Quantum Information Theory

oliverobrien111

July 2021

1 Shannon entropy

Alternative definition is using the surprisal $\gamma(x) = -\log p(x)$ which measures how likely an event is. The Shannon entropy is the expected value of the surprisal. A good way of thinking about it is the information gained on average when you learn the value of X.

For a string chosen from the binary alphabet the number of distinct strings of length n is the binomial coefficient $\binom{n}{np}$ which is approximated by Strilings approximation $\log n! = n \log n - n$ to give

$$\log \binom{n}{np} \approx nH(p)$$

for the entropy function:

$$H(p) = -p \log p - (1-p) \log(1-p) \tag{1}$$

Therefore, the strings can be specified by a block code of $2^{nH(p)}$ letters. As $H(p) \leq 1$, the block code is shorter than the message. This reflects the fact that for large n the chance of an unlikely string like 1111111111... becomes very small, so can be left out of the block code. This generalises to n letters to give Shannon entropy:

$$H(X) = \sum_x -p(x) \log p(x) \tag{2}$$

1.1 Properties of H

$$H(X) = H(\{p(x)\}) = H(p_x)$$

Permutation invariant. If you have a π permutation of X then:

$$H(p_X \cdot \pi) = H(p_X) = -\sum_{x \in J} p(\pi(x)) \log p(\pi(x)) = -\sum_{x \in J} p(x) \log p(x)$$

$H(X) \geq 0$

1.2 Formal Proof

A typical sequence $\boldsymbol{u}$ is defined as one which for which the probability of occurrence satisfies:

$$2^{-n(H(X)+\epsilon} \leq p(\boldsymbol{u}) \leq 2^{-n(H(X)-\epsilon} \tag{3}$$

Typical sequence theorem states that the probability of getting any typical sequence can be made arbitrarily close to 1 for large enough n. Let $T_\epsilon^{(n)}$ be the set of typical sequences, then the probability of getting any typical sequence is bounded by:

$$2^{-n(H(X)+\epsilon}|T_\epsilon^{(n)}| \leq \sum_{\boldsymbol{u}\in T_\epsilon^{(n)}} p(\boldsymbol{u}) \leq 1$$

using the left hand inequality of 3. Therefore, $|T_\epsilon^{(n)}| \to 2^{nH(X)}$ as $\epsilon \to 0$.

Pick ϵ such that $R > H(X) + \epsilon$. Break set of possible sequences into typical set and its complement A_ϵ^n. Encode any value in A_ϵ^n to flag bit 0 and assign a codeword of length nR to elements in $T_\epsilon^{(n)}$ (which is possible as $|T_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon} < 2^{nR}$). Probability of failure is:

$$\sum_{\boldsymbol{u}\in A_\epsilon^n} p(\boldsymbol{u}) \tag{4}$$

which by typical sequence theorem can be made arbitrarily small (as the probability of being in $T_\epsilon^{(n)}$ tends to 1).

1.2.1 Converse

Pick a subset of the typical set S^n with $|S^n| = 2^{nR}$ with $R < H$. The probability of a sequence being in S^n is:

$$P(S^n) = \sum_{\boldsymbol{u}} p(\boldsymbol{u}) = 2^{nR}2^{-nH(X)}$$

($p(\boldsymbol{u}) = 2^{-nH(X)}$ in limit as $n \to \infty$). As $H(X) - R > 0$, this tends to 0 as n tends to infinity.

Intuitively it doesn't work as the number of sequences we missed grows exponentially as n heads to infinity.

Joint Entropy:

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

$$H(X,Y) = H(X) + H(Y)$$

Conditional Entropy (imagine X are the bits received over a noisy channel so this is the entropy of the source Y given the data received):

$$H(Y|X) = \sum_{\boldsymbol{x}} p(x)H(Y|X=x) = \sum_{\boldsymbol{xy}} p(x)p(y|x)\log p(y|x) \tag{6}$$

After every letter is received a new optimal code can be found that will specify the remaining string with $H(Y|X)$ bits per letter. Therefore, **chain rule**

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

Relative Entropy (defines a sort of distance between two probability distributions but is not a metric as not symmetric and does not satisfy triangle inequality):

$$D(p||q) = \sum_{\boldsymbol{x}} p(x) \frac{p(x)}{q(x)} \tag{8}$$

only 0 for $p = q$. The above is only well-defined if $p << q$ (meaning that $q(x) = 0 \Rightarrow p(x) = 0$).

Proof of Gibbs inequality

$$A = \{x \in J, p(x) > 0\}$$

$$D(p||q) = \sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} = -\sum_{x \in A} p(x) \log \frac{q(x)}{p(x)}$$

Define a random variable to take values $\log \frac{p(x)}{q(x)}$ with probablity $p(x)$ then.

$$-D(p||q) = \mathbb{E}_p(\log \frac{q(X)}{p(X)})$$

using jensens inequality

$$-D(p||q) \leq \log \mathbb{E}_p(\frac{q(X)}{p(X)}) = \log \sum_A p(x) \frac{q(X)}{p(X)} = \log \sum_A q(x) \leq \log \sum_J q(x) = 0$$

so

$$D(p||q) \geq 0$$

Mutual information:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{9}$$

If you were to fill out a venn diagram with overlapping circles $H(X)$ and $H(Y)$ the union would be $H(X,Y)$, the intersection would be $I(X,Y)$ and the remainder of the circle $H(X)$ would be $H(X|Y)$ as once you are given one of the values there is no longer any entropy coming from its circle.

Neat result. As $D(p||q) \geq 0$ and given $q(x) = \frac{1}{|J|}$ for alphabet J,

$$D(p||q) = \sum p(x) \log \frac{p(x)}{\frac{1}{|J|}} = -H(X) + \sum p(x) \log |J|$$

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

Jensen's Inequality: (for concave functions)

$$\mathbb{E}(f(X)) \leq f(\mathbb{E}(X)) \tag{10}$$

Need to learn what concavity really means as used lots in quantum part of course. Important to know that $H(X)$ is concave.

Subadditivity Prove this with $D(p||q) \geq 0$ and jensens inequality. Do it on example sheet 1.

$$H(X,Y) \geq H(X) + H(Y) \tag{11}$$

5.28 of chapter 5 of Caltech I don't understand

Can use relative entropy as a parent quantity to get all the types of entropy from above. But only if we lift the restriction of the distributions having total probablity 1. For example if we take $q(x) = 1 \forall x$ then $D(P||Q) = -H(X)$. Also, $I(X:Y) = D(p(x,y)||p(x)p(y))$ and $H(Y|X) = D(p(x,y)||p(x))$. Prove and check these on example sheet.

1.3 Shannon's Noisy Channel Theorem

For certain codewords, their images after applying the channel map will represent disjoint subsets in the asymptotic limit. The typical number of sequences that will be received is $|T_n| \approx 2^{nH(Y|X)}$, whereas the size of the range is $2^{nH(Y)}$, so the maximum achievable rate (number of bits communicated per use of channel) is $\frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(Y;X)}$.

If a message of length n is sent along a channel with an error rate of p. Then roughly np bits will flip, leading to $2^{nH(p)}$ typical output strings (it would be 2^{np} strings but we have to account for the encoding reducing the number of strings). In order for this input to be accurately distinguished from any other this "sphere" of possibilities must be distinct from the rest. Therefore, there must be at least $2^{nR}2^{nH(p)}$ possible output strings. Therefore, $R \leq 1 - H(p) = C$.

Can be shown that even picking random codewords gives the optimal rate in the asymptotic limit. If we adopt the decoding method of drawing a "Hamming sphere" of radius $2^{n(H(p)+\delta}$ around the received string and looking for a codeword within this radius. We would typically expect their to be at least one or our assumption about the error in the channel is wrong/we need a bigger delta. The chance of there being two can be calculated as the fraction of space occupied by the s"sphere" is:

$$\frac{2^{n(H(p)+\delta)}}{2^n} = 2^{-n(C-\delta)}$$

so the chance of one of the $2nR$ codewords lieing there is:

$$2^{-n(C-R-\delta)}$$

As δ can be as small as we like, we can pick R as close as we want to C and this will still vanish asymptotically. APPARENTLY THE AVERAGE IS TAKEN HERE BUT I DON'T SEE WHERE.

1.4 Shannon's Noisy Channel Coding Theorem - lectures

1.4.1 Discrete Memoryless Channel (DMC)

Action of each successive uses of $\mathfrak{N}$ is identical and independent to the previous use/the noise affecting each successive inputs in uncorrelated.

$$p(u^{(n)}|x^{(n)}) = \sum_{i=1}^{n} p(u_i|x_i)$$

Might as well restrict to only considering a single use of the channel. Can write channel matrix as $p_{ij} = p(y_i|x_i)$. The channel matrix is symmetric if the rows are permutations of each other.

1.4.2 Example - Memoryless Binary Symmetric Channel (m.b.s.c)

$$J_x = \{0, 1\} = J_y$$

Flips the bit with probability p. So we need an error-correcting code, e.g. the repetition code using three bits at once.

Rate: The encoding decoding pair is said to have a rate R if $|M|$ (number of possible messages) $= 2^{nR}$ for a given number of channel uses n.

Maximum probability of error corresponding to C_n:

$$p_{err}^{(n)}(C_n) = P(\mathfrak{D}_n(Y^{(n)}) \neq m|X^{(n)} = \epsilon_n(m))$$

Achievable rate: A $R \in \mathbb{R}$ is said to be an achievable rate if there exists a sequence of codes $((C_n)_n)$ of rate R s.t. $p_{err}^{(n)} \to 0$ as $n \to \infty$.
Channel Capacity: Maximum rate of reliable transmission of information $C(\mathfrak{N}) = \sup\{R : \text{R is an achievable rate}\}$

Shannon's Theorem says that $C(\mathfrak{N}) = max_{\{p(x)\}} I(X : Y)$. Not going to do this proof as it is not very connected to the quantum proof.

For m.b.s.c

$I(X : Y) = H(Y) - H(Y|X) = H(Y) - (-(1-p)\log(1-p) - p\log p) = H(Y) - h(p) \leq \log|J_Y| - h(p) = 1 - h(p)$

Does there exist some distribution $\{p(x)\}$ for which $H(Y) = 1$ because if there is then this bound is saturated.

$$H(Y) = -\sum p(y)\log p(y)$$

$$p(y) = \sum_x p(x,y) = \sum_x p(x)p(y|x)$$

So look for $p(x)$ that makes $p(y)$ equiprobable. Try $p(x) = \frac{1}{2}$, and this indeed gives $H(Y) = 1$. Therefore $C(\mathfrak{N}) = 1 - h(p)$. (I am confused about this as the H(Y) only reaches its maximum for $p = 1/2$ when $h(p)$ is also 1 so surely this doesn't work? ask in office hours).

2 Quantum

Can associate many ensembles with the same state. You can use any $\{p_i, |\psi_i\rangle\}$ as long as $p_i \geq 0$, $\sum p_i = 1$ and $\langle\psi_i|\psi_i\rangle = 1$. You can find an infinite number of these ensembles that give the same spectral decomposition (are identical) as the $|\psi_i\rangle$ need not even by orthogonal and can basically be anything as long as they have norm 1. e.g. $\rho = \frac{I}{d} = \sum_j \frac{1}{d} |e_j\rangle\langle e_j| = \sum_k \frac{1}{d} |\psi_k\rangle\langle\psi_k|$ so ensembles $\{\frac{1}{d}, |e_j\rangle\}$ and $\{\frac{1}{d}, |\psi_k\rangle\}$ are both equally valid.

Expectation of observable A in state ρ: $< A >=< A >_\rho = Tr(A\rho)$ which is a positive linear functional.

System of interest to us in the course is often a subsystem S of a composite system SE. The density matrix formalism provides a description of states of subsystems. Consider a comp. system AB then the underlying hilbert space is $\mathcal{H}_A \otimes \mathcal{H}_B$. If AB is in the state $\rho_{AB} \in \mathfrak{D}(\mathcal{H}_A \otimes \mathcal{H}_B) then state A is given by the reduce state : \rho_A = Tr_B \rho_{AB}. Consider orthonomal basis \{ |i_A\rangle\}$ in $\mathcal{H}_A$ and $\{|\alpha_B\rangle\}$ in $\mathcal{H}_B$ then we have $\{|i_A\rangle \otimes |\alpha_B\rangle\}$ in $\mathcal{H}_A \otimes \mathcal{H}_B$. Can always write $A = \sum a_{ij} |i\rangle\langle j|$ with $a_{ij} = \langle i| A |j\rangle$.

$$\rho_{AB} = \sum_{i,j=1}^{d_A} \sum_{\alpha,\beta=1}^{d_B} r_{i\alpha,j\beta} |i_A\rangle |\alpha_B\rangle \langle j_A| \langle\beta_B|$$

$$\rho_A = Tr_B \rho_A B = Tr_B(\sum_{i,j=1}^{d_A} \sum_{\alpha,\beta=1}^{d_B} r_{i\alpha,j\beta} |i_A\rangle\langle j_A| \otimes |\alpha_B\rangle\langle\beta_B|) = \sum_{i,j=1}^{d_A} \sum_{\alpha=1}^{d_B} r_{i\alpha,j\alpha} |i_A\rangle\langle j_A|$$

Consider an observable $M_{AB} = M_A \otimes I_B$ then we claim that:

$$< M_{AB} >_{\rho_{AB}} = Tr(M_{AB}\rho_{AB})$$

Proof:

$$< M_{AB} = Tr(M_{AB} \sum_{i,j=1}^{d_A} \sum_{\alpha,\beta=1}^{d_B} r_{i\alpha,j\beta} |i_A\rangle\langle j_A| \otimes |\alpha_B\rangle\langle\beta_B|) = \sum_{i,j=1}^{d_A} \sum_{\alpha=1}^{d_B} r_{i\alpha,j\alpha} Tr(M_A |i\rangle\langle j| = Tr(M_A\rho_A)$$

Example: Consider state of 2 qubits $\rho_{AB} = |phi^+_{AB}\rangle\langle\phi^+_{AB}|$ with $|phi^+_{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Therefore, $\rho_A = Tr_B\rho_{AB} = \frac{I}{2}$ which is an example of when

we know everything about the system but still have no information about A or B.

A pure state is known as a product state if it can be written: $\psi_{AB} = |\psi_A\rangle \otimes |\psi_B\rangle$ and if it cannot be written like this then it is called entangled.

A mixed state is a seperable state if it can be written as an ensemble with pure states: $\rho_{AB} = \sum_i p_i \omega_A^i \otimes \tau_B^i . Otherwise it is called an entangled state.$

$The four pure states in \mathrm{C}^2 \otimes \mathbb{C}^2$ are called Bell states and EPR states.

2.1 Schmidt decomposition

Any state in space $H_A \otimes H_B$ can be described using coefficients of basis states of the form $|i_A\rangle \otimes |i_B\rangle$ where $|i_A\rangle$ for some set of basis eigenstates $|i_A\rangle$, $|i_B\rangle$. The schmidt rank is the number of positive Schmidt coefficents.

There always exists a set of orthonormal vectors $\{|i_A\rangle\}$ and $\{|i_B\rangle\}$ such that:

$$|\psi_{AB}\rangle = \sum_{i=1}^{\kappa = min(d_a, d_b)} \lambda_i |i_A\rangle |i_B\rangle$$

Proof:

$$\psi_{AB} = \sum_{r,\alpha} a_{r\alpha} |\gamma_A\rangle |\alpha_B\rangle$$

Singular Value Decompisition (SVD): Need to know proof look up in notes There exists unitaries U $d_A \times d_A$ and V $d_B \times d_B$ s.t $A = UDV$ with D a diagonal matrix $d_A \times d_B$. Therefore,

$$\alpha_{r,\alpha} = \sum_{i=1}^{d_A} \sum_{\beta=1}^{d_B} u_{\gamma i} d_{i\beta} u_{\beta\alpha}$$

$$|\psi_{AB}\rangle = \sum \sum d_{i\beta} (\sum u_{ri} |r_A\rangle)(\sum v_{\beta\alpha} |\alpha_B\rangle)$$

as $d_{i\beta} =_{i\beta} d_{ii}$

$$|\psi_{AB}\rangle = \sum_i d_{ii} (|i_A\rangle)(|i_B\rangle) = \sum_i \lambda_i (|i_A\rangle)(|i_B\rangle)$$

Now need to check these $\{|i_A\rangle\}$, $\{|i_B\rangle\}$ are an orthonormal basis. So need to check they are orthonormal and complete ($\sum_i^{min(d_A, d_B)} |i_A\rangle \langle i_B| = 1$). To check completeness consider:

$$\rho_{AB} = |\psi_{AB}\rangle \langle \psi_{AB}|$$

$$\rho_A == Tr_B |\psi_{AB}\rangle \langle \psi_{AB}| = Tr_B (\sum_i \lambda_i |i_A\rangle |i_B\rangle)(\sum_j \lambda_j |j_A\rangle |j_B\rangle) = \sum_{ij} \lambda_{ij} |i_A\rangle \langle j_A| \otimes Tr_B |j_A\rangle \langle j_B| = \sum_i \lambda_i^2$$

This means that if AB is in a pure state $|\psi_{AB}\rangle$ then ρ_A and ρ_B have identical sets of non-zero eigenvalues. If $d_A > d_B$ then ρ_A has set of eigenvalues $\{\frac{2}{i}\}^{min(d_A,d_B)}$ and the rest are zero.

Is the schmidt deompistion unique? No if the eginvalues are degenerate. As you can generate the schmidt decompistion by diagonlising ρ_A and ρ_B then matching eigenvectors of the same eigenvalues (which is non-unique if there are degenerate eigenstates.

The set of $\{\lambda_i\}$ are called Schmidt coefficents, and the set $\{|i_A\rangle\}$, $\{|i_B\rangle\}$ are the schmidt baseses, and $n(\psi_{AB})$ (the number of non-zero schmidt coefficents) is called the Schmidt rank. The schmidt rank is the simplest signature of entanglement as $|\psi_{AB}\rangle$ is entangled iff $n(\psi_{AB}) > 1$. Easy to see that $n(\psi_{AB}) = 1 \implies$ product state as then can write $\psi_{AB} = |i_A\rangle \otimes |j_B\rangle$ as only one non-zero schmidt coefficent.

2.2 Purification

It is possible to convert a mixed state into a pure state by adding a purifying reference system R with Hilbert space H_R, and defining a pure state $|\psi_{AR}\rangle \in H_A \otimes H_B$ such that:

$$\rho_A = Tr_R |\psi_{AR}\rangle \langle\psi_{AR}| = \sum_{i=1} \lambda_i^2 |i_A\rangle \langle i_A|$$

Given a ρ_A we write its spectral decompistion (diagonlise it):

$$\rho_A = \sum_{i=1}^{d_A} p_i |i_a\rangle \langle i_A|$$

Define

$$|\psi_{AR}\rangle = \sum_{i=1}^{d_A} \sqrt{p_i} |i_A\rangle |i_R\rangle$$

then it is easy to check this statifies the relation above. More generally there is a canonical way of writing a purification:

$$|\psi_{AR}\rangle = (\sqrt{\rho_A} \otimes I) |\tilde{\Omega}\rangle$$

for $|\tilde{\Omega}\rangle = \sqrt{d} |\Omega\rangle$ and $|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i_A\rangle |i_R\rangle$. Need to check that $\rho_A = Tr_R |\psi_{AR}\rangle \langle\psi_{AR}|$ Usefulness of pure states:

$$\rho = \sum_i |i\rangle \langle i| \implies f(\rho) = \sum f(\lambda_i) |i\rangle \langle i| \implies \sqrt{\rho} = \sum \sqrt{\lambda_i} |i\rangle \langle i|$$

2.3 No-cloning theorem (N.C thm)

There does not exist a universal quantum copier. You cannot make perfect copies of arbitary unknown quantum states. You can copy some states under-some conditions.

Proof by contradiction:

Assume there exists a universal quantum copier which takes arbitary states $|\psi\rangle$ or $|\phi\rangle$ and blank slate $|s\rangle$. Assume there exists U s.t.

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

Take inner product of LHS

$$\langle\psi|\langle s|U^\dagger U|\psi\rangle|s\rangle = \langle\psi|\langle\psi|\phi\rangle|\phi\rangle$$

$$\langle\psi|\phi\rangle\langle s|s\rangle = \langle\psi|\phi\rangle^2$$

so cannot copy arbitary states can only copy orthogonal states with the same copier as must have $\langle\psi|\phi\rangle = 0$ (orthogonal) or $\langle\psi|\phi\rangle = 1$ (identical states).

Second proof by contradiction Use qubits: $\psi = \alpha|0\rangle + \beta|1\rangle$, and $|s\rangle = |0\rangle$. Therefore assume that U exists such that:

$$U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle = \alpha^2|0\rangle|0\rangle + \alpha\beta|0\rangle|1\rangle + \alpha\beta|1\rangle|0\rangle + \beta^2|1\rangle|1\rangle$$

$$U|0\rangle|0\rangle = |0\rangle|0\rangle, U|1\rangle|0\rangle = |1\rangle|1\rangle$$

So,

$$U(|\psi\rangle|0\rangle) = U(\alpha|0\rangle + \beta|1\rangle)|0\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle$$

So this is a contradiction unless $\alpha = 1, \beta = 0$ or $\alpha = 0, \beta = 1$.

An implication of the no-cloning theorem is that superluminal communication is impossible.

2.4 Maximally Entangled States

$$|\psi_{AB}\rangle, \lambda_i = \frac{1}{\sqrt{d}}, d = min(d_A, d_B)$$

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d}|i_A\rangle|i_B\rangle$$

Its reduced states are completely mixed states if $d_A = d_B = d$:

$$\rho_A = \frac{1}{d}\sum|i_A\rangle\langle i_A|, \rho_B = \frac{1}{d}\sum|i_B\rangle\langle i_B|$$

If $d_A < d_B$ then ρ_A is a completely mixed state and ρ_B is a completely mixed state on their supports:

$$\rho_A = \frac{1}{d_A}\sum|i_A\rangle\langle i_A|, \rho_B = \frac{1}{d_A}\sum^{d_A}|i_B\rangle\langle i_B|$$

Properties
$|\psi_{AB}\rangle = \mathrm{H}_A \otimes \mathrm{H}_B \approx \mathbb{C}^d \otimes \mathbb{C}^d$ These are two qudits. Fix the orthonormal basis on $\mathbb{C}^2$ of $\{|i\rangle\}_{i=1}^d$ and let:

$$|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle |i\rangle$$

Lemma 1: $\forall A, B \in \mathfrak{B}(\mathbb{C}^d)$ we have $\langle\Omega| A \otimes B |\Omega\rangle = Tr(A^T B)$ where T is the transposition in the chosen basis (schmidt basis of $|\Omega\rangle$

Lemma 2: Ricoche trick $(A \otimes I)|\Omega\rangle = (I \otimes A^T)|\Omega\rangle$
Proof:

$$A = \sum_{i=1}^{d} a_{jk} |j\rangle \langle k|$$

LHS of lemma 2:

$$= \frac{1}{\sqrt{d}} \sum_i [(\sum_{jk} a_{jk} |j\rangle \langle k|) \otimes I] |i\rangle \otimes = \frac{1}{\sqrt{d}} \sum_{ij} a_{ji} |j\rangle |i\rangle$$

$$A^T = \sum_{kj} |j\rangle \langle k|$$

RHS of lemma 2:

$$= \frac{1}{\sqrt{d}} \sum_{ij} |i\rangle \otimes a_{ij} |j\rangle = \frac{1}{\sqrt{d}} \sum_{ij} a_{ij} |j\rangle |i\rangle$$

Remember these lemmas we will use it over and over again.
Implications of Lemma 1
Given a state ρ_A then the canonical purification is$|\psi_{AB}\rangle = \sqrt{d}(\sqrt{\rho_A} \otimes I)|\Omega\rangle$ as:

$$\rho_A = Tr_B |\psi_{AB}\rangle \langle\psi_{AB}| = Tr_B d(\sqrt{\rho_A} \otimes I)|\Omega\rangle \langle\Omega| (\sqrt{\rho_A} \otimes I) = Tr_B \sum_{ij} \sqrt{\rho_A} |i\rangle \langle j| \sqrt{\rho_A} \otimes |i\rangle \langle j| = \sum_i \sqrt{\rho_A} |i\rangle \langle$$

Every biparitate state can be written:

$$|\psi_{AB}\rangle = (I \otimes R)|\Omega\rangle$$

for some operator R. Proof by construction:

$$|\psi_{AB}\rangle = \sum \lambda_i |i_A\rangle |i_B\rangle \in \mathrm{H}_A \otimes \mathrm{H}_B$$

Choose two isometries U, V s.t.

$$U^\dagger U = V^\dagger V = I, U|i\rangle = |i_A\rangle, V|i\rangle = |i_B\rangle, D = \sum \lambda_k |k\rangle \langle k|$$

Let $R = \sqrt{d} V D U^T$ therefore:

$$(I \otimes R)|\Omega\rangle = (I \otimes VDU^T)|\Omega\rangle = (I \otimes VD)(I \otimes U^T)|\Omega\rangle =^{\text{lemma 2}} (U \otimes VD)|\Omega\rangle$$

$$(I\otimes R)\left|\Omega\right\rangle = \sum_i (U\otimes VD)\left|i\right\rangle\otimes\left|i\right\rangle = \sum_i \left|i_A\right\rangle\otimes V\sum_k \lambda_k \left|k\right\rangle\left\langle k\right|\left|i\right\rangle = \sum_i \left|i_A\right\rangle\otimes V\lambda_i\left|i\right\rangle = \sum_i \lambda_i \left|i_A\right\rangle\otimes\left|i_B\right\rangle$$

Aside about Schmidt
$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\xi_B\rangle \neq \sum_{i>1} \lambda_i |i_A\rangle |i_B\rangle$ as the schmidt coefficents are unique as they are the eigenvalues of ρ_A and ρ_B, so the schmidt rank is unique.

3 Time evolution of open quantum system

Dynamics is given by a quantum operation (also called a quantum channel).
Quantum Operation: $\Lambda : \mathrm{D(H)} \to \mathrm{D(K)}$ and is a completely positive, trace-preserving map (CPTP) map
- It enables a description of discrete state changes

$$\Lambda\rho_{t=0} \to \rho'_{t>0} = \Lambda(\rho)$$

Λ is a superoperator and can map from operators in general H to other operators on H: $\Lambda : \mathrm{B(H)} \to \mathrm{B(K)}$.
Properties of Λ
- Linearity $\Lambda(p_1\rho_1 + p_2\rho_2) = p_1\Lambda(\rho_1) + p_2\Lambda(\rho_2)$
- Trace preserving: $Tr\Lambda(\rho) = Tr\rho = 1$, $\rho \in \mathrm{D(H)}$ this corresponds to the conservation of probablity.
- Positivity [positive(-semidefiniteness) preserving] $\rho \geq 0 \implies \Lambda(\rho) \geq 0$
- Completely positive (CP):

$$(\Lambda \otimes id_B) : \mathrm{D(H}_A) \otimes \mathrm{B(H}_B) \to \mathrm{D(K)} \otimes \mathrm{B(H}_B)$$

is completely positive if $(\Lambda \otimes id_B)$ is positive for all B:

$$(\Lambda \otimes id_B)\rho_{AB} \geq 0$$

Further consideration of completely positiveness (CP)

$$\mathrm{H} = \mathbb{C}^m, \mathrm{K} = \mathbb{C}^n$$

use notation: M_m is set of complex $m \times m$ matrices, M_m^+ is set of positive semi definite complex $m \times m$ matrices, $\mathrm{B}(\mathbb{C}^n) = M_m$, $\mathrm{D(H)} = \{\rho \in M_m^+, Tr\rho = 1\}$.

A linear map $\Lambda : M_m \to M_n$ is positive if $\Lambda(A) \in M_n^+$ if $A \in M_m^+$.

k-positive for any $k \geq 1$ if $(\Lambda \otimes id_k)$ is positive

Is completely positive if it is k-positive for all positive integres k

3.0.1 Necessary and sufficent condtion for complete positivity

A linear map $\Lambda : \mathrm{B(H)} \to \mathrm{B(K)}$ with $\mathrm{H} = \mathbb{C}^d$ and $\mathrm{K} = \mathbb{C}^{d'}$ is completely positive iff

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

Proof: If it is completely positive then it is obvious as it is a completely positive operator acting on a positive state so it will be positive. Now proof opposite direction:

Consider an arbitary $k \geq 1$ and a bipartite state $\rho \in M_d \otimes M_k$ with spectral decomposition: $\rho = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle \langle\psi_i|$.

$$(\Lambda \otimes id_k)\rho \geq 0 \iff \sum \lambda_i (\Lambda \otimes id_k) |\psi_i\rangle \langle\psi_i| \geq 0 \iff (\Lambda \otimes id_k) |\psi_i\rangle \langle\psi_i| \geq 0 \forall i$$

second iff holds as $\lambda_i \geq 0$. As any bipartite state can be written $|\psi\rangle = (I \otimes R) |\Omega\rangle$. Since $|\Omega_i\rangle \in \mathbb{C}^d \otimes \mathbb{C}^k$ there exists R_i s.t. $|\psi_i\rangle = (I \otimes R_i) |\Omega\rangle$. So the above can be written as:

$$(\Lambda \otimes id_k)(I \otimes R_i) |\Omega\rangle \langle\Omega| (I_d \otimes R_i^\dagger) \geq 0$$

Note that $R_i \in \mathcal{B}(\mathbb{C}^d, \mathbb{C}^d)$. Define a superoperator $Q_i : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^k)$ such that $Q_i(.)R_i(.)R_i^\dagger$. So:

$$(\Lambda \otimes id_k)(I \otimes R_i) |\Omega\rangle \langle\Omega| (I_d \otimes R_i^\dagger) = (\Lambda \otimes id_k)(id_d \otimes Q_i) |\Omega\rangle \langle\Omega| = (id_{d'} \otimes Q_i)(\Lambda \otimes id_d)(|\Omega\rangle \langle\Omega| \geq 0$$

$$(id_{d'} \otimes R_i)(\Lambda \otimes id_d)(|\Omega\rangle \langle\Omega| (id_{d'} \otimes R_i^\dagger) \geq 0$$

let $A = (id_{d'} \otimes R_i)$ and $B = (\Lambda \otimes id_d) |\Omega\rangle \langle\Omega|$ as

$$B \geq 0 \implies ABA^\dagger \geq 0$$

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