of I into a finite number of positive semi-definite operators satisfying the above four properties. We call a POVM **pure** if all of its operators are of rank 1 meaning that there exists states $|\varphi_a\rangle$ such that $E_a = |\varphi_a\rangle \langle\varphi_a|$.

Theorem 14. *Then **Neumann's theorem** states that any pure POVM acting on an N dimensional Hilbert space $\mathcal{H}$ consisting of $n \geq N$ operators E_i can be realised by extending $\mathcal{H}$ to dimension n and performing a projective measurement.*

There are some exercises concerning these on the example sheet.

Finally, we check what it means to be a projective operator in the context of POVMs. Here we notice that

$$E_M = P_m^\dagger P_m = P_m^2 = P_m \tag{41}$$

so we see that in the case of projective measurements, the projections are the same as the POVMs.

2.11.2 Implementation of a generalised measurement

To implement a genrealised measurement, we implement the idea stated in Neumann's theorem by using unitary operators and projective measurements. Our procedure is as follows

1. start with a system $A, \mathcal{H}_A$ and POVMs $\{H_a\}$
2. add an auxilliary space $B, \mathcal{H}_B$ with an orthonormal basis corresponding to outcomes of H_a, labelled as $|e_a\rangle$.
3. Implement operator U such that for a state $|\varphi\rangle$

$$U |\psi\rangle |\varphi\rangle = \sum_a H_a |\phi\rangle |e_a\rangle \tag{42}$$

meaning that if we define

$$|\Psi\rangle = U |\psi\rangle |\varphi\rangle \tag{43}$$
$$|\Phi\rangle = U |\phi\rangle |\varphi\rangle \tag{44}$$

we find that

$$\langle\Psi|\rangle \Phi = \langle\psi|\phi\rangle \tag{45}$$

allowing us to use the inner product of the smaller space in the larger space. [End of lecture 11]

4. Now define $P_a = I_A \otimes |e_a\rangle |e_a\rangle$ so that we see that

$$\mathbb{P}(a) = \langle\Psi| P_a |\Psi\rangle = \langle\phi| H_a^\dagger H_a |\psi\rangle \tag{46}$$

and

$$\frac{P_a |\Psi\rangle}{\sqrt{\langle\Psi| P_a |\Psi\rangle}} = \frac{H_a |\psi\rangle |\varphi\rangle}{\sqrt{\langle\psi| H_a^\dagger H_a |\psi\rangle}} \tag{47}$$

Furthermore, we notice that for the set $\{|\phi_i\rangle\}$ representing the decomposition of maximally mixed state ρ deines a pure POVM. Moreover, for a POVM $\{E_i\}$,

$$\rho_i = \frac{E_i}{\operatorname{tr} E_i}; \qquad p_i = \frac{1}{N}\operatorname{tr}(E_i) \tag{48}$$

defines a maximally mixed state (what is a maximally mixed state?). For the converse, we can use $E_i = Np_i\rho_i$ which satisfies $\sum E_i = I$.

2.12 Distances between states

2.12.1 Trace Distance

We define the

Definition 27. trace distance between two states to be

$$D(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1 \tag{49}$$

where $\|A\|_1 = \operatorname{tr}|A|$ for $|A| = \sqrt{A^TA}$.

Here we notice that if $A = \sum a_o |\varphi_i\rangle\langle\varphi_i| = Q - R$ is the eigenvalue decomposition of A where Q contains the nonnegative eigenvalues and R contains the negative eigenvalues, meaning $Q, R \geq 0$ (R can be 0 if $A \geq 0$). We see that $Q \perp R$ meaning that

$$D(\rho,\sigma) = \frac{1}{2}(\operatorname{tr} Q + \operatorname{tr} R) \tag{50}$$

Lemma 2.

$$D(\rho,\sigma) = \max_{0\leq P\leq 1} \operatorname{tr} P(\rho-\sigma) \tag{51}$$

Proof. For quantum states ρ,σ we see that $\operatorname{tr} Q = \operatorname{tr} R$, meaning that $D(\rho,\sigma) = \operatorname{tr} Q = \operatorname{tr} R$. Then if $0 \leq P \leq 1$ projects onto the support of Q we get that $\operatorname{tr} P(\rho-\sigma) = \operatorname{tr} P(Q-R) = \operatorname{tr} PQ = \operatorname{tr} Q = D(\rho,\sigma)$. Conversely, for any $0 \leq P \leq 1$, we have $\operatorname{tr} P(\rho-\sigma) = \operatorname{tr} P(Q-R) \leq \operatorname{tr} PQ \leq \operatorname{tr} Q = D(\rho,\sigma)$. □

We remark that the trace distance is a metric in the topological sense, and that

Lemma 3. *we have* ***monotonicity over quantum operators*** *meaning that for any CPTP* Λ

$$D(\Lambda\rho,\Lambda\sigma) \leq D(\rho,\sigma) \tag{52}$$

Proof. Pick P such that $D(\Lambda\rho,\Lambda\sigma) = \operatorname{tr}(P(\Lambda\rho-\Lambda\sigma))$ then Λ trace preserving implies $\operatorname{tr}\Lambda Q = \operatorname{tr} Q$ meaning that

$$D(\rho,\sigma) = \operatorname{tr} Q = \operatorname{tr}\Lambda Q \geq \operatorname{tr} P\Lambda Q \geq \operatorname{tr} P(\Lambda Q - \Lambda R) = D(\Lambda\rho,\Lambda\sigma)$$

□

Example 5. *As an example where the trace distance has operational significance, we consider* ***(Binary) Quantum Hypothesis Testing****. Here we have A prepare one of the states ρ_0, ρ_1 with equal probability, which is then sent to B through a noiseless quantum channel. B wishes to determine which of the two states was prepared. This can be done using POVM $E_0, E_1 = I - E_0$. Then the probability of error is given by*

$$p_e(E_0, E_1) = \frac{1}{2}(\operatorname{tr} E_0\rho_1 + \operatorname{tr} E_1\rho_0) = \frac{1}{2}(I - \operatorname{tr} E_0(\rho_0 - \rho_1)) \tag{53}$$

and so we have minimum probability of error given by

$$p_e^* = \frac{1}{2}(1 - D(\rho_0, \rho_1) \tag{54}$$

and similarly the maximum probability of success is

$$p_{success}^* = \frac{1}{2}(1 + D(\rho_0, \rho_1). \tag{55}$$

We finally note that when the maximum success probability is obtained we can use the projective measurements given by the nonnegative negative parts $P_{\geq}, P_{<}$ of the operator $P = \rho_0 - \rho_1 = \sum_i a_i |\varphi_i\rangle \langle\varphi_i|$. [End of lecture 12]

2.12.2 Fidelity

Definition 28. The **fidelity** of ρ, σ is

$$F(\rho, \sigma) = \operatorname{tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \|\sigma^{1/2}\rho^{1/2}\|_1. \tag{56}$$

If then $\rho = \sum \lambda_i |e_i\rangle \langle e_i|, \sigma = \sum M_i |e_i\rangle \langle e_i|$ (so they share the same eigenvectors) then

$$F(\rho, \sigma) = \sum \sqrt{\lambda_i M_i} = F_{cl}(\lambda, M) \tag{57}$$

for the **classical fidelity** $F_{cl}(\lambda, M) = \sum \sqrt{\lambda_i M_i}$. If instead one of the states is rank 1, we find

$$F(|\varphi\rangle \langle\varphi|, \sigma) = \sqrt{\langle\varphi| \sigma |\varphi\rangle} \tag{58}$$

meaning that

$$F(|\psi\rangle \langle\psi|, |\varphi\rangle \langle\varphi|) = |\langle\psi|\varphi\rangle|. \tag{59}$$

As an exercise, the lecturer proposes proving that fidelity is invariant under unitary transformations $\rho \mapsto U\rho U^\dagger$.

Theorem 15. ***Uhlmann's theorem*** *states that*

$$F(\rho, \sigma) = \max_{|\psi_\rho\rangle, |\psi_\sigma\rangle} |\langle\psi_\rho|\psi_\sigma\rangle| \tag{60}$$

where $|\psi_\rho\rangle, |\psi_\sigma\rangle$ are purifications of the states ρ, σ.

but before proving it, we consider the lemma

Lemma 4. $\forall A \in \mathcal{B}(\mathcal{H})$

$$\|A\|_1 = \max_{\text{unitary} U} |\operatorname{tr}(UA)| \tag{61}$$

Proof. Note from the example sheet that for any unitary U, $|\operatorname{tr}(AU)| \leq \operatorname{tr}|A|$ (this can be justified using the polar decomposition together with Cauchi-Schwarz). Using the polar decomposition we can also show that this maximum is attained for some U, so done. □

And now to prove Uhlmann's theorem

Proof. We note (from ES 2) that all purifications are equivaleent up to a unitary transformations on the reference system, so

$$\begin{aligned} F(\rho,\sigma) &= \max_{U_R^\rho, U_R^\sigma} |\, |\psi_\rho\rangle\, (U_R^\rho \otimes I_A)(U_R \otimes I_A)\, |\psi_\sigma\rangle \\ &= \max_U |\, \langle\psi_\rho|\, (U \otimes I_A)\, |\psi_\sigma\rangle\, | \end{aligned}$$

which if we pick $|\psi_\rho\rangle, |\psi_\sigma\rangle$ to be the canonical purifications

$$|\psi_\rho\rangle = \sqrt{d}(I_R \otimes \sqrt{\rho_A})\, |\Omega\rangle \tag{62}$$

$$|\psi_\sigma\rangle = \sqrt{d}(I_R \otimes \sqrt{\sigma_A})\, |\Omega\rangle \tag{63}$$

then we find $|\, \langle\psi_\rho|\, U \otimes I_A\, |\psi_\sigma\rangle\, | = |\operatorname{tr}\sqrt{\rho_A}\sqrt{\sigma_A}U^\dagger|$. □

Uhlmann's theorem has the following handy consequences:

- $0 \leq F(\rho,\sigma) \leq 1$ and $F(\rho,\sigma) = 1$ iff $\rho = \sigma$
- $F(\rho,\sigma) = F(\sigma,\rho)$
- The

  Lemma 5. ***monotonicity under partial trace*** *of fidelity, meaning that for bipartite states ρ_{AB}, σ_{AB} with reduces states ρ_A, σ_A we have*

  $$F(\rho_{AB}, \sigma_{AB}) \leq F(\rho_A, \sigma_A) \tag{64}$$

  Proof. By Uhlmann, we find purifications $\psi_\rho^{ABC}, \psi_\sigma^{ABC}$ such that $F(\rho_{AB}, \sigma_{AB}) = |\, \langle \psi_\rho^{ABC} | \psi_\sigma^{ABC} \rangle\, |$. We notice that these are also purifications of ρ_A, σ_A meaning that by Uhlmann,

  $$F(\rho_A, \sigma_A) = \max_{|\psi_\rho\rangle, |\psi_\sigma\rangle} |\, \langle\psi_\rho|\psi_\sigma\rangle\, | \geq |\, \langle \psi_\rho^{ABC} | \psi_\sigma^{ABC} \rangle\, | = F(\rho_{AB}, \sigma_{AB}) \tag{65}$$

  □

2.12.3 Entanglement Fidelity

We would like to know how much entanglement is preserved by a quantum operation in some sense. As such we define the

Definition 29. entanglement fidelity to be

$$F_e(\rho, \Lambda) = \langle\psi_{RA}| \left((\mathrm{id}_R \otimes \Lambda) |\psi_{RA}\rangle \langle\psi_{RA}|\right) |\psi_{RA}\rangle \tag{66}$$

so if $\psi_{RA} = |\psi_{RA}\rangle \langle\psi_{RA}|$ then

$$F_e(\rho, \Lambda) = (F(\psi_{RA}, (\mathrm{id}_R \otimes \Lambda)\psi_{RA}))^2 \tag{67}$$

Lemma 6. *For a CPTP $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ with Krauss representation $\Lambda(\rho) = \sum A_k \rho A_k^\dagger$ we have*

$$F_e(\rho, \Lambda) = \sum |\operatorname{tr}(A_k \rho)|^2 \tag{68}$$

Proof.

$$\begin{aligned} F_e(\rho, \Lambda) &= \sum_k \langle\psi_{RA}| (I \otimes A_k)\psi_{RA}(I \otimes A_R^\dagger) |\psi_{RA}\rangle \\ &= \sum_k |\langle\psi_{RA}| (I \otimes A_k) |\psi_{RA}\rangle|^2 \end{aligned}$$

so considering the Schmidt decomposition $|\psi_{RA}\rangle = \sum \sqrt{\lambda_i} |i_R\rangle |i_A\rangle$, so if $\rho = \operatorname{tr}_R |\psi_{RA}\rangle \langle\psi_{RA}| = \sum \lambda_i |i_A\rangle \langle i_A|$ meaning that

$$F_e(\rho, \Lambda) = \sum_k \left| \sum_i \langle i_A| A_k \lambda_i |i_A\rangle \langle i_A| \, |i_A\rangle \right|^2 = \sum_k |\operatorname{tr}(A_k \rho)|^2$$

□

Lemma 7. *For CPTP $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ we have $\forall \rho \in \mathcal{D}(\mathcal{H})$ we have $F_e(\rho, \Lambda) = (F(\rho, \Lambda(\rho)))^2$.*

This is clear from an earlier remark. [End of lecture 13]

2.13 Quantum Entropy

Having built up the framework of quantum information, we introduce concepts from classical information theory for quantum information. We begin with the

Definition 30. quantum entropy or **von Neumann entropy** (which incidentally was defined before the Shannon entropy even though its significance was only realised much later) to be

$$S(\rho) = -\operatorname{tr} \rho \log \rho \tag{69}$$

When $\rho = \sum \lambda_i |\psi_i\rangle \langle\psi_i|$ this reduces to the Shannon entropy on the distribution given by the $\{\lambda_i\}$. It satisfies the properties

- $S(\rho) \geq 0$ with equality iff ρ is a pure state.
- $S(\rho) = S(U\rho U^\dagger)$ for any unitary U, since an operation of this kind corresponds only to a relabelling of states.
- $S(\rho) \leq \log d$ for $\dim(\mathcal{H}) = d$ with equality iff ρ is the completely mixed state I/d.
- S is concave meaning that for probabilities p_i and densities ρ_i we have

$$S(\sum p_i \rho_i) \geq \sum p_i S(\rho_i) \tag{70}$$

  (proof on example sheet 3)

We similarly define

Definition 31. the **quantum relative entropy** (QER) or **quantum Kullback-Leibler divergence** to be

$$D(\rho||\sigma) = \operatorname{tr} \rho(\log \rho - \log \sigma) \tag{71}$$

for $\text{supp}\rho \subseteq \text{supp}\sigma$.

which has the properties

- **Klein's inequality** which states the QRE is nonnegative so $D(\rho||\sigma) \geq 0$ with equality iff $\rho = \sigma$.

  Proof. $\rho = \sum \lambda_i |i\rangle \langle i|, \sigma = \sum q_\alpha |\alpha\rangle \langle \alpha|$ so $D(\rho||\sigma) = \sum \lambda_i \log \lambda_i - \sum_{i\alpha} \lambda_i |\langle i|\alpha\rangle|^2 \log q_\alpha$. Thus if $p_{i\alpha} = |\langle i|\alpha\rangle|^2$ then $p_{i\alpha} \geq 0$ and $\sum_i p_{i\alpha} = \sum_\alpha p_{i\alpha} = 1$ (forming a "double-stochastic" matrix). Consequently, $D(\rho||\sigma) = \sum_i \lambda_i (\log \lambda_i - \sum_\alpha p_{i\alpha} \log q_\alpha)$, but by the concavity of the logarithm and for $r_i == \sum_\alpha p_{i\alpha} q_\alpha$

$$D(\rho||\sigma) \geq \sum_i \lambda_i (\log \lambda_i - \log r_i) = \sum_i \log(\lambda_i / r_i) \geq 0 \tag{72}$$

  since the classical relative entropy is non-negative. □

  This property allows us to prove property 3) of the von Neumann entropy using $\rho = \rho, \sigma = I/d$ to see that

$$\begin{aligned} D(\rho||I/d) &= \operatorname{tr} \rho(\log \rho - \log I/d \\ &= \operatorname{tr} \rho \log \rho - \log(1/d) \operatorname{tr} \rho \\ &= -S(\rho) + \log d \\ &\geq 0 \end{aligned}$$

-

Lemma 8. *The QRE is* ***monotonous under quantum operations*** *which is also just a quantum version of the data processing inequality*

$$D(\Lambda(\rho)||\Lambda(\sigma)) \leq D(\rho||\sigma) \tag{73}$$

- The QRE is **jointly convex** meaning that for a probability distribution p_i
$$D(\sum p_i\rho_i||\sum p_i\sigma_i) \leq \sum p_i D(\rho_i||\sigma_i) \tag{74}$$
- $D(\rho\otimes\omega||\sigma\otimes\nu) = D(\rho||\sigma) + D(\omega\otimes\nu)$
- $D(U\rho U^\dagger||U\sigma U^\dagger) = D(\rho||\sigma)$ for any unitary U
- (proof as exercise)
$$\frac{1}{d}\sum_{km} W_{km} A W_{km}^\dagger = \operatorname{tr} A\cdot\tau \tag{75}$$
for $\tau = I/d$ and the (all unitary) **Heisenberg-Weyl operators** $W_{km} = X^k Z^m$ for higher dimension generalisations of the σ_X, σ_Z Pauli operators
$$X^k|j\rangle = |j +_d k\rangle \qquad\qquad Z^m|j\rangle = e^{\frac{2\pi i m j}{d}}|j\rangle \tag{76}$$

Then we wrap up with the definitions of

Definition 32. the **quantum join entropy**

$$S(A,B) = \operatorname{tr}\rho_{AB}\log\rho_{AB}, \tag{77}$$

Definition 33. the **quantum conditional entropy**

$$S(A|B) = S(A,B) - S(B) \tag{78}$$

which has no clear analogue if not defined in terms of the join entropy in this way. Here $S(B) = S(\rho_B)$ where $\rho_B = \operatorname{tr}_A\rho_{AB}$.

Definition 34. And finally the **quantum mutual information**

$$I(A:B) = S(A) + S(B) - S(A,B) = S(A) - S(A|B) = S(B) - S(B|A) \tag{79}$$

As in the classical case, all of these can be written in terms of the QRE, and it is instructive to find out how to do so. [End of lecture 14]