CSC2429 Introduction to Quantum Information Theory

1 Classical Information Theory

1.1 Shannon Entropy and Source Coding Theorem

***Definition:* 1.1: Alphabet and Closure**

Let $\Sigma = \{0, 1\}$ be an alphabet, the Klein Closure $\Sigma^* = \{0, 1\}^*$ is the set of strings of arbitrary length, containing alphabet 0 and 1.

Examples:

- set of string of length 2: $\{0, 1\}^2 = \{00, 01, 10, 11\}$.
- set of string of length at most 2: $\emptyset \cup \{0, 1\} \cup \{0, 1\}^2$
- $\Sigma^* = \emptyset \cup \{0, 1\} \cup \{0, 1\}^2 \cup \cdots$

***Definition:* 1.2: Entropy**

Let P be a probability distribution of a r.v. X over a set S, then **Entropy** of X is

$$H(X) = -\sum_{x \in S} P(x) \log P(x) = -\mathbb{E}(\log P(x)) \tag{1}$$

.

If an event has probability 0, then including it or not in the set does not matter. Define $0 \log 0 = \lim_{x \to 0} x \log x = 0$ when $P(x) = 0$. $H(X)$ gives the number of bits that is required at least to encode X.

Example: Assume $P(X = 0) = 1$, $P(X = 1) = 0$, then $H(X) = -0 \log 0 - 1 \log 1 = 0$. If a seqnece of signals containing 0 only is sent, the optimal encoder doesn't require any input.

Example: Assume $P(X = 0) = \frac{1}{2}$, $P(X = 1) = \frac{1}{2}$, then $H(X) = 2\left(-\frac{1}{2} \log \frac{1}{2}\right) = 1$. If a string is of length n, there are 2^n equally likely strings, and it requires n bits to encode.

***Definition:* 1.3: Entropy Rate**

Let $P^{(n)}$ be a family of distributions on S^n. The **Entropy Rate** of the source X of the signals is

$$\lim_{n \to \infty} \frac{1}{n} \sum_{x \in S^n} -P^{(n)}(x) \log(P^{(n)}(x)) \tag{2}$$

.

The sum is the number of bits required to encode a specific string x of size n. The limit gives the asymptotic average of entropy on each bit.

Definition: 1.4: $\epsilon-$typical Sequence

Let $\epsilon > 0$, if the string $(x_1, ..., x_n)$ are drawn i.i.d. from a distribution, then the sequence $(x_1, ..., x_n)$ is $\epsilon-$**typical** if $2^{-n(H(X)+\epsilon)} \leq P(x_1, ..., x_n) \leq 2^{-n(H(X)-\epsilon)}$ for some ϵ.

For a uniform distribution of a space of size D, $H(X) = \log D$, then the sequence is $\epsilon-$typical if $\frac{2^{-n\epsilon}}{D^n} \leq P \leq \frac{2^{n\epsilon}}{D^n}$.

Lemma 1. *$\forall \delta > 0$, $\exists n$ s.t. with probability $> 1 - \delta$, $(x_1, ..., x_n)$ will be $\epsilon-$typical, for any $\epsilon > 0$.*

Proof. Take log in the inequalities defining $\epsilon-$typical sequence, $-n(H(X) + \epsilon) \leq \log P \leq -n(H(X) - \epsilon)$.
Then, $H(X) - \epsilon \leq -\frac{1}{n} \log P \leq H(X) + \epsilon$.
For i.i.d. drawn strings $P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i)$, so $\log P = \log \prod_{i=1}^{n} P(x_i) = \sum_{x \in S} \log P(x)$
Thus, we have $H(X) - \epsilon \leq -\frac{1}{n} \sum_{x \in S} \log P(x) \leq H(X) + \epsilon$.
Pick an x_q, check the number of times x_q ends up appearing in $(x_1, ..., x_n)$.
By Central Limit Theorem[1]: for $\hat{P}(x_q)$, $\mu_{x_q} = P(x_q)$, $\sigma^2_{x_q} = \frac{\sigma^2}{n}$, *i.e.* $x_q \sim \mathcal{N}\left(P(x_q), \frac{\sigma^2}{n}\right)$
$H(X) - \epsilon \leq -\sum_{x \in S} P(x) \log \hat{P}(x) + o(1) = H(X) + o(1) \leq H(x) + \epsilon$
By Chebyshev's inequality[2], choose $k = (1 - \delta)^{-\frac{1}{2}}$, $\Pr\left(\left|\sum_{x_q} \hat{P}(x_q) - H(X)\right| \geq (1 - \delta)^{-\frac{1}{2}} \frac{\sigma}{\sqrt{n}}\right) \leq 1 - \delta$.
Thus for $n \geq \sigma^2(1 - \delta)/\epsilon$, it follows that the error will be at most ϵ with probability greater than $1 - \delta$. □

1.2 Compression

Definition: 1.5: Reliable Compression Scheme

Let $x_1, ..., x_n \in \Sigma^*$, a compression scheme C of rate R is a map from $(x_1, ..., x_n)$ to nR bits. It is **reliable** if there exists a decoding map D s.t. $D(C(x_1, ..., x_)) = x_1, ..., x_n$ with $P \approx 1$.

Theorem: 1.1: Existence of Reliable Compression Scheme

Let $x_1, ..., x_n$ be i.i.d. drawn from X, and C be a compression scheme of rate R.

- If $R > H(X)$, then a reliable C exists.
- If $R < H(X)$, then no reliable C exists.

i.e. it is not possible to compress information further than the entropy while still having a one-to-one reliable decoding map.

Proof. If $R > H(X)$, we can construct an encoder as follows:

1: **function** ENCODER($x_1, ..., x_n$)
2: Determine if $(x_1, ..., x_n)$ is $\epsilon-$typical for some $\epsilon > 0$
3: If it is typical, then it must be one of at most $2^{n(H(x)+\epsilon)}$ strings[3]. By definition of entropy rate, $R = \lim_{n \to \infty} \frac{1}{n}(n(H(X) + \epsilon)) = H(X) + \epsilon > H(X)$.

[1] For i.i.d. distributed r.v.s, the sampling distribution of the standardized sample mean tends towards the standard normal distribution

[2] $\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

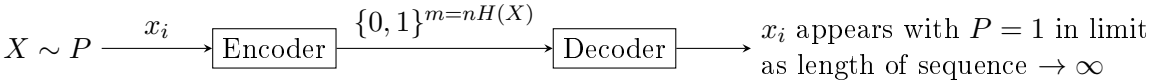
[3] We know that $P(x_1, ..., x_n) \geq 2^{-n(H(x)+\epsilon)}$ and $\sum_{x_1, ..., x_n} P(x_1, ..., x_n) = 1$. Thus, $\sum_{x_1, ..., x_n} 2^{-n(H(x)+\epsilon)} \leq 1$, then the number of elements $\leq 2^{n(H(x)+\epsilon)}$.

4: If it is not typical, then return a $nH(X)$ bit string (failed), but the probability of this outcome is near 0.

5: **end function**

If $R < H(X)$, then $\leq nR$ bits are needed to encode typical sequence.
There are at least $2^{n(H(X)-\epsilon)}$ typical strings.
To make the encoder reliable, we must need at least $nH(X)$ bits to encode. □

Final comments: Shannon entropy gives the optimal # bits for an encoder to encode the strings $X \to (x_1, ..., x_n)$.



Note: Most likely string may not be a typical string.

1.3 Mutual Information

Information is meant to be shared. Mutual information or information gain measures the number of bits of information that A has about B's samples given A's samples.

***Theorem:* 1.2: Jenson's inequality**

If $F(x)$ is concave, then $\mathbb{E}(F(x)) = \sum_x P(x)F(x)$ obeys $\mathbb{E}(F(x)) \leq F(\mathbb{E}(x))$. If $F(x)$ is convex, then $\mathbb{E}(F(x)) \geq F(\mathbb{E}(x))$.

***Definition:* 1.6: Mutual Information**

Let X, Y be two r.v.s over Σ_X, Σ_Y respectively. The mutual information is $I(X;Y) = H(X) + H(Y) - H(X,Y)$[a].

[a] If $(X,Y) \sim P(X,Y)$, then $H(X,Y) = -\sum_{X,Y} P(X,Y)\log(P(X,Y))$

Lemma 2. *Mutual information satisfies the following:*

1. $I(X;Y) \geq 0$ *(can be proved by Jenson's inequality)*
2. $I(X;Y) = 0$ *if and only if the distributions* X *and* Y *are independent*
3. $I(X;Y) = I(Y;X)$

The distance measure on distributions $D(X,Y) = H(X,Y) - I(X;Y)$ can be interpreted as the information needed to encode samples from the joint distribution minus the information shared in common between the marginal distributions over X and Y individually. This quantity actually can be seen to serve as a metric in that it is symmetric, obeys the triangle inequality and non-negativity etc. This quantity is known as the *variation of information*.

1.4 Relative Entropy (KL Divergence)

Relative entropy measures the number of bits of information that A would need to encode B's samples given A's own data.

***Definition:* 1.7: Relative Entropy (Kullback-Leibler Divergence)**

Let X, Y be two r.v.s over Σ_X, with probability distributions $X \sim P$, $Y \sim Q$, then the relative entropy is

$$H(X||Y) = D_{KL}(X||Y) = -H(X) - \sum_i P(x_i)\log(Q(x_i)) = \sum_i P(x_i)\log\left(\frac{P(x_i)}{Q(x_i)}\right) \tag{3}$$

.

Relative entropy measures the statistical distance of two distributions, *i.e.* how distinguishable two distributions are from each other.

Lemma 3. *Relative entropy satisfies the following:*

1. $H(X||Y) \geq 0$
2. $H(X||Y) \neq H(Y||X)$
3. $I(X;Y) = H(P(X,Y)||P(X)P(Y))$

Proof. (Lemma 1)
$H(X||Y) = \sum_i P(x_i)\log\left(\frac{P(x_i)}{Q(x_i)}\right) = -\sum_i P(x_i)\log\left(\frac{Q(x_i)}{P(x_i)}\right) \geq -\log\left(\sum_i \frac{PQ}{P}\right) = -\log(\sum_i Q) \geq 0$ by Jenson's inequality. □

$H(X||Y)$ quantifies the additional bits needed to encode X if the encoder is able to encode Y.

Lemma 4. *Let B_n be any set of sequences $x_1, ..., x_n$ such that for some distribution P_1, $P_1(B_n) \geq 1 - \epsilon$, and let P_2 be any other distribution s.t. $H(P_1||P_2) < \infty$, then $P_2(B_n) > (1-2\epsilon)2^{-n(H(P_1||P_2)+\epsilon)}$.*

Note: if $H(X||Y) = \infty$, we can distinguish $X$and Y by a finite number of observations.

2 Quantum Formalism

In classical mechanics, suppose we want to measure the position x and momentum p from a joint distribution $\Pr((x,p))$, the order of measurement doesn't matter. We can always measure both x and p exactly. However, in quantum mechanics, Heisenberg uncertainty tells us that once we measure position, the probability of position is concentrated, but the probability of momentum becomes arbitrarily spreaded. The same applies when we measure momentum first. We cannot measure both position and momentum exactly at the same time.

A quantum state is like a probability distribution except for two things:

1. It can describe the outcome of every possible experiment that could be performed on the system rather than simply reporting the probability of a single measurement outcome.
2. The probability of outcomes depend on the order of questions.

Definition: 2.1: Adjoint

$(\cdot)^\dagger$ is the adjoint operator. if $\rho = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $\rho^\dagger = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix}$, i.e. the complex transpose. If $|q\rangle = \begin{pmatrix} a \\ b \end{pmatrix}$ is a state vector, then the adjoint $|q\rangle^\dagger = (a^*, b^*)$.

Example: Suppose $v = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}$ a unit vector, $v^\dagger v = 1$.

Note: $(AB)^\dagger = B^\dagger A^\dagger$.

Definition: 2.2: Quantum State/Density Matrix

Let D be a non-negative integer, then the quantum state (density matrix) is a mapping ρ from the complex Euclidean space to itself. *i.e.* $\rho : \mathbb{C}^D \to \mathbb{C}^D$, such that:

1. ρ is linear. *i.e.* ρ can be represented by a matrix.
2. $\mathrm{Tr}(\rho) = \sum_i e_i^\dagger \rho e_i = \sum_i \lambda_i(\rho) = 1$, so it has a probability interpretation.
3. $\rho = \rho^\dagger$ if $\rho \geq 0$ (PSD).

Definition: 2.3: Hermitian Operator

A is Hermitian if it is a linear operator on a complex Euclidean space s.t. $A^\dagger = A$, *i.e.* All eigenvalues of A are real-valued.

Example: $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and Pauli matrices $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, all have $\lambda = \pm 1$.

Definition: 2.4: Expectation of an operator

Let A be an operator on $\mathbb{C}^D$ and ρ be a density operator. Then $\mathbb{E}(A) = \mathrm{Tr}(A\rho)$.

Example: Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\rho = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, then $\mathbb{E}(A) = \mathrm{Tr}\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \right) = \mathrm{Tr}\left(\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix} \right) = \frac{1}{2}$. A projects the outcome $|0\rangle$ and measures its probability. Similarly, $\Pr(|1\rangle) = \frac{1}{2}$.

Note for the trace operator: $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, $\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA)$. The order of matrix multiplication doesn't matter for the final trace calculation.

Definition: 2.5: Normal and unitary operators

Let M be a normal matrix, there exists change of basis matrix U and $D = \mathrm{diag}(\lambda_1, ..., \lambda_n)$ s.t. $M = UDU^\dagger$ and U is unitary, i.e. $U^\dagger = U^{-1}$, $U^\dagger U = UU^\dagger = I$.

For a normal matrix M, $\mathrm{Tr}(M) = \mathrm{Tr}(UDU^\dagger) = \mathrm{Tr}(UU^\dagger D) = \mathrm{Tr}(D) = \sum_i \lambda_i$.
Let V be any unitary operator, then $\mathrm{Tr}(VMV^\dagger) = \mathrm{Tr}(M)$. The formalism doesn't depend on the choice of basis.

Definition: 2.6: Pure and Mixed States

A pure state is a density operator that is rank $= 1$, *i.e.* it has 1 non-zero eigenvalue. A mixed state has rank > 1.

If A is a pure state, then $A = VV^\dagger$ for unit column vector V, e.g. $A = e_0 e_0^\dagger$ is pure. In Dirac's notation $A = |v\rangle\langle v|$. If $\rho = uu^\dagger$, then $\Pr(\rho = VV^\dagger) = \mathrm{Tr}(VV^\dagger\rho) = V^\dagger\rho V = |V^\dagger u|^2$.

2.1 Projector and Measurement

Definition: 2.7: Projective measurement operator

Let Π be a Hermitian operator such that $\Pi^2 = \Pi$ on a complex Euclidean space $\mathbb{C}^\Sigma$ we call Π a projective measurement operator

Suppose $P = \begin{pmatrix} p_1 \\ p_2 \\ 1-p_1-p_2 \end{pmatrix}$ a classical probability vector, the corresponding quantum state is $\rho = \begin{pmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & 1-p_1-p_2 \end{pmatrix}$. Suppose $\Pi = e_0 e_0^\dagger$ is a projective measurement operator. After the measurement, the state will transite to:

$$\rho \to \begin{cases} \frac{\Pi\rho\Pi}{\mathrm{Tr}(\Pi\rho\Pi)} & \text{with probability } \mathrm{Tr}(\Pi\rho\Pi) = p_0 \\ \frac{(I-\Pi)\rho(I-\Pi)}{1-\mathrm{Tr}(\Pi\rho\Pi)} & \text{with probability } 1 - \mathrm{Tr}(\Pi\rho\Pi) = 1 - p_0 \end{cases}$$

A projector either projects the state to the subspace that's compatible with the projector's eigenspace, or the remainder.
For $P = \begin{pmatrix} p_1 \\ p_2 \\ 1-p_1-p_2 \end{pmatrix}$, classically, if we measure 0 using Π, we get $P = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ with probability p_1, and $P = \begin{pmatrix} 0 \\ p_2 \\ 1-p_1-p_2 \end{pmatrix}$ with probability $1 - p_1$.
In quantum, the state collapses with the same measurement probability.

Example: Assume $\rho \in \mathbb{C}^{2\times 2}$ (often written as $\rho \in L(\mathbb{C}, \mathbb{C})$). Let $\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = e_0 e_0^\dagger$ be the state.

$\Pi_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = e_0 e_0^\dagger$ and $\Pi_+ = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = H\Pi_0 H$[4] be 2 projector measurements.

a. Measure Π_0 first, then Π_+
   Apply Π_0, $\rho \to e_0 e_0^\dagger$ with probability $\mathrm{Tr}(\Pi_0 \rho) = \mathrm{Tr}(e_0 e_0^\dagger e_0 e_0^\dagger) = 1$
   Then apply Π_+ to $e_0 e_0^\dagger$, it becomes $e_+ e_+^\dagger$ with probability $\mathrm{Tr}(\Pi_+ e_0 e_0^\dagger) = \mathrm{Tr}\left(\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) = \frac{1}{2}$.
   Since the probability is not 1, we must be able to get the compliment $\mathbb{1} - e_+ e_+^\dagger$ with probability $\frac{1}{2}$

b. Measure Π_+ first, then Π_0
   Apply Π_+, $\rho \to e_+ e_+^\dagger$ with probability $\mathrm{Tr}(\Pi_+ \rho) = \frac{1}{2}$
   And it can become $\mathbb{1} - e_+ e_+^\dagger$ with probability $\mathrm{Tr}(\Pi_+ \rho) = \frac{1}{2}$
   Then apply Π_0
   If in the previous step, we get $e_+ e_+^\dagger$ as the final state, then it will change to $e_0 e_0^\dagger$ with probability $\mathrm{Tr}(e_0 e_0^\dagger e_+ e_+^\dagger) = \frac{1}{2}$, and change to $1 - e_0 e_0^\dagger$ with probability $\frac{1}{2}$
   Otherwise, it will change to $e_0 e_0^\dagger$ with probability $\mathrm{Tr}(e_0 e_0^\dagger (1 - e_+ e_+^\dagger)) = \frac{1}{2}$, and change to $1 - e_0 e_0^\dagger$ with probability $\frac{1}{2}$

We can see the difference of quantum and classical probability here. In classical experiments, the experiments always reveal the truth regardless of the order. In quantum experiments, the order of experiments changes the truth.

2.2 Tensor Product

Let X, Y be two uncorrelated r.v.s, s.t. $X \sim \begin{pmatrix} p_1 \\ 1 - p_1 \end{pmatrix}$, $Y \sim \begin{pmatrix} p_2 \\ 1 - p_2 \end{pmatrix}$. Consider the space (X, Y). The probability distribution will be $(X, Y) \sim \begin{pmatrix} p_1 p_2 \\ (1-p_1)p_2 \\ p_1(1-p_2) \\ (1-p_1)(1-p_2) \end{pmatrix} = \begin{pmatrix} p_2 \begin{pmatrix} p_1 \\ 1-p_1 \end{pmatrix} \\ (1-p_2) \begin{pmatrix} p_1 \\ 1-p_1 \end{pmatrix} \end{pmatrix}$.

***Definition:* 2.8: Tensor Product**

Let $A \in L(X, Y)$, $B \in L(W, Z)$, the tensor project is defined as

$$A \otimes B = \begin{pmatrix} A_{00}B & A_{01}B & \cdots & A_{0,Y-1}B \\ A_{10}B & A_{11}B & \cdots & A_{1,Y-1}B \\ \vdots & \vdots & \ddots & \vdots \\ A_{X-1,0}B & A_{X-1,1}B & \cdots & A_{X-1,Y-1}B \end{pmatrix}$$

Tensor products have the following properties:

1. For a constant a, $A \otimes aB = a(A \otimes B)$
2. $A \otimes (B + C) = A \otimes B + A \otimes C$
3. $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$
4. $\mathrm{Tr}(A \otimes B) = \mathrm{Tr}(A)\mathrm{Tr}(B)$

[4] $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = e_+ e_+^\dagger$, where $e_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, is the Hadamard (multidimensional DFT) operator

5. $(A \otimes B)(C \otimes D) = AC \otimes BD$

Suppose $A \in \mathbb{C}^{4\times 4} = \mathbb{C}^{2\times 2} \otimes \mathbb{C}^{2\times 2}$ is a density operator.
Note that $\mathbb{1}_4 = \text{diag}(1,1,1,1) = \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & \mathbb{1}_2 \end{pmatrix} = I_2 \otimes I_2 = \sum_{i=0}^{1} e_i e_i^\dagger \otimes \sum_{j=0}^{1} e_j e_j^\dagger = \sum_{i,j}(e_i e_i^\dagger \otimes e_j e_j^\dagger)$
Then
$A = \mathbb{1}_4 A \mathbb{1}_4 = \sum_{i,j}(e_i e_i^\dagger \otimes e_j e_j^\dagger) A \sum_{k,l}(e_k e_k^\dagger \otimes e_l e_l^\dagger) = \sum_{i,j,k,l}(e_i \otimes e_j)\left[(e_i^\dagger \otimes e_j^\dagger)A(e_k \otimes e_l)\right](e_k^\dagger \otimes e_l^\dagger)$
$= \sum_{i,j,k,l}(e_i \otimes e_j)(e_k^\dagger \otimes e_l^\dagger)A_{i,j,k,l}$
A density operator can be written as $\rho = \sum_{i,j,k,l}(e_i \otimes e_j)(e_k^\dagger \otimes e_l^\dagger)\rho_{i,j,k,l}$ as long as the dimensions are not prime.

Definition: 2.9: Quantum register

Let $\rho \in \mathbb{C}^{p^n \otimes p^n}$, where n is a non-negative integer and p is a prime number. $\rho \in \mathbb{C}^{p\times p} \otimes \mathbb{C}^{p\times p} \otimes \cdots = (\mathbb{C}^{p\times p})^n$ can be written as the tensor product of n copies of density operator. Then ρ is a quantum register.

Note: In quantum computing, $p = 2$ and n is the number of qubits combined.

Definition: 2.10: Subspace

S_A is a subsystem of S if $\exists S_B$ s.t. $S = S_A \otimes S_B$

Example: each copy $\mathbb{C}^{p\times p}$ ia a subsystem of the register.

2.3 Partial Trace

Definition: 2.11: Partial Trace

Let S be a complex Euclidean space and let S_A and S_B be subsystems such that $S = S_A \otimes S_B$. The partial trace of a linear transformation on S over subsystem S_B is denoted $\text{Tr}_B(\cdot)$ and it has the following properties:

a. Let $A : S_A \mapsto S_A$ and $B : S_B \mapsto S_B$ be linear transformations then $\text{Tr}_B(A \otimes B) = A\text{Tr}(B)$
b. For linear transformation C and constant a $\text{Tr}_B(aC) = a\text{Tr}_B(C)$
c. For linear transformations C, D we have $\text{Tr}_B((C + D)) = \text{Tr}_B(C) + \text{Tr}_B(D)$
d. For linear transformation C, $\text{Tr}_A\text{Tr}_B(C) = \text{Tr}(C)$.

Example: Let ρ and σ be pure states (*i.e.* $\text{Tr}(\rho^2) = 1$ or $\text{rank}(\rho) = 1$ or $\exists$ unit vector v s.t. $\rho = vv^\dagger$).
$\rho \otimes \sigma$ is pure.
$\text{Tr}[(\rho \otimes \sigma)^2] = \text{Tr}[(\rho \otimes \sigma)(\rho \otimes \sigma)] = \text{Tr}(\rho^2 \otimes \sigma^2) = \text{Tr}(\rho^2)\text{Tr}(\sigma^2) = 1$

Example (entanglement): Let $\rho = \frac{1}{2}(\overbrace{e_0}^{S_A} \otimes \overbrace{e_0}^{S_B} + e_1 \otimes e_1)(e_0 \otimes e_0 + e_1 \otimes e_1)^\dagger$

$\text{Tr}_B(\rho) = \frac{1}{2}\text{Tr}_B(\overbrace{e_0 e_0^\dagger}^{S_A} \otimes \overbrace{e_0 e_0^\dagger}^{S_B} + e_0 e_1^\dagger \otimes e_0 e_1^\dagger + e_1 e_0^\dagger \otimes e_1 e_0^\dagger + e_1 e_1^\dagger \otimes e_1 e_1^\dagger)$
Note that $\text{Tr}_B(e_0 e_1^\dagger \otimes e_0 e_1^\dagger) = e_0 e_1^\dagger \text{Tr}(e_0 e_1^\dagger) = 0$, and similarly $\text{Tr}_B(e_1 e_0^\dagger \otimes e_1 e_0^\dagger) = 0$.
We get $\text{Tr}_\rho = \frac{1}{2}(e_0 e_0^\dagger + e_1 e_1^\dagger) = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
And $\text{Tr}(\rho_A^2) = \frac{1}{2}$.
This state is not a pure state. In fact, in two dimensions this would be the maximally mixed state which is as far from any pure state as you can possibly get. The information is neither in subsystem A nor B.

It's in the correlation of A and B. This is called *entanglement*.

2.4 Purification

Given an arbitrary quantum state, we can always view the quantum state to be a pure state in a higher dimensional space. The process of introducing (a potentially fictitious) subsystem to create such a pure state is called a purification.

Definition: 2.12: Purification

Let X and Y be complex Euclidean spaces and let $P \geq 0$ be in the set of positive definite operators acting on X ($P \in \text{Pos}(X)$). A vector $u \in X \otimes Y$ is said to be a purification of P if $\text{Tr}(uu^\dagger) = P$ and the state operator ρ is said to be a purification if $\rho = uu^\dagger$ for unit vector u.

Theorem: 2.1: Exisitence of Purification

Let X, Y be complex Euclidean spaces and let $P \in \text{Pos}(X)$. There exists a purification of P, $u \in X \otimes Y$, s.t. $P = \text{Tr}_Y(uu^\dagger)$ if and only if $\dim(Y) \geq \text{rank}(P)$.

Proof. ($\Rightarrow$) Write $P = \sum_{i=1}^{r} \lambda_i v_i v_i^\dagger$, since $P \geq 0$, $\lambda_i \geq 0$ for all i.
Thus, we can define $w = \sum_{i=1}^{r} w_i$, with $w_i = \sqrt{\lambda_i} v_i \otimes e_i$. We claim that $P = \text{Tr}_Y(ww^\dagger)$.

$$
\begin{aligned}
\text{Tr}_Y(ww^\dagger) &= \text{Tr}_Y\left[(\sum_{i=1}^{r} \sqrt{\lambda_i} v_i \otimes e_i)(\sum_{i=1}^{r} \sqrt{\lambda_i} v_i \otimes e_i)^\dagger\right] \\
&= \text{Tr}_Y\left[(\sum_{i=1}^{r} \sqrt{\lambda_i} v_i \otimes e_i)(\sum_{i=1}^{r} \sqrt{\lambda_i} v_i^\dagger \otimes e_i^\dagger)\right] \\
&= \text{Tr}_Y\left[(\sum_{i,j} \sqrt{\lambda_i \lambda_j} v_i v_j^\dagger \otimes e_i e_j^\dagger)\right] \\
&= \sum_{k=1}^{\dim(Y)} \mathbb{1} \otimes e_k^\dagger (\sum_{i,j} \sqrt{\lambda_i \lambda_j} v_i v_j^\dagger \otimes e_i e_j^\dagger)) \mathbb{1} \otimes e_k \\
&= \sum_{k=1}^{\dim(Y)} \lambda_k v_k v_k^\dagger \text{ (only } i = j = k \text{ are preserved in the sum)}
\end{aligned}
$$

Note this is true for any unitary transformation also of the e_i since we only need above that the vectors are orthogonal.

($\Leftarrow$)Now we need to show the reverse direction is impossible by contradiction.
If $P \geq 0$then it follows that there exists a matrix A such that $P = AA^\dagger$ where $A \in L(X, Y)$ can be interpretted as a square root of the operator P. Further it follows that $\text{rank}(P) = \text{rank}(A)$.
Because the rank of a matrix is at most the dimension of the image of the operator, we have that $\text{rank}(A) \leq \dim(Y)$. Thus by contradiction we must have that $\dim(Y) \geq \text{rank}(P)$ □

3 Quantum Channels

A channel describes a process that sends information, or signals, from Alice to Bob. Alternatively, channels describe any transformation that you can perform on data. An arbitrary computation can be thought of a channel.

***Definition:* 3.1: Quantum Channel/CPTP map**

Let X, Y be complex Euclidean spaces. A map $\Phi : L(X) \to L(Y)$ (for linear vector spaces $L(X)$ and $L(Y)$) is a quantum channel if and only if

1. Φ is a completely positive map meaning that for $\rho \in L(X)$ with $\rho \geq 0$ (ie is a positive semi-definite matrix), $\Phi(\rho) \geq 0$.
2. Φ is trace preserving, which means that for $\rho \in L(X)$, $\text{Tr}(\Phi(\rho)) = \text{Tr}(\rho)$.

The channel is also called a Completely Postive Trace-Preserving (CPTP) map.

The CPTP requirements guarantee that if the input to the channel Φ is a valid quantum state, then the output will also be a valid quantum state.

Examples:

- **Identity:** $\text{id} : L(X) \to L(X)$ that for $\rho \in L(X)$, $\text{id}(\rho) = \rho$. This is the ideal transmission channel. It introduces no error. The Shannon entropy is unchanged.
- **Unitary quantum channel:** Given U a unitary matrix ($U^\dagger = U^{-1}$), $\Phi : L(X) \to L(X)$ that for $\rho \in L(X)$, $\Phi(\rho) = U\rho U^\dagger$
- **Change of basis channel:** unitary channels are change of basis channel, since $U\rho U^\dagger$ describes a basis transformation. The eigenvalues of ρ are unchanged[5]. We can also compose change of basis by $\Phi(\rho) = \sum_k \rho_k U_k \rho U_k^\dagger$, which is also a valid channel.
- **Replacement channel:** $\Phi : L(X) \to L(Y)$, $\sigma \in L(Y)$, s.t. $\rho \in L(X)$, $\phi(\rho) = \text{Tr}(\rho)\sigma$. Throws away a state and replaces it with a new one.
- **Depolarizing channel**: $\Phi : L(X) \to L(X)$, s.t. $\rho \in L(X)$, $\Phi(\rho) = \alpha\rho + (1-\alpha)\text{Tr}(\rho)\frac{\mathbb{1}_X}{\dim(X)}$, with $\alpha \in [0, 1]$. If $\alpha = 1$, it is a complete depolarization, giving the highest entropy.
- **Convex combination of channels:** Let $\alpha \in [0, 1]$, $\Phi_0, \Phi_1 : L(X) \to L(Y)$, then $\alpha\Phi_0 + (1-\alpha)\Phi_1$ is a CPTP map.
- **Composition of channels:** If $\Phi : L(X) \to L(Y)$ and $\Psi : L(Y) \to L(Z)$, then $\Psi \circ \Phi : L(X) \to L(Z)$ is a CPTP map.
- **Tensor product channels:** Let $\Phi : L(X) \to L(Y)$ and $\chi : L(v) \to L(W)$ be CPTP maps. Define $\Phi \otimes \chi : L(X) \otimes L(V) \to L(Y) \otimes L(W)$ s.t. for $\rho \in X$ and $\sigma \in V$, $\Phi \otimes \chi(\rho \otimes \sigma) = \Phi(\rho) \otimes \chi(\sigma)$ is a CPTP map.

[5] All quantum computing that can be described using unitary channels can be viewed as reversible classical computing that can be described using permutations.

3.1 Representation of quantum channels

3.1.1 Natural Representation

The natural representation represents all quantum states as vectors (not just the pure states as in Dirac notation). In this representation, we rewrite the density matrix $\rho \in \mathbb{C}^{nm \times nm}$ by:

$$\text{vec}(\rho) = \begin{pmatrix} \rho_0 0 \\ \vdots \\ \rho_{n-1,0} \\ \rho_{0,1} \\ \vdots \\ \rho_{n-1,m-1} \end{pmatrix}$$

The higher dimensional vector contains all of the information of the quantum state, but with a different orientation.

Definition: 3.2: Natural Representation

$K(\Phi) \in L(X \otimes X, Y \otimes Y)$ is a natural representation of $\Phi : L(X) \to L(Y)$ if for all $\rho \in L(X)$, $\text{vec}(\Phi(\rho)) = K(\Phi)\text{vec}(\rho)$

This is often called *Superoperator Representation* and these vectorized states are sometimes written $\text{vec}(\rho) = |\rho\rangle\rangle$

3.1.2 Choi Representation

Definition: 3.3: Choi Representation

The Choi representation of an operator is $J(\Phi) : T(X, Y) \to L(Y \otimes X)$, s.t. $J(\Phi) = (\Phi \otimes I)(\text{vec}(\mathbb{1}_X)\text{vec}(\mathbb{1}_X)^\dagger) = \sum_{a,b} \Phi(E_{a,b}) \otimes E_{a,b}$, where $E_{a,b} = e_a e_b^\dagger$. The action of the map $\Phi(\rho)$ is $\Phi(\rho) = \text{Tr}_X(J(\Phi)(\mathbb{1}_X) \otimes \rho^T)$.

We are using the normal transpose operation here. The concept of states and channels need not to be seen as totally separate. We can use states to represent a channel and operationalize the action of a channel. This is useful for understanding how gate teleportation and magic state injection works in quantum computing wherein a resource state that represents a channel is provided and a protocol reminiscent of the one above is used to perform the action of the state-encoded channel on a target.

3.1.3 Kraus Representation

Definition: 3.4: Kraus Representation

Let $a \in \Sigma$, $A_a, B_a \in L(X, Y)$ be two sets of operators. $\Phi(\rho) = \sum_{a \in \Sigma} A_a \rho B_a^\dagger$.

If the channel is CPTP, then Kraus operators have $A_a = B_a$. But for an arbitrary transformation, A, B don't have to be the same.

3.2 Discrimination

Suppose Alice has a state $\rho = \lambda\rho_0 + (1 - \lambda)\rho_1$. Alice sends ρ_0 to Bob with probability λ and ρ_1 with probability $1 - \lambda$. Bob measures the state using projective measures Π_0, Π_1, with $\Pi_0 + \Pi_1 = \mathbb{1}$. We want to maximize the probability of correct measurement.

		Alice	
		0	1
Bob	0	$\lambda\langle\Pi_0,\rho_0\rangle$	$(1-\lambda)\langle\Pi_0,\rho_1\rangle$
	1	$\lambda\langle\Pi_1,\rho_0\rangle$	$(1-\lambda)\langle\Pi_1,\rho_1\rangle$

We want to maximize the probability of correct measurement, i.e. $\max(\lambda\langle\Pi_0,\rho_0\rangle + (1-\lambda)\langle\Pi_1,\rho_1\rangle)$

$$\begin{aligned}
\lambda\langle\Pi_0,\rho_0\rangle + (1-\lambda)\langle\Pi_1,\rho_1\rangle &= \langle\Pi_0,\lambda\rho_0 + (1-\lambda)\rho_1 - (1-\lambda)\rho_1\rangle + \langle\Pi_1,(1-\lambda)\rho_1 + \lambda\rho_0 - \lambda\rho_0\rangle \\
&= \langle\Pi_0,\lambda\rho_0\rangle + \langle\Pi_0,(1-\lambda)\rho_1\rangle - \langle\Pi_0,(1-\lambda)\rho_1\rangle + \langle\Pi_1,(1-\lambda)\rho_1\rangle + \langle\Pi_1,\lambda\rho_0\rangle - \langle\Pi_1,\lambda\rho_0\rangle \\
&= \langle\Pi_0+\Pi_1,\lambda\rho_0\rangle + \langle\Pi_0+\Pi_1,(1-\lambda)\rho_1\rangle - \langle\Pi_0,(1-\lambda)\rho_1\rangle - \langle\Pi_1,\lambda\rho_0\rangle \\
&= \lambda\mathrm{Tr}(\rho_0) + (1-\lambda)\mathrm{Tr}(\rho_1) - \langle\Pi_0,(1-\lambda)\rho_1\rangle - \langle\Pi_1,\lambda\rho_0\rangle \\
&= 1 - \langle\mathbb{1}-\Pi_1,(1-\lambda)\rho_1\rangle - \langle\Pi_1,\lambda\rho_0\rangle \\
&= 1 - \langle\mathbb{1},(1-\lambda)\rho_1\rangle + \langle\Pi_1,(1-\lambda)\rho_1\rangle - \langle\Pi_1,\lambda\rho_0\rangle \\
&= \lambda - \langle\Pi_1,\lambda\rho_0 - (1-\lambda)\rho_1\rangle \\
&= 1 - \lambda + \langle\Pi_0,\lambda\rho_0 - (1-\lambda)\rho_1\rangle \text{ (replace } \Pi_1 \text{ with } \Pi_0)
\end{aligned}$$

So we are maximizing $\langle\Pi_0,\lambda\rho_0-(1-\lambda)\rho_1\rangle = \mathrm{Tr}\left[\Pi_0(\lambda\rho_0-(1-\lambda)\rho_1)\right]$ $\lambda\rho_0-(1-\lambda)\rho_1$ is Hermitian, since it is a linear combination of Hermitians. We can then write $\lambda\rho_0-(1-\lambda)\rho_1 = \sum_i \lambda_i v_i v_i^\dagger$, with $\lambda_i \in \mathbb{R}$.

$\mathrm{Tr}\left[\Pi_0(\lambda\rho_0-(1-\lambda)\rho_1)\right] = \mathrm{Tr}\left[\sum_i(\lambda_i\Pi_0 v_i v_i^\dagger)\right] = \sum_{i,j}(\lambda_i v_j^\dagger\Pi_0 v_i v_i^\dagger v_j) = \sum_i \lambda_i v_i^\dagger\Pi_0 v_i$

Let $\Pi_0 = \sum_{i \text{ s.t. } \lambda_i\geq 0} v_i v_i^\dagger - \sum_{i \text{ s.t. } \lambda_i\leq 0} v_i v_i^\dagger$, we get the maximum value of the trace to be $\sum_i|\lambda_i| = \|\lambda\rho_0-(1-\lambda)\rho_1\|_1$.

3.3 More Examples of Quantum Channels

Bit flip channel: If we have a classical bit with probability vector $p = \begin{pmatrix} a \\ 1-a \end{pmatrix}$, the equivalent quantum state is $\rho = \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix}$.

Classically, a bit flip can be applied as $Xp = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a \\ 1-a \end{pmatrix} = \begin{pmatrix} 1-a \\ a \end{pmatrix}$.

In quantum, we have $\Lambda : \rho \to X\rho X^\dagger$ for the bit flip, it gives
$\Lambda(\rho) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1-a & 0 \\ 0 & a \end{pmatrix}$.
Note that this is just a change of basis, no information is changed in this unitary channel. If we redefine the basis to be $e_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, we get to the original state.

Depolarizing channel: It completely throw away the information by $\Lambda : \rho \to \frac{\mathbb{1}}{D}$, where D is the dimension.
Note that unitary channels do nothing to a completely mixed channel $\Lambda_U(\frac{\mathbb{1}}{D}) = \frac{U\mathbb{1}U^\dagger}{D} = \frac{\mathbb{1}}{D}$. No matter what unitary we use to measure the state, we get no knowledge about it, as any measurement gives the same result.

Side Note: Any classical computation is a unitary channel with linear permutation operators. Any quantum computation is a unitary channel.

Isometry Channel: Let S be a complex Euclidean space, $\Lambda : S \to S \otimes Y$ a CPTP map for some complex Euclidean space Y. Λ is an isometry channel if $\Lambda(\rho) = A\rho A^\dagger$ with $AA^\dagger = I$ and $A \in L(S, S \otimes Y)$.

3.4 Generalized Measurement

Recall that a measurement is the expected value of an operator on a quantum state $\langle \Pi \rangle = \text{Tr}(\Pi\rho)$. Suppose we have 2 outcomes Π and $1 - \Pi$, then

$$M : \rho \to \begin{cases} \frac{\Pi\rho\Pi}{\text{Tr}(\Pi\rho\Pi)} & \text{with probability } \text{Tr}(\Pi\rho\Pi) = p_0 \\ \frac{(I-\Pi)\rho(I-\Pi)}{1-\text{Tr}(\Pi\rho\Pi)} & \text{with probability } 1 - \text{Tr}(\Pi\rho\Pi) = 1 - p_0 \end{cases}$$

The measurement may be inconclusive, or a single measurement tells us information about other measurement. For example, if $\rho = \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix}$, and we measure 3 outcomes $e_0 e_0^\dagger$, $\frac{1}{2}(e_0 + e_1)(e_0 + e_1)^\dagger$, $e_1 e_1^\dagger$. The second measurement contains information about the other two measurements.

We want to model all possible measurements to know the ultimate limitation of potential outcomes. Potentially we can increase the dimension of the space.

***Definition:* 3.5: Positive Operator Valued Measure**

Let S be a complex Euclidean space and let $\{P_i : i \in \Sigma\}$ be positive semi-definite linear operators mapping S to S, where Σ is an alphabet. $\mu : \Sigma \to \text{Pos}(S)$ is a Positive Operator Valued Measure (POVM) if

1. $\sum_i P_i^2 = 1$
2. $P_i \geq 0$ are PSD
3. $\mu(a) = P_a$, for all $a \in \Sigma$

Measuring a state operator $\rho : S \to S$ using the POVM $\{P_i\}$ is defined to yield index i with probability $\text{Tr}(P_i\rho)$ and results in the transformation $\rho \to P_i\rho P_i / \text{Tr}(P_i\rho P_i)$

***Theorem:* 3.1: Naimark's Dilation Theorem**

Let S be a complex Euclidean space, Σ be an alphabet and μ be a POVM on the space S. There exists an isometry channel $\Lambda : L(S) \to L(S, S \otimes Y)$ s.t. $\mu(a) = A^\dagger(I_X \otimes E_{aa})A$, where E_{aa} is elementary projector corresponding to $a \in S$ and $\Lambda(\rho) = A^\dagger \rho A$.

3.5 Generalized Discrimination

***Definition:* 3.6: Trace Distance**

Let ρ, σ be bounded linear operators on a d-dimensional complex Euclidean space for positive integer d. We then define the trace distance $D_1(\rho, \sigma) = \frac{1}{2}\text{Tr}(|\rho - \sigma|) = \frac{1}{2}\text{Tr}(\sqrt{(\rho-\sigma)(\rho-\sigma)^\dagger}) = \frac{\|\rho-\sigma\|_1}{2}$.

Analog: The toal variation distance in classical probability is $\text{TVD}(P, Q) = \frac{1}{2}\sum_i |P_i - Q_i|$.

***Definition:* 3.7: Holder's Inequality**

Let S be a finite-dimensional complex Euclidean space and let $A : S \to S$ and $B : S \to S$ be linear operators and let $\|\cdot\|_a$ be the Schatten $a-$norm defined for a map G via $\|G\|_a = (\mathrm{Tr}(|G|^a))^{1/a}$ (a-norm of singular values of operator). We then have that the trace of the composition of two linear operators is then, for any $p, q \in [1, \infty]$ st $1/p + 1/q = 1$, bounded above by the Schatten norms of the two individual operators via

$$\mathrm{Tr}(AB) \leq \|A\|_p \|B\|_q.$$

***Theorem:* 3.2: Holevo Helstrom Bound**

Let S be a complex Euclidean space, ρ, σ be state operators acting on S. Given uniform prior probability over the two states, (which can be thought of as providing a state $(\sigma+\rho)/2$), there exists a POVM $\mu : \{0, 1\} \to \mathrm{Pos}(S)$ with POVM elements $\{E_0, E_1\}$ such that the probability of successfully assigning 0 to σ and 1 to ρ is $\leq \frac{1}{2} + \frac{1}{2} D_1(\rho, \sigma)$.

Proof. Suppose we measure $\{E_0, E_1\}$, the success probability is
$P = \frac{1}{2}\mathrm{Tr}(E_0\sigma) + \frac{1}{2}\mathrm{Tr}(E_1\rho) = \frac{1}{4}\mathrm{Tr}((E_0 + E_1)(\sigma + \rho)) + \frac{1}{4}\mathrm{Tr}((E_0 - E_1)(\sigma - \rho))$
Since $\mu(0) + \mu(1) = E_0 + E_1 = 1$, $\frac{1}{4}\mathrm{Tr}((E_0 + E_1)(\sigma + \rho)) = \frac{1}{4}\mathrm{Tr}(\sigma + \rho) = \frac{1}{4}\mathrm{Tr}(\sigma) + \frac{1}{4}\mathrm{Tr}(\rho) = \frac{1}{2}$
For $\mathrm{Tr}((E_0 - E_1)(\sigma - \rho))$, we consider $A = E_0 - E_1$, $B = \sigma - \rho$, and apply Holder's Inequality with $p = \infty$ and $q = 1$,
$\|E_0 - E_1\|_\infty \leq 1$, and $\mathrm{Tr}((E_0 - E_1)(\sigma - \rho)) \leq 1\|\sigma - \rho\|_1 = 2D_1(\rho, \sigma)$.
Thus $P \leq \frac{1}{2} + \frac{1}{2} D_1(\rho, \sigma)$ □

4 Quantum Entropy and Source Coding Theorem

Definition: 4.1: Von Neumann Entropy

Let S be a complex Euclidean space and $\rho : S \to S$ be a quantum state operator. Then $H(\rho) = -\text{Tr}(\rho \log \rho)$.

Here $\rho \geq 0$ by defintion, so we can write $\rho = UDU^\dagger$ for unitary U and $D = \text{diag}(\lambda_1, ..., \lambda_n)$.
If $\rho = \text{diag}(p_1, ..., p_n)$ the classical probabilities, then $H(\rho) = -\text{Tr}(\rho \log \rho) = -\sum_i p_i \log p_i$ which matches the classical Shannon entropy.

Note that change of basis (changing the representation of the information) won't change the entropy, since change of basis is performed by unitaries U.

Lemma 5. *Let $\Lambda : S \to S \otimes Y$ be an isometry channel, then $H(\rho) = H(\Lambda(\rho))$.*

Proof. If Λ is isometry, then there exists A s.t. $\Lambda(\rho) = A^\dagger \rho A$
$H(\Lambda(\rho)) = H(A^\dagger \rho A) = -\text{Tr}(A^\dagger \rho A \log A^\dagger \rho A)$
Since $\rho \geq 0$, write $\rho = \sum_j \lambda_j v_j v_j^\dagger$ for $\lambda_j \geq 0$ and $v_i^\dagger v_j = \delta_{ij}$
Then $A^\dagger \rho A = \sum_j \lambda_j A^\dagger v_j v_j^\dagger A = \sum_j \lambda_j y_j y_j^\dagger$, where $y_j = A^\dagger v_j$
Define $\sigma = \sum_j \lambda_j y_j y_j^\dagger$. Then $H(\Lambda(\rho)) = H(\sigma) = -\sum_j \lambda_j \log(\lambda_j) = H(\rho)$ □

Other properties of Von Neumann Entropy:

1. For independent states ρ and σ, $H(\rho \otimes \sigma) = H(\rho) + H(\sigma)$.
2. Subadditivity: Let S be complex Euclidean space with subsystems S_A and S_B and let $\rho_{AB} : S \mapsto S$ be a quantum state operator, then $H(\rho_{AB}) \leq H(\rho_A) + H(\rho_B)$.
3. Concavity: for $\lambda \in [0, 1]$, $H(\rho\lambda + \sigma(1 - \lambda)) \geq \lambda H(\rho) + (1 - \lambda)H(\sigma)$.

As an analog to classical source coding theorem, we want to achieve something similar in quantum. Let $\sigma \in L(S)$ be a density operator, where $S \in \mathbb{C}^\Sigma$. (If $\Sigma = \{0, 1\}$ as in quantum computation,
$S = \text{span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right)$) We want to compress $\rho = \sigma \otimes \sigma \otimes \cdots \otimes \sigma = \sigma^{\otimes n}$ to a $m-$qubit space
$Y = (\mathbb{C}^\Sigma)^{\otimes m}$ reliably. In ideal world, we want $\Psi(\Phi(\rho)) = \rho$, where $\Phi : S^{\otimes m} \to Y$ and $\Psi : Y \to S^{\otimes m}$.

4.1 Fidelity

Definition: 4.2: Fidelity

Let S be a complex Euclidean space. Let $\rho, \sigma \in L(S)$ be quantum sates ($\text{Tr}(\rho) = 1$, $\rho \geq 0$). The Fidelity between ρ and σ is

$$F(\rho, \sigma) = \left(\text{Tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right)\right)^2 = F(\sigma, \rho).$$

Note: If $\sigma = yy^\dagger$ a pure state, $(yy^\dagger)^2 = y(y^\dagger y)y^\dagger = yy^\dagger$, since $y^\dagger y = 1$. Then $\sqrt{yy^\dagger} = yy^\dagger$. *i.e.* Every pure state is a projector.
If $\rho = vv^\dagger$, $\sigma = yy^\dagger$ pure states, we can do the projection $\langle v, y\rangle$ to find the similarity between them. Fidelity is a general extension.

Properties of Fidelity:

1. If ρ, σ are pure states, then $F(\rho, \sigma) = \text{Tr}(\rho, \sigma)$
2. $F(\rho, \sigma) \in [0, 1]$ always.

Lemma 6. *Let $\rho = vv^\dagger$ be a pure state on complex Euclidean space S, $\sigma \in L(S)$. Then $F(\rho, \sigma) = v^\dagger \sigma v = Tr(vv^\dagger \sigma) = \langle vv^\dagger, \sigma \rangle$.*

Proof.

$$F(\rho, \sigma) = \text{Tr}^2\left(\sqrt{\sqrt{vv^\dagger}\sigma\sqrt{vv^\dagger}}\right) = \text{Tr}^2\left(\sqrt{vv^\dagger \sigma vv^\dagger}\right)$$
$$= \text{Tr}^2\left(\sqrt{vv^\dagger (v^\dagger \sigma v)}\right) = \left(\sqrt{v^\dagger \sigma v}\text{Tr}(vv^\dagger)\right)^2 = v^\dagger \sigma v \text{ note that } \dim(v^\dagger \sigma v) = 1$$

Also, $\langle vv^\dagger, \sigma \rangle = \text{Tr}(vv^\dagger \sigma) = \text{Tr}(v\sigma v^\dagger) = v^\dagger \sigma v$
Thus, $F(\rho, \sigma) = F(vv^\dagger, \sigma) = \langle vv^\dagger, \sigma \rangle$ □

Lemma 7. *$F(\rho, \rho) = 1$, always*

Proof. Note that $\sqrt{\rho}\rho = \sqrt{\rho}(\sqrt{\rho})^2 = (\sqrt{\rho})^2\sqrt{\rho} = \rho\sqrt{\rho}$.
$\text{Tr}^2(\sqrt{\sqrt{\rho}\rho\sqrt{\rho}}) = \text{Tr}^2(\sqrt{\rho\sqrt{\rho}\sqrt{\rho}}) = \text{Tr}^2(\sqrt{\rho\rho}) = \text{Tr}^2(\rho) = 1$ □

We can also define infidelity to be $1 - F(\rho, \sigma)$. Fidelity is a similarity measure, while infidelity is a distance measure.

Lemma 8 (Fuchs Van de Graat)**.** *Let $\rho, \sigma \in L(S)$ quantum states in a complex Euclidean space S. Then,* $1 - \frac{1}{2}\|\rho - \sigma\|_1 \leq F(\rho, \sigma) \leq \sqrt{1 - \frac{1}{4}\|\rho - \sigma\|_1^2}$

This is equivalent to $1 - D_{tr}(\rho, \sigma) \leq F(\rho, \sigma) \leq \sqrt{1 - D_{tr}^2(\rho, \sigma)}$.
When $\rho = 0$, $1 \leq F(\rho, \sigma) \leq 1$, since $D_{tr}(\rho, \rho) = 0$
If we reverse the inequality to infidelity, we get $1 - \sqrt{1 - D_{tr}^2(\rho, \sigma)} \leq 1 - F(\rho, \sigma) \leq D_{tr}(\rho, \sigma)$. By Taylor expansion of $\sqrt{x}$, we get $1 - F(\rho, \sigma) \in \mathcal{O}(D_{tr}(\rho, \sigma))$ and $1 - F(\rho, \sigma) \in \Omega(D_{tr}^2(\rho, \sigma))$.

Now we also generalize the concept of fidelity to quantum channels.

***Definition:* 4.3: Channel Fidelity**

Let $\rho \in L(S)$ be a quantum state. Fidelity of a quantum channel (CPTP map) $\Phi : L(S) \to L(S)$ is $F(\Phi, \rho) = F(\text{vec}(\sqrt{\rho})\text{vec}(\sqrt{\rho})^\dagger, (\Phi \otimes I)(\text{vec}(\sqrt{\rho})\text{vec}(\sqrt{\rho})^\dagger))$.

4.2 Quantum Source Coding Theorem

***Definition:* 4.4: Quantum Coding Scheme**

Let $X = \mathbb{C}^\Sigma$, $Y = \mathbb{C}^{\{0,1\}^{\otimes m}}$ be complex Euclidean spaces where Σ is an alphabet and let $\rho \in L(X)$ be a quantum state. The quantum coding scheme (Φ, Ψ) is a pair of channels s.t. $\Phi : X^{\otimes n} \to Y^{\otimes m}$ and $\Psi : Y^{\otimes m} \to X^{\otimes n}$. Let $m = \lfloor \alpha n \rfloor$, then (Φ, Ψ) is a (n, α, δ) coding scheme if $F(\Psi\Phi, \rho^{\otimes n}) > 1 - \delta$.

Here α can be viewed as the compression rate, n is the number of sequences.

E.g. The quantum auto-encoder is compressing $X \to Y$ and then decompress $Y \to X$. Quantum Source Coding Theorem imposes a limitation on the error bounds of the machine learning algorithm.

Any $\rho = \sum_v \lambda_v e_v e_v^\dagger = \text{diag}(\lambda_0, \lambda_1, ..., \lambda_{n-1})$ where $\sum_i \lambda_i = 1$ can be viewed as classifcal probability distribution over pure states. $H(\rho) = H(vv^\dagger) = -\text{Tr}(vv^\dagger \log(vv^\dagger)) = 0$.

Theorem: 4.1: Quantum Source Coding Theorem

Let Σ be an alphabet, $S = \mathbb{C}^\Sigma$, $\rho \in L(S)$. Let $\alpha > 0$, $\delta \in (0, 1)$.

1. If $\alpha > H(\rho)$, then there exists an (n, α, δ) coding scheme for all but a finite number of n.
2. If $\alpha < H(\rho)$, then there exists an (n, α, δ) coding scheme for at most a finite number of n.

Proof. **Case 1:** if $\alpha > H(\rho)$.
Then there exists $\epsilon > 0$, s.t. $\alpha > H(\rho) + 2\epsilon$
Assume that $n > \frac{1}{\epsilon}$. This is true for all but finitely many n.
Let $T_{n,\epsilon} \subset \Sigma^n$ for a probability distribution P s.t. we can decompose $\rho = \sum_{a \in \Sigma} P(a) u_a u_a^\dagger$
$T_{n,\epsilon}$ is the $\epsilon-$typical set for the classical probability distribution P.
Define $\Pi_{n,\epsilon} = \sum_{a_1 ... a_n \in T_{n,\epsilon}} a_1 a_1^\dagger \otimes a_2 a_2^\dagger \otimes \cdots \otimes a_n a_n^\dagger$ where each component is a projector onto $\epsilon-$typical set.
$\text{Tr}(\Pi_{n,\epsilon} \rho^{\otimes n}) = \sum_{a_1 ... a_n \in T_{n,\epsilon}} P(a_1) \cdots P(a_n)$
We can enumerate each instance of the $\epsilon-$typical set and map over to it and fail if we don't get an $\epsilon-$typical string.
Define $A_n = \sum_{a_1 ... a_n \in T_{n,\epsilon}} (\underbrace{e_{f_n}(a_1 ... a_n)}_{\text{Encoding of } a_1 ... a_n})(\underbrace{u_{a_1} \otimes \cdot \otimes u_{a_n}}_{\text{eigenvectors of } \rho})^\dagger$
We can now build the encoder $\Phi_n(X) = A_n X A_n^\dagger + \langle I - A_n^\dagger A_n, X \rangle \sigma$, where $X \in L(S^{\otimes n})$, σ can be any state,
and the decoder $\Psi_n(Y) = A_n^\dagger Y A_n + \langle I - A_n A_n^\dagger, Y \rangle \xi$, where $Y \in L(Y^{\otimes n})$, ξ can be any state
Then $(\Psi_n \Phi_n)(X) = (A_n^\dagger A_n) X (A_n^\dagger A_n)^\dagger + \sum_k C_{n_k} X C_{n_k}^\dagger$.
$F(\Psi_n \Phi_n, \rho^{\otimes n}) \geq \langle \rho^{\otimes n}, A_n^\dagger A_n \rangle$, where $A_n^\dagger A_n$ is a projector onto an $\epsilon-$typical space $T_{n,\epsilon}$.
By argument about $\epsilon-$typicality, we have that as $n \to \infty$, the inner product (probability) $\to 1$.

Case 2: if $\alpha < H(\rho)$.
Recall the Kraus representation: $\Phi_n(X) = \sum_{k=1}^{N_1} A_k X A_k^\dagger$, $\Psi_n(X) = \sum_{k=1}^{N_2} B_k X B_k^\dagger$. Note that we can simply append $0X0$ to the smaller term to reach $\max(N_1, N_2)$ terms.
Assume (Φ, Ψ) is an $(n, \alpha, \delta)-$reliable encoding scheme.
$\Psi_n \Phi_n = \sum_{j \geq 1, k \leq N} (B_k A_j) X (B_k A_j)^\dagger$ where $N = \max(N_1, N_2)$
Assume $Y = \{0, 1\}$ (encode using quantum bits), $\dim(B_k A_j) = 2^m$, $\text{rank}(B_j A_j) \leq 2^m$
Consider a projector operation Π_k, with $\Pi_k B_k = B_k$
Fidelity squared $F^2(\Psi_n \Phi_n, \rho^{\otimes n}) = \sum_{jk} |\langle B_k A_j, \rho^{\otimes n} \rangle|^2 = \sum_{jk} \left| \langle B_k A_j \sqrt{\rho^{\otimes n}}, \Pi_k \sqrt{\rho^{\otimes n}} \rangle \right|^2$
$\leq \sum_{jk} \text{Tr}(B_k A_j \rho^{\otimes n} A_j^\dagger B_k^\dagger) \langle \Pi_k, \rho^{\otimes n} \rangle$ by Cauchy Schwarz inequality.
For a projector, $\langle \Pi_k, \rho^{\otimes n} \rangle \leq \sum_i \lambda_i = \sum_{\vec{a}} P(a_1) \cdots P(a_n)$
Also $\sum_{jk} \text{Tr}(B_k A_j \rho^{\otimes n} A_j^\dagger B_k^\dagger) = 1$
So $F^2 \leq \sum_{\vec{a}} P(a_1) \cdots P(a_n)$
As $n \to \infty$, $F^2 \to 0$, because $P(a_i)$ are drawn from typical sequences.
This means that the map is only reliable for at most a finite number of n □

Von Neumann Entropy characterize the minimum bits (information) required to describe a quantum state.

5 Quantum Entanglement

If we have one copy of $|+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$, then once we measure and fine tune a projector, we are sure to always get $|+\rangle$, and the state is easily distinugishable.
However, when we have multiple copies $|+\rangle^{\otimes n}$, we cannot distinuish it with $(|+\rangle + \epsilon v)^{\otimes n}$, where v is a unit vector. This is because $D_{\mathrm{Tr}}(|+\rangle^{\otimes n}, (|+\rangle + \epsilon v)^{\otimes n}) \in \mathcal{O}(n\epsilon)$.
Entanglement gives information about correlation. Define $e_{00} = e_0 \otimes e_0$. Consider the maximally mixed state $\rho = \frac{1}{2}(e_{00}e_{00}^\dagger + e_{00}e_{11}^\dagger + e_{11}e_{00}^\dagger + e_{11}e_{11}^\dagger)$. $P(e_{00}) = \mathrm{Tr}(e_{00}e_{00}^\dagger\rho) = \frac{1}{2}$. Similarly,
$P(e_{11}) = \mathrm{Tr}(e_{11}e_{11}^\dagger\rho) = \frac{1}{2}$.

Suppose Alice has the state ρ. Denote $e_{00} = e_0^1 \otimes e_0^2$. Alice keeps e_0^1 and sents e_0^2 to Bob. If Alice measures her qubit and gets 0, then when Bob measures his qubit, he gets 0 with $P = 1$. The state changes instantly, regardless of the distance between A and B.
Note that no information is sent (faster than speed of light) during measurement. We just observe the correlations and the correlation is stored in the entanglement, rather than a single space.

Bipartite Entanglement is a method to describe the degree of correlation between two quantum systems. If we can apply a unitary basis transformation U to $\Phi\Phi^\dagger$ and get a tensor product such as $e_0e_0^\dagger \otimes e_0e_0^\dagger$, then the two subsystems are independent and have no correlation.

***Definition:* 5.1: Separable Operators**

For any choice of Euclidean spaces X, Y the set $Sep(X : Y)$ contains the set of all positive semidefinite operators $R \in Pos(X \otimes Y)$ for which there exists an alphabet Σ and two sets of of positive semidefinite operators $\{P_a : a \in \Sigma\} \subset Pos(X), \{Q_a : a \in \Sigma\} \subset Pos(Y)$ such that $R = \sum_{a\in\Sigma} P_a \otimes Q_a$. and the elements of the set $Sep(X : Y)$ are called separable operators.

Example: $e_0e_0^\dagger \otimes e_0e_0^\dagger \in Sep(X : Y)$, but $\rho = \frac{1}{2}(e_{00}e_{00}^\dagger + e_{00}e_{11}^\dagger + e_{11}e_{00}^\dagger + e_{11}e_{11}^\dagger) \notin Sep(X : Y)$.

5.1 Notions of Entanglement

***Definition:* 5.2: Entropy of Entanglement**

Let $\rho_{AB} \in L(X \otimes Y)$ be a pure quantum state operator acting on complex Euclidean space $X \otimes Y$. The Entropy of entanglement is defined to be $S = H(\mathrm{Tr}_A(\rho_{AB})) = H(\mathrm{Tr}_B(\rho_{AB}))$

This notion of entanglement describes the information loss in the state that occurs when we throw away one of the two subsystems that the state is supported on. In particular, if ρ_A is the maximally entangled state considered in the above example the entropy of entanglement is maximum, meaning that our knowledge of state is so small after throwing away part of it that we would need the maximum possible amount of information to code samples drawn from the distribution per the quantum source coding theorem.

Example: if $\rho_{AB} = \frac{1}{2}I \otimes \frac{1}{2}I$, then $\rho_B = \mathrm{Tr}_A(\rho_{AB}) = \frac{1}{2}\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$, $S = H(\rho_B) = 1$.

Example: if $\rho = \frac{1}{2}\mathrm{diag}(1, 1)$ a maximally mixed state. Then
$H(\rho) = -\mathrm{Tr}(\mathrm{diag}\left(\frac{1}{2}, \frac{1}{2}\right)\log \mathrm{diag}\left(\frac{1}{2}, \frac{1}{2}\right)) = -\mathrm{Tr}(\mathrm{diag}\left(\frac{1}{2}, \frac{1}{2}\right)\mathrm{diag}\left(-1, -1\right)) = 1$

In classical information, $H(\rho) = 0$ means no uncertainty. In quantum information, if $H(\rho) = 0$ for an entangled state, then we have perfect quantum certainty/knowledge.
Example: $\rho = \frac{1}{2}(e_0e_0^\dagger \otimes e_0e_0^\dagger + e_0e_1^\dagger \otimes e_1e_0^\dagger + e_1e_0^\dagger \otimes e_0e_1^\dagger + e_1e_1^\dagger \otimes e_1e_1^\dagger)$, $H(\rho) = 0$, we have perfect knowledge about the entanglement. If A gets e_0, then B must get e_0. However, $H(\rho_A) = 1$, is a

maximally mixed state after we trace out B.

Definition: 5.3: Schmidt Decomposition

Let $A \in L(X,Y)$ with a SVD $A = \sum_{k=1}^{r} s_k(x_k y_k^\dagger)$, where s_k are singular values, x_k, y_k are corresponding singular vectors. $\text{vec}(A) = \sum_{k=1}^{r} s_k x_k \otimes y_k^T$ is the Schmidt decomposition.
If $r = 1$, then the state is separable. If $r \geq 2$, the state is entangled.

Pitfall: If the Schmidt decomposition is $(1-\epsilon)x_0 \otimes y_0^T + \epsilon x_1 \otimes y_1^T$, the state is classified as entangled even if there is minimal information from $x_1 \otimes y_1^T$ as $\epsilon \to 0$.

5.2 LOCC (Local Operations and Classical Communication)

Suppose A, B are subspaces, we want the channel operations on a state $X \in (A \otimes B)$ to be separable. i.e. $\Phi(\rho_A \otimes \rho_B) = \Phi_A(\rho_A) \otimes \rho_B$.
This makes the bipartite entanglement of two spaces invariant under LOCC operations.

Definition: 5.4: Separable Channels

The class of separable channels for complex Euclidean spaces X, Y, Z, W is $\Theta = C(X \otimes Y : W \otimes Z)$ if and only if there exists an alphabet Σ and collections of operators $\{A_a : a \in \Sigma\} \subset L(X,Z)$, $\{B_a : a \in \Sigma\} \subset L(Y,W)$ such that $\Theta(\rho) = \sum_{a\in\Sigma}(A_a \otimes B_a)\rho(A_a \otimes B_a)^\dagger$.

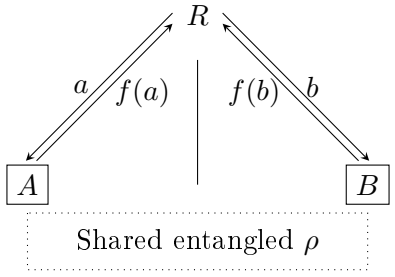
With $\rho = \rho_A \otimes \rho_B$, $\Theta(\rho) = \sum_{a\in\Sigma}(A_a\rho_A A_a^\dagger \otimes B_a \rho_B B_a^\dagger)$. Kraus representation factorizes. Equivalently, channel acts independently on each of the two subspaces that form a larger vector space.

Claim 1. *All sensible metrics of entanglement are non-increasing under LOCC.*

5.3 CHSH Game

CHSH stands for John Clauser, Michael Horne, Abner Shimony, and Richard Holt. In the game, we have:

- A referee R
- Two players A, B, who can only communicate with R.
- R sends signals 00, 01, 10, 11 from uniform distribution randomly
- A, B take the bits a, b seperately, compute $f(a) \in \{0,1\}$, $g(b) \in \{0,1\}$ and return the results to the referee.
- A and B win the game if
  (i) $a = b = 1$ and $f(a) \neq g(b)$
  (ii) $ab = 0$ and $f(a) = g(b)$



Classically, the winning rate is 75%. If they share a quantum entanglement, the winning rate can be $\approx 83\%$.

Theorem: 5.1: Classical winning rate of CHSH game

There exists a deterministic classical strategy that wins CHSH game with $P = \frac{3}{4}$. No deterministic strategy exists that can win with $P > \frac{3}{4}$.

Proof. If A and B constantly output the same value, $f(a) = 0$, $g(b) = 0$ for any a, b, A and B win if $a \neq 1$ and $b \neq 1$, otherwise they lose.
Since R sends 00, 01, 10, 11 uniformly with probability $\frac{1}{4}$, then A, B must win with probability $\frac{3}{4}$
Now we show that P cannot be greater than $\frac{3}{4}$
Assume $\exists f, g$ s.t. A and B win with $P > \frac{3}{4}$. Then A and B must always win. *i.e.* they must win for all 4 cases 00, 01, 10, and 11.
Then we need $f(0) = g(0)$, $f(0) = g(1)$, $f(1) = g(0)$, and $f(1) \neq g(1)$
The first 3 equality gives that $f(1) = g(0) = f(0) = g(1)$. Contradiction with the final inequality.
i.e. winning rate $= 1$ is impossible. And $\frac{3}{4}$ is optimal winning rate. □

Note that no such mixed/probabilistic strategy can hold also because the set of mixed strategies are found by convex combinations of these deterministic strategies and as all deterministic strategies succeed with probability at most $\frac{3}{4}$ we cannot succeed here with probability greater than $\frac{3}{4}$.

Now, we consider the **quantum CHSH games**.

Definition: 5.5: Deterministic Correlation Operator

The operator C is a deterministic correlation operator if $C = \sum_{(a,b)\in\Sigma_a\times\Sigma_b} E_{ab} \otimes E_{f(a)g(b)}$, where E_{xy} refers to the projection matrix with entry 1 at position x, y and f and g are functions. Probabilistic strategies can be formed by convex combinations of the correlation operator.

Definition: 5.6: Quantum Correlation Operator

The operator C is a quantum correlation operator if there exist complex Euclidean spaces X, Y and a state $\rho \in L(X \otimes Y)$ and two collections of measurements $\mu_a : \Gamma_a \mapsto L(X)$, $\nu_b : \Gamma_b \mapsto L(Y)$ such that $C((a,c),(b,d)) = \mathrm{Tr}(\mu_a(c) \otimes \nu_b(d)\rho) = \langle \mu_a(c) \otimes \nu_b(d), \rho \rangle$ for every $a \in \Sigma_a, b \in \Sigma_B, c \in \Gamma_a, d \in \Gamma_b$.

Definition: 5.7: Bell Inequality Violation

Bell inequality violation occurs for an operator K that describes the win/loss strategy for a game if for classical deterministic correlation operator C, $\langle C, K \rangle \leq \alpha$ and if there exists a quantum strategy D, s.t. $\langle D, K \rangle > \alpha$.

For CHSH games, classically $\alpha = 2$. In quantum version, $\langle D, K \rangle = 2\sqrt{2}$.

This violation can make the computation of two physically separated untrusted quantum computers perform trustworthy computation.

Note that $(b_0, b_1) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$ gives the answers. If we get 1, this means that answers from A and B are the same. If we get -1, this means that answers are different.

Define $K = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. The first operator acts on the question space. The second operator acts on the answer space. The -1 flips the sign for $f(a) = f(b) = 1$, which means A and B lose the game.
Choose the deterministic correlation operator $C = E_{00} \otimes E_{00} + E_{01} \otimes E_{00} + E_{10} \otimes E_{00} + E_{11} \otimes E_{00}$ to represent the game, where A and B only output $f(a) = f(b) = 0$ deterministically.
Then, $\langle C, K \rangle = \left\langle \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right\rangle = 2$
i.e. The classical CHSH game has $\langle C, K \rangle = 2$.

Definition: 5.8: Commutators

The communtator of two operators A, B is $[A, B] = AB - BA$. For A, B acting on different spaces, $[A, B] = 0$.

Definition: 5.9: Rotation Projector

$\Pi_\theta = \begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{pmatrix}$ is a rotation projector, with rotation angle θ.

Theorem: 5.2: Quantum Bell Inequality

There exists a quantum strategy that achieves $\alpha = 2\sqrt{2}$, and no quantum strategy exists s.t. $\alpha > 2\sqrt{2}$.

Proof. We can decompose $K = A_0B_0 + A_0B_1 + A_1B_0 - A_1B_1$, where A_i, B_j are reflexive operators $A_i^2 = \mathbb{1}$. A_i gives A's answer and B_i gives B's answer.
$K^2 = (A_0B_0)^2 + (A_0B_0A_0B_1) + \cdots + A_0B_0A_1B_0 + \cdots - A_1B_1A_1B_0 + (A_1B_1)^2$
Since A_i, B_j act on different spaces, $[A_i, B_j] = A_iB_j - B_jA_i = 0$.
This implies that $(A_0B_0)^2 = A_0^2B_0^2 = \mathbb{1}$
$K^2 = 4\mathbb{1} - [A_0, A_1][B_0, B_1]$
Note then that for $\|\cdot\|$ the Schatten ∞-norm (otherwise called the spectral norm or the largest singular value of a matrix),

$$\begin{aligned} \|K^2\| &\leq 4 + \|[A_0, A_1][B_0, B_1]\| \leq 4 + \|[A_0, A_1]\|\|[B_0, B_1]\| \\ &\leq 4 + (\|A_0\|\|A_1\| + \|A_1\|\|A_0\|)(\|B_0\|\|B_1\| + \|B_1\|\|B_0\|) \\ &= 4 + (1+1)(1+1) = 8 \end{aligned}$$

Then $\langle K \rangle^2 \leq \|K^2\| \leq 8$
Thus, $\langle K \rangle \leq 2\sqrt{2}$

Next we show that there exists a quantum strategy that can equal this score of $2\sqrt{2}$.
We specifically take,

$$\begin{aligned} \mu_0(0) &= \Pi_0 & \mu_0(1) &= \Pi_{\pi/2} \\ \mu_1(0) &= \Pi_{\pi/4} & \mu_1(1) &= \Pi_{3\pi/4} \\ \nu_0(0) &= \Pi_{\pi/8} & \nu_0(1) &= \Pi_{5\pi/8} \\ \nu_1(0) &= \Pi_{7\pi/8} & \nu_1(1) &= \Pi_{3\pi/8} \end{aligned}$$

and we take ρ to be the maximally entangled state

$$\rho = \frac{1}{2}(E_{00} + E_{11} + e_0 e_1^\dagger \otimes e_1 e_0^\dagger + e_1 e_0^\dagger \otimes e_0 e_1^\dagger) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

These values are chosen specifically to have the property that

$$\langle \mu_a(c) \otimes \nu_b(d), \rho \rangle = \frac{1}{2}\langle \mu_a(c), \nu_b(d) \rangle.$$

Using this property, we can drop the ρ from the calculation. Filling in the table of elements in the state yields

$$C = \begin{pmatrix} \frac{2+\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} & \frac{2+\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} \\ \frac{2-\sqrt{2}}{8} & \frac{2+\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} \\ \frac{2+\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} & \frac{2+\sqrt{2}}{8} \\ \frac{2-\sqrt{2}}{8} & \frac{2+\sqrt{2}}{8} & \frac{2-\sqrt{2}}{8} & \frac{2+\sqrt{2}}{8} \end{pmatrix}$$

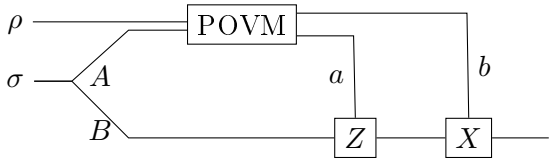
Then plugging this in allows us to verify that this case corresponds to $\langle K, D \rangle = 2\sqrt{2}$. □

6 Superdense coding

6.1 Review of Teleportation

Teleportation gave a protocol for using entanglement for a useful task. It can be used to in some sense pre-cache quantum communication between two parties so that any time after sharing an entangled state Alice and Bob can consume that state to send a quantum state between them using only LOCC (local operations and classical communication).

Given $V_{00} = \frac{1}{\sqrt{2}}(e_{0A} \otimes e_{0B} + e_{1A} \otimes e_{1B})$, $\sigma = V_{00}V_{00}^\dagger$, A holds the state σ_A, B holds σ_B. Now, A sends 2 secret bits constructed by σ_A and some quantum state ρ to B. B can use the 2 bits to correct the error in the shared qubit σ.



Note that 1 quantum bit can only transfer 1 classical bit of information. If we want to share multiple bits, we need to teleport multiple times, but one bit at a time.

6.2 Superdense coding

***Definition:* 6.1: Superdense Coding**

Let $S_{A,0}, S_{B,0}, S_{B,1}$ be complex Euclidean spaces over the alphabet $\{0,1\}$. Further, let $\Lambda : L(S_{A,0}) \to L(S_{B,1})$ be a CPTP map that acts as the identity isomorphism between linear operators acting on the two spaces. Finally let $\Phi : (L(S_{A,0}), \{0,1\}^2) \to L(S_{A,0})$ be for any bit string in $\{0,1\}^2$ a CPTP map such that

$$\Phi(\rho, ab) = Z^a X^b \rho X^b Z^a$$

Let $\{\Pi_{00}, \Pi_{01}, \Pi_{10}, \Pi_{11}\}$ be a four-outcome projective measurement on the space $S_{B,0} \otimes S_{B,1}$ corresponding to the pure state vectors $v_{ab} = \frac{1}{\sqrt{2}}((-1)^{ab} e_0^{1-b} e_1^b \otimes e_0 + (-1)^{a+ab} e_0^b e_1^{1-b} \otimes e_1)$

***Theorem:* 6.1: Superdense Coding Theorem**

Let A have a bit string $h = (a, b)$. The following will send h to B with probability 1.

1. Prepare the state $\rho = V_{00}V_{00}^\dagger$ on $S_{A,0} \otimes S_{B,0}$
2. Apply the transformation $\Phi(\rho, h)$ to encode the hidden string in $S_{A,0}$
3. Apply the Λ channel to ρ to send half the state to B
4. Measure the POVM $\{\Pi_{AB}\}$ on B's qubits and set the measurement result to be h'

Proof.

$$\begin{aligned}
\Phi(V_{00}V_{00}^\dagger) &= (Z^a \otimes I)(X^b \otimes I)V_{00} \\
&= \frac{1}{\sqrt{2}}(Z^a X^b e_0 \otimes e_0 + Z^a X^b e_1 \otimes e_1) \\
&= \frac{1}{\sqrt{2}}((-1)^{ab} e_0^{1-b} e_1^b \otimes e_0 + (-1)^{a+ab} e_0^b e_1^{1-b} \otimes e_1)
\end{aligned}$$

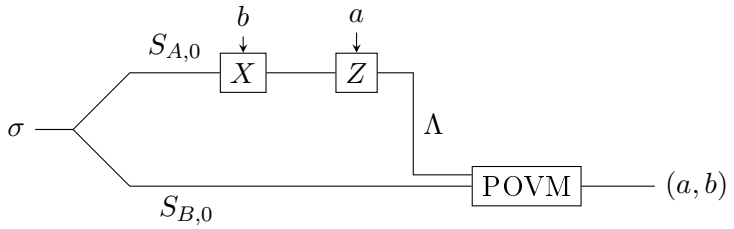
Now if we apply the channel Λ to this, which transfers the qubit in $S_{A,0}$ to $S_{B,1}$.

$$\Lambda(\frac{1}{2}((-1)^{ab}e_0^{1-b}e_1^b \otimes e_0 + (-1)^{a+ab}e_0^b e_1^{1-b} \otimes e_1)((-1)^{ab}e_0^{1-b}e_1^b \otimes e_0 + (-1)^{a+ab}e_0^b e_1^{1-b} \otimes e_1)^\dagger)$$
$$= \frac{1}{2}(e_0 \otimes e_0^{1-b}e_1^b + (-1)^a e_1 \otimes e_1^{1-b}e_0^b)(e_0^{1-b}e_1^b \otimes e_0 + (-1)^a e_1^{1-b}e_0^b \otimes e_1)^\dagger$$
$$= v_{ab}v_{ab}^\dagger.$$

Then B measures a POVM with elements $\Pi = \{V_{ab}V_{ab}^\dagger : (a,b) \in \{0,1\}^2\}$
B gets (a,b) with probability $= \mathrm{Tr}(V_{ab}V_{ab}^\dagger V_{ab}V_{ab}^\dagger) = \mathrm{Tr}(V_{ab}V_{ab}^\dagger) = 1$ □

Note: $V_{00} = \frac{1}{\sqrt{2}}(e_{00} + e_{11})$, $V_{01} = \frac{1}{\sqrt{2}}(e_{00} - e_{11})$, $V_{10} = \frac{1}{\sqrt{2}}(e_{01} + e_{10})$, $V_{11} = \frac{1}{\sqrt{2}}(e_{01} - e_{10})$, and the $V_{ab}V_{ab}^\dagger$ forms the basis for Bell measurements.

Suppose A prepares $\sigma = V_{00}V_{00}^\dagger$ and has a hidden bit stream $h = (a,b)$. A shares σ_B to B first. Then encode $h = (a,b)$ in σ_A, and send σ_A tp B. B measures through POVM and get the 2 bit info h.



If A wants to send 2 classical bits to B, A can send 1 qubit to B first, then decide what classical information $h = (a,b)$ to share, encode (a,b) with the qubit A has, and then send to B. When B measures the qubits, B gets (a,b).

Side Notes: (Entanglement swapping) If A share a bell state σ_1 with B, B share a bell state σ_2 with C. If B performs a joint measure on σ_1 and σ_2. A, C now share an entangled state.

6.3 Accessible Information and Holevo's Theorem

Recall that classical information in quantum format must be diagonal, *e.g.*, $\rho = \frac{1}{2}(e_0e_0^\dagger + e_1e_1^\dagger)$. If A, B communicates classical information through quantum channel Y, what's the mutual information $I(A;B)$?

***Definition:* 6.2: Accessible Information**

Let X, Z be classical registers with states drawn from the alphabets Σ, Γ respectively. Let Y be a quantum register whose state is described by a complex Euclidean space S_Y and let $\mu : \Gamma \to Pos(S_Y)$ be a measurement and let $\eta : \Sigma \to Pos(S_Y)$ be an ensemble of classical states. Let q be a probability distribution on $\Gamma \times \Sigma$ such that $q(a,b) = \langle \mu(b), \eta(a) \rangle = \mathrm{Tr}(\mu(b)\eta(a))$ for any $(a,b) \in \Gamma \times \Sigma$ and let $q[X]$ and $q[Z]$ be the corresponding distributions for registers X and Z. The accessible information, I_{acc}, is defined to be the maximum mutual information over all measurements μ which can be expressed in terms of KL Divergence/relative entropy as

$$I_{\mathrm{acc}} := \sup_\mu I_\mu(\eta) := \sup_\mu D_{KL}(q||q[X] \otimes q[Z])$$

Definition: 6.3: Holevo Information

Let X be a register with states taken over the finite alphabet Σ, and let Y be a quantum register on which we have a complex Euclidean space S_Y with elements drawn from Σ and let $\rho \in L(S_Y)$ be a density operator that can be interpreted as an ensemble of classical states meaning that $\rho = \sum_{i\in\Sigma} P_i\sigma_i$ for non-negative P_i and quantum state operators σ_i such that $\sum_i P_i = 1$ and the total quantum state defined on the classical/quantum space is $\sum_{i\in\Sigma} P_i E_{ii} \otimes \sigma_i$. We then define the Holevo information of ρ to be

$$\begin{aligned}\chi(\rho) = I(X;Y) &= H(\sum_{a\in\Sigma} P_a\sigma_a) + \sum_{a\in\Sigma} P_a H(\sigma_a) \\ &= D(\rho||\rho[X]\otimes\rho[Y])\end{aligned}$$

Theorem: 6.2: Holevo's Theorem

Let $\eta : \Sigma \to Pos(S_Y)$ for alphabet Σ and S_Y a complex Euclidean space. Then $I_{\text{acc}}(\eta) \leq \chi(\eta)$

Proof. Let X be a classical register with state set Σ, S_Y be a complex Euclidean space, and $\sigma \in L(S_X \otimes S_Y)$ be a quantum state s.t. $\sigma = \sum_{a\in\Sigma} E_{aa} \otimes \eta(a)$

$\chi(\eta) = D(\sigma||\sigma[X]\otimes\sigma[Y])$
$\Phi(\rho \in S_Y) = \sum_{b\in\Gamma}\langle\mu(b),\rho\rangle E_{bb}$.
Note that while we have used quantum language for the register the output can be interpretted as an ensemble over the alphabet Γ and so it can be thought of as a classical distribution.

Next consider applying the channel to the quantum information in the state σ,

$(I_X \otimes \Phi)(\sigma) = \sum_{a\in\Sigma}\sum_{b\in\Gamma}\langle\mu(b),\eta(a)\rangle E_{bb} = \text{diag}(q)$,
where q is the probability distribution defined above where $q(a,b) = \langle\mu(b),\eta(a)\rangle$.

The accessible information is again defined as the maximization of the mutual information over all such measurements,
$\forall\mu$, $I_\mu(\eta) = D_{KL}(q||q[X]\otimes q[Y]) = D((I_X\otimes\Phi)(\sigma)||(I_X\otimes\Phi)(\sigma[x]\otimes\sigma[Y]))$
Note that the quantum relative entropy D is non increasing under CPTP maps. Thus,
$T_\mu \leq D(\sigma||\sigma[X]\otimes\sigma[Y]) = \chi(eta)$ □

Corollary 1. *Let Σ be an alphabet, Y be a complex Euclidean space and $\eta : \Sigma \to Pos(Y)$ be an ensemble of states. Then $I_{acc}(\eta) \leq \log(\dim(Y))$*

Example: For superdense coding, the dimension $\dim(Y) = 4$, so $I_{\text{acc}} = 2$ and 2 qubits are needed.

Note: Quantum machine learning is just a constant factor better than classical machine learning based on Holevo theorem.

7 Channel Capacity

Suppose there exists a channel $\Phi^{\otimes n}$ between A and B. How many bits of information can A send to B per use of Φ? *i.e.* What is $\sup_{n,\rho\in L(S_X^{\otimes}n\otimes S_A)} \left[\frac{1}{n}\text{Info}\left[\left(\Phi^{\otimes n}\otimes I_A\right)(\rho)\right]\right]$?

We can have entanglement between $S_X^{\otimes}n$ and the ancillary space S_A in state ρ. Also, we can have arbitrary encoder Ξ_E and decoder Ξ_D, and build a *emulation* channel $\Psi = \Xi_D(\Phi^{\otimes n}\otimes I_A)\Xi_E$.

Taxonomy of quantum channel capacities

- Classical capacity (classical bits, classical channels)
- Classical capacity of a quantum channel (classical in, quantum channel, classical out)
- Entanglement assisted classical capacity
- Holevo capacity
- Quantum channel capacity (qubits in, quantum channel, qubits out)

***Definition:* 7.1: Channel Norm and Channel Distance**

Given two channels Φ and Ψ, we can define the distance to be

$$\sup_\rho \|\Phi(\rho)-\Psi(\rho)\|_1 = \sup_\rho 2D_{tr}(\Phi(\rho),\Psi(\rho)).$$

Diamond Norm: $\|\Phi\|_\diamond = \sup_{\rho\in S_X\otimes S_A, S_A} \|(\Phi\otimes I_A)(\rho)\|_1$, S_A is an ancillary (garbage) space.

Diamond Distance: $d_\diamond(\Phi,\Psi) = \frac{1}{2}\|\Phi-\Psi\|_\diamond$.

Recall that the Holevo Helstrom bound, the optimal probability of distinguishing two quantum states $\geq \frac{1}{2}+\frac{1}{2}D_{tr}(\rho,\sigma)$
Similarly, the diamond distance gives the optimal probability of distinguishing two quantum channels Φ and Ψ is $\geq \frac{1}{2}+\frac{1}{4^c}\|\Phi-\Psi\|_\diamond$.

We consider the following two channel examples:

***Definition:* 7.2: Quantum Erasure Channel**

Let $X=\{e_0,e_1\}$ and $Y=\{e_0,e_1,e_2\}$ and let S_X and S_Y be complex Euclidean spaces over symbols from these alphabets. Let the channel $\Phi: L(S_X)\to L(S_Y)$ be a CPTP map such that for $a\in[0,1]$ and any state $\rho\in S_X$ we have that $\Phi(\rho)=(1-a)\rho + a\text{Tr}(\rho)e_2e_2^\dagger$.

As A sends a state ρ through Φ to B, B gets either ρ with probability $1-a$ or $e_2e_2^\dagger$ with probability a.
In the erasure channel, we know when there is an error in the channel transmission, because we get a completely different output.
All quantum channel capacities are known for the erasure channel.

Definition: 7.3: Depolarizing Channel

Let S_X be a complex Euclidean space. Let the channel $\Phi : L(S_X) \to L(S_X)$ be a CPTP map such that $\Phi(\rho) = (1-a)\rho + a\mathrm{Tr}(\rho)\frac{I}{2}$.

The depolarizing channel either sends the information correctly, or mixes the state completely.
We don't know the quantum channel capacities for the depolarizing channel.

Definition: 7.4: Channel Approximation

We say quantum channel Φ is an ϵ-approximation to Ψ if $\exists \epsilon \geq 0$, $\|\Phi - \Psi\|_\diamond \leq \epsilon$.

Definition: 7.5: Completely Dephasing Channel

Let S_X and S_Y be complex Euclidean spaces, $\Phi : L(S_X) \to L(S_Y)$ be a quantum channel. Let $\Gamma = \{0,1\}$ and S_Z be a complex Euclidean space on $\mathbb{C}^\Gamma$. The completely dephasing channel is:

$$\Delta : \rho \in L(S_Z) \to \frac{1}{2c}\int_0^{2\pi} (e^{-i\theta e_0 e_0^\dagger}\rho e^{i\theta e_0 e_0^\dagger} + e^{-i\theta e_1 e_1^\dagger}\rho e^{i\theta e_1 e_1^\dagger})d\theta,$$

where c is some constant.

$\Delta(e_0 e_0^\dagger) = \frac{1}{2c}\int_0^{2\pi}(e^{-i\theta e_0 e_0^\dagger} e_0 e_0^\dagger e^{i\theta e_0 e_0^\dagger} + e^{-i\theta e_1 e_1^\dagger} e_0 e_0^\dagger e^{i\theta e_1 e_1^\dagger})d\theta = \frac{\pi}{c} e_0 e_0^\dagger$
$\Delta(e_1 e_1^\dagger) = \frac{\pi}{c} e_1 e_1^\dagger$
$\Delta(e_0 e_1^\dagger) = \Delta(e_1 e_0^\dagger) = \int_0^{2\pi} e^{-i\theta} d\theta e_0 e_1^\dagger = 0$

Δ removes the off-diagonal elements in a quantum state ρ, which converts ρ to an ensemble of classical states. But Δ is a perfect map for $e_0 e_0^\dagger$ and $e_1 e_1^\dagger$.

Definition: 7.6: Achievable Rate

The rate $\alpha \geq 0$ for the channel Φ is said to be an achievable rate for classical information transfer if
1. $\alpha = 0$ (completely block the channel)
2. or $\alpha > 0$ and for all but a finite number of n, $\Phi \otimes n$ emulates an ϵ-approximation to $\Delta^{\otimes \lfloor \alpha n \rfloor}$. Φ emulates Ψ if $\exists \Xi_D, \Xi_E$ s.t. $\Psi = \Xi_D \Phi \Xi_E$.

Definition: 7.7: Classical Capacity

The classical capacity of a channel Φ, $C(\Phi)$, is the supremum over α.

Theorem: 7.1: Multi-Channel Capacity

For any positive integer k, $C(\Phi^{\otimes k}) = kC(\Phi)$.

Proof. Assume α is achievable, then $k\alpha$ is achievable with k copies of the channel acting on parallel inputs. If $\alpha > 0$, $\Phi^{\otimes n} = \Phi^{\otimes k\lfloor n/k \rfloor}$ emulates an ϵ-approximation to $\Delta^{\otimes \alpha \lfloor \alpha \lfloor n/k \rfloor \rfloor}$.
For any $n \geq k$, the channel is trivially emulated by $\Phi^{\otimes k\lfloor n/k \rfloor}$ and for $\delta \in (0, \alpha/k)$, $\alpha\lfloor n/k \rfloor \geq (\alpha/k - \delta)$ for all but finitely many n.
Thus, for any $\epsilon > 0$ and all but finitely many n, $\Phi^{\otimes n}$ emulates an ϵ-approximation to $\Delta^{\otimes m}$ for $m = \lfloor (\alpha/k - \delta)n \rfloor$. Therefore, rate is achievable for all $\alpha > 0$.

For $\alpha = 0$, it is totally achievable, so rate α/k is achievable.

Then, $C(\Phi) \geq \frac{1}{k} C(\Phi^{\otimes k})$, *i.e.* $kC(\Phi) \geq C(\Phi^{\otimes k})$.

And therefore, $kC(\Phi) = C(\Phi^{\otimes k})$ □

The rate α is dependent on n, the number of copies of channels. $\alpha(n) = \frac{\#\text{bits transmitted}}{n}$. What we are interested in is the asymptotic behavior $\lim_{n\to\infty} \alpha(n)$. Since number of bits transmitted is $\mathcal{O}(n)$ and is monotonically increasing, the limit must exist, but may be hard to compute.

Theorem: 7.2: Classical Capacity of Quantum Erasure Channel

The classical capacity of the quantum erasure channel is $C(\Phi) = 1 - a$

Proof. The fraction of states that is not erased is $1 - a$. From Holevo's theorem that each qubit received carries at most 1 bit of information, then $C(\Phi) \leq 1 - a$
To show that the maximum is achievable, we design a protocol.
Consider the completely dephasing channel which brings a quantum state into a classical state by removing the off-diagonal elements
Suppose A applies Δ after encoding the input and B applies Δ after getting it and measures using codewords e_0, e_1.
On average $\frac{1}{1-a}$ attempts will be needed before a success is attained (mean of a geometric r.v.).
Then Φ^n can emulate an ϵ-approximation to the channel $\Delta^{\lfloor (1-a)n \rfloor}$
$\alpha = \lim_{n\to\infty} \frac{\lfloor (1-a)n \rfloor}{n} = 1 - a.$ □

The quantum capacity of the quantum erasure channel is also $C(\Phi) = 1 - a$. However, in general, the classical capacity is a lower bound for quantum capacity.

Definition: 7.8: Holevo Capacity

The Holevo Capacity of a channel Φ is defined as $\chi(\Phi) = \sup_{\eta} \chi(\Phi(\eta))$, where η is a quantum state in the domain of Φ, and $\chi(\eta)$ is the Holevo information of η.

Theorem: 7.3: Holevo-Schumacher-Westmorelan

$$C(\Phi) = \lim_{n\to\infty} \frac{\chi(\Phi^{\otimes n})}{n} \geq \chi(\Phi)$$

This provides a lower bound for channel capacity.

8 Quantum Cryptography

In classical cryptography, we make computational assumptions. For example, RSA assumes that factoring integers is hard. *i.e.* There is no polynomial time algorithm exists for factoring. To retrieve the key from cipher, the cost $>$ poly(n). With an increase in number of bits (*e.g.* from 2048 to 4096), the difficulty of breaking increases exponentially. The second example is the ECC (Elliptic Curve Cryptography), which computes discrete logarithms on elliptic curves.

However, if the attacker gets a scalable quantum computer, the problems can be solved in polynomial time (*e.g.* the Shor's algorithm), Although less efficient schemes based on lattice based cryptography remain immune to all known quantum attacks.

This raises a question about whether quantum information is solely a weapon that can be used to thwart secure communication. We will see here that in fact it isn't. Quantum cryptography offers tools that enable communication between two parties whose security is guaranteed by the laws of quantum mechanics and even an adversary with unbounded computational power is unable to crack such a code provided that the laws of quantum mechanics are correct (and the protocol is perfectly implemented).

***Definition:* 8.1: One-Time-Pad**

One-Time-Pad is a code such that A and B wish to encode $V \in \{0,1\}^*$ and share a key $K \in \{0,1\}^{\text{len}(V)}$. Encoding scheme: $f_E(V,K) = V \oplus K$; Decoding scheme: $f_D(V,K) = f_E(V,K) = V \oplus K$.

The intuition behind the same encoding/decoding function is:

$$V \xrightarrow[A]{} V \oplus K \xrightarrow[B]{} (V \oplus K) \oplus K = V$$

This requires the key to have the same length as the message they want to encode, but it is relatively secure.

The OTP is one time because the password must be different for each transmission. Otherwise, $K \oplus W$ and $K \oplus V$ can give much information $(K \oplus V) \oplus (K \oplus W) = V \oplus W$. However, no one can reliably guess the message.

***Theorem:* 8.1: Probability of Random Guessing OTP**

Assume that A and B wish to communicate $V \in \{0,1\}^L$ or $V \in \{0,1\}^L$ and that they share a secret key K chosen uniformly at random. If V and W are chosen with uniform probability, then there does not exist an unbiased estimator of the message that will succeed in deciding the identity of the message with probability greater than 50%.

Proof. This is a proof sketch
Assume WLOG A sends V, then the cipher text is $f_E(V,K) = V \oplus K$.
The adversarial party gets a copy of $V \oplus K$. $P(K) = P(W \oplus V \oplus K)$.
$f_D(V \oplus K, W \oplus V \oplus K) = W$
The likelihood ratio is $\frac{P(W|W \oplus K)}{V|V \oplus K} = 1$.
Assume that the adversarial party has a strategy that will correctly identify V with probability greater than 50%. If true, then it must misidentify W with probability greater than 50%.
Thus the estimator is biased and so no unbiased estimator with $P > 50\%$ is possible. □

8.1 Quantum Key Distribution (QKD)

The Quantum Key Distribution (QKD) gives a secure protocol for A and B to grow key between them given an untrusted quantum channel. If the channel is authenticated, A and B can identify whether the message has been changed. Through QKD, the two parties either share a random key between two parties (given that you can authenticate the person you are talking to is the intended person) or determine that an eavesdropper is closely monitoring the system and you cannot securely communicate.

The best an adversarial party can do has two possibilities:

1. Jam the communication
2. Not jam the communication, but learn 2^{-m} bits of the information with m predefined by A and B.

Lemma 9. *Using only classical information, A and B cannot securely share random bits.*

Proof. The attacker uses an intercept-resend attack. They measure the classical codeword, stores a copy and resends it to B.

$$I(A:B) = I(E:B)$$

Further the decoding information is sent over the public channel, so the attacker can copy the decoding. □

Bennett and Brassard, the creators of the quantum teleportation protocol, developed the BB84 protocol. Here we consider the simplified BB84 protocol.

***Definition:* 8.2: Simplified BB84 Protocol**

1. A and B talk over the classical channel and agree on a number of qubits to send to each other, L.
2. For each L_i, A picks a codeword uniformly from $\{E_0, E_1, E_+, E_-\}$ and sends to B using Eve's quantum channel Λ_E, where $E_+ = \frac{1}{2}(e_0+e_1)(e_0+e_1)^\dagger$, $E_- = \frac{1}{2}(e_0-e_1)(e_0-e_1)^\dagger$. The states are not perfectly distinguishable.
3. B measures the state he receives $\Lambda_E(\rho)$ using the elements $\{E_0/2, E_1/2, E_+/2, E_-/2\}$ with outcomes 00, 01, 10, 11. Bit value recorded is the second bit, *i.e.* 0 for E_0 and E_+, 1 for E_1 and E_-.
4. A and B announce which of the qubits were prepared in E_0, E_1 and which were prepared in E_+, E_- and discard any that do not match.
5. Fixed a fraction of these basis reconciled qubits to measure and publicly compare their results over the authenticated channel.
6. If there are discrepancies, then there is some attacker watching.
7. If there are more than L^* discrepancies, they abort the protocol.
8. If they notice fewer than L^* discrepancies then they apply an agreed upon cryptographic hash function to reduce the attacker's residual knowledge of the shared bit strings to an exponentially small level (2^{-m} bits).

Example: A: I sent a $+/-$ state. B: I measured $-$. A: I actually sent $+$. An error happened. There could be eavesdropper.

Information Disturbance principle: In quantum, generically information comes with a disturbance to the state.

8.2 Entanglement-Based QKD

The proof of security for BB84 shows that BB84 is equivalent to Entanglement-Based QKD after quantum correction ideas are applied. We can directly prove security for Entanglement-Based QKD.

We want to come up with a scheme that allows A and B to share a bell state $\rho = \frac{1}{2}(e_0 \otimes e_0 + e_1 \otimes e_1)(e_0 \otimes e_0 + e_1 \otimes e_1)^\dagger$. If A measures 0, then B also measures 0. Same for measurement of 1. Also, if A and B share perfect entanglement states as random bit key. E can learn nothing about the bit shared. $\mathrm{Tr}_E(\rho) = (\Phi_{00})_{AB}$.

Theorem: 8.2: Accessible Information of Entanglement-Based QKD

Let S_A, S_B, S_E be finite dimensional complex Euclidean space. Assume they share a state $\rho_{ABE} \in L(S_A \otimes S_B \otimes S_E)$ s.t. $\rho_{ABE} = \rho_{AB} \otimes \rho_E$. The accessible information of E about A is $I(A:E) \leq H(\rho_{AB})$.

i.e. If ρ_{AB} is pure, then the Von-Neumann Entropy $H(\rho_{AB}) = 0$, and E can learn 0 bit (nothing) about A.

Proof. Assume WLOG, ρ_E is pure. Since if ρ_E is mixed, we can dilate it to a pure state by extending the dimension.

Theorem 6.2 tells us that $I(A:E) \leq H(\rho_{ABE}) - \sum_i P_i H(\rho_i)$ where $\rho_{ABE} = \sum_i P_i \rho_i$ for probability $P_i > 0$ and density operator ρ_i.

Since $H(\rho_i) \geq 0$, $I(A:E) \leq H(\rho_{ABE})$.

By Subadditivity property of H, $H(\rho_{ABE}) \leq H(\rho_{AB}) + H(\rho_E) = H(\rho_{AB})$, since $H(\rho_E) = 0$ as a pure state.

Thus, $I(A:E) \leq H(\rho_{AB})$.

Information is measured in bits, $I(A:E) \geq 0$. If ρ_{AB} is pure, then $0 \leq I(A:E) \leq 0$, which implies that $I(A:E) = 0$. □

8.3 Entanglement Distillation

Let $\Phi_{00} = \frac{1}{2}(e_0 \otimes e_0 + e_1 \otimes e_1)(e_0 \otimes e_0 + e_1 \otimes e_1)^\dagger$, $\Phi_{01} = \frac{1}{2}(e_0 \otimes e_0 - e_1 \otimes e_1)(e_0 \otimes e_0 - e_1 \otimes e_1)^\dagger$, $\Phi_{10} = \frac{1}{2}(e_0 \otimes e_1 + e_1 \otimes e_0)(e_0 \otimes e_1 + e_1 \otimes e_0)^\dagger$, $\Phi_{11} = \frac{1}{2}(e_0 \otimes e_1 - e_1 \otimes e_0)(e_1 \otimes e_0 - e_0 \otimes e_1)^\dagger$. These are the bell states.

Note: $\Lambda(\rho_B) = XZ\rho_B ZX$ converts Φ_{11} to Φ_{00}.

Twirling process: $\Lambda_{\text{twirl}}(\rho_A \otimes \rho_B) = \int_{Haar} U\rho_A U^\dagger \otimes U^* \rho_B U^{*\dagger} dU$, where the Haar integral gives the uniform probability distribution on unitary operators. U^* is the complex conjugate of U.

1. $\Lambda_{\text{twirl}}(\Phi_{11}) = \Phi_{11}$
2. $\Lambda_{\text{twirl}}(\Phi_{ij}) = \frac{1}{3}(\Phi_{00} + \Phi_{01} + \Phi_{10})$ for $(i,j) \neq (1,1)$

Theorem: 8.3: Entanglement Distillation

Let ρ_{AB} be a quantum state $\rho_{AB} \in L(S_A \otimes S_B)$ s.t. $F = \text{Tr}(\rho_{AB}\Phi_{11})$ is the fidelity of ρ_{AB} of the ideal state Φ_{11}.

1. A and B share 2^m copies of ρ_{AB} between them and repeat the following protocol to each of the 2^{m-1} pairs of qubits.
2. A and B publicly agree on 2^m random single qubit channels U_i to apply to qubit i. *e.g.* U_1 applies to A's first qubit and B's first qubit.
3. Apply XOR channels (CNOT gates) to their qubits. $\Lambda_{\text{XOR}}(e_i \otimes e_j) = e_i \otimes e_{i\oplus j}$ in a pairwise fashion. *i.e.* between qubit 1&2, qubit 3&4.
4. Measure the target (even number) qubits and communicate result.
5. If they measure the same, they keep the odd number qubits. Otherwise they discard both qubits.
6. A and B randomly apply a single qubit Unitary channel to the state.

The above protocol yields at most 2^{m-1} states that have fidelity F' which obeys

$$F' = \frac{F^2 + (1-F)^2/9}{F^2 + 2F(1-F)/3 + 5(1-F)^2/9}$$

Remark 1. The infidelity $1 - F' = \frac{2}{3}(1-F) + \mathcal{O}((1-F)^2)$. If we repeat, the infidelity decreases exponentially, thus the fidelity increases exponentially.

Proof. Φ_{11} is invariant under Haar twirling. Unitary channel is just basis transformation:

1. $\Lambda_U(e_0) = \cos(\theta)e_0 + \sin(\theta)e^{i\phi}e_1$
2. $\Lambda_U(e_1) = \cos(\theta)e_1 - \sin(\theta)e^{i\phi}e_0$
3. $\Lambda_U \otimes \Lambda_U(e_0 \otimes e_1) = \cos^2(\theta)e_0 \otimes e_1 - \sin^2(\theta)e_1 \otimes e_0 + e^{-i\phi}\cos(\theta)\sin(\theta)e_0 \otimes e_0 - e^{i\phi}\cos(\theta)\sin(\theta)e_1 \otimes e_1$
4. $\Lambda_U \otimes \Lambda_U(e_1 \otimes e_0) = \cos^2(\theta)e_1 \otimes e_0 - \sin^2(\theta)e_0 \otimes e_1 + e^{-i\phi}\cos(\theta)\sin(\theta)e_0 \otimes e_0 - e^{i\phi}\cos(\theta)\sin(\theta)e_1 \otimes e_1$

Then $\Lambda_U \otimes \Lambda_U$ maps $(e_0 \otimes e_1 - e_1 \otimes e_0)$ to
$(\cos^2(\theta) + \sin^2(\theta))e_0 \otimes e_1 - (\cos^2(\theta) + \sin^2(\theta))e_1 \otimes e_0 = e_0 \otimes e_1 - e_1 \otimes e_0$. *i.e.* $\Lambda_U \otimes \Lambda_U(\Phi_{11}) = \Phi_{11}$.

Because Haar is unitary invariant, $\int_{Haar} \Lambda_U \otimes \Lambda_U(\Phi_{11}) = \Phi_{11}$.

Then after Haar operation, ρ is mapped to $F\Phi_{11} + \frac{1}{3}(1-F)\sum_{(i,j)\neq(1,1)}\Phi_{ij}$ and $\rho \otimes \rho$ is mapped to $F^2\Phi_{11}^{\otimes 2} + \frac{1}{9}(1-F)^2\sum_{(i,j)\neq(1,1)}\Phi_{ij}^{\otimes 2}$.

Apply the XOR channel and get $\rho' = \frac{F^2\Phi_{11}^{\otimes 2} + \frac{1}{9}(1-F)^2\sum_{(i,j)\neq(1,1)}\Phi_{ij}^{\otimes 2}}{\text{Tr}(\rho')}$. This gives the final state fidelity. □

The problem is then deciding how many iterations are needed to ascertain whether the state has been distilled sufficiently. There are multiple ways that we can do this, but an easy way to perform this is by modifying the protocol to test to see what the error is. One way to do so is to use a procedure known as quantum state tomography. *Quantum state tomography* simply aims to find a reconstruction of a quantum state via measurement of local operators. The simplest way to achieve this is via Pauli matrices. Note that the Pauli matrices I, X, Y, Z form an orthonormal and complete operator basis for 1 qubit and $\{P_i^{\otimes m}\}$ forms a basis for m qubits. Specifically, $\rho = \sum_{ij}\frac{\text{Tr}(\rho_{AB}P_i \otimes P_j)}{4}$, where P_i are the Pauli matrices.

Thus all that A and B need to do to recover the state is to measure a POVM consisting of the $+1/-1$ eigenvector of both of these operators and from these expectation values A and B can identify the state. The only problem involves deciding which state they reconstruct. If E knew which of the states was being

used to address the fidelity, then an optimal strategy would be to distribute all of her measurements on the states that are being used to assess the fidelity. In order to deal with this, the natural approach is to estimate the fidelity using a subset of the overall quantum states.

The idea to address this is to divide the data up randomly into two parts. The first is the part that is actually used to teleport data. The second is used only to perform tomography and learn what the resultant state is. If a state with fidelity $< \frac{1}{2}$ is obtained then the above protocol will not succeed in boosting the fidelity through distillation and the result will be a failure. If the fidelity is greater than $\frac{1}{2}$ then we can always boost the probability of success and asymptotically at most a logarithmic number of repetitions are needed in order to reduce the infidelity to 2^{-m}. Thus we can achieve unconditional security with these protocols.