Proof: 1) if $R > S(\rho)$ choose $\varepsilon > 0$ such that $R \geq S(\rho) + \varepsilon$.

For $\eta > 0$ choose a typical subspace $T^{(n)}_{\varepsilon,\eta} \subseteq Sp(\rho)^{\otimes n}$ such that

$$\#\left(T^{(n)}_{\varepsilon,\eta}\right) \leq 2^{n(S(\rho)+\varepsilon)} \leq 2^{nR}.$$

Use the compression-decompression scheme introduced before the Theorem. Then

$$F_n \geq 2\sum_\alpha d^{(n)}_\alpha \mu^2_\alpha - 1 = 2\sum_\alpha d^{(n)}_\alpha \langle \psi^{(n)}_\alpha | \tilde{P}^{(n)}_{\varepsilon,\eta} | \psi^{(n)}_\alpha \rangle - 1$$

$$= 2\,\mathrm{Tr}\left(\rho^{(n)} \tilde{P}^{(n)}_{\varepsilon,\eta}\right) - 1 = 2\,\mathrm{Prob}\left(T^{(n)}_{\varepsilon,\eta}\right) - 1 \geq 1 - 2\eta$$

2) if $R < S(\rho)$ choose $\varepsilon > 0$ so that $R < S(\rho) - \varepsilon$.

We may assume that $\rho^{(n)} \in M_{2^n}(\mathbb{C})$ and that

$\mathcal{E}^{(n)} : M_{2^n}(\mathbb{C}) \longrightarrow M_{d_{c(n)}}(\mathbb{C})$, $d_{c(n)} = 2^{nR}$

with $\mathbb{C}^{d_{c(n)}} \subseteq \mathbb{C}^{2^n}$. Set $P^{(n)} : \mathbb{C}^{2^n} \longrightarrow \mathbb{C}^{d_{c(n)}}$, $\tilde{\rho}^{(n)}_\alpha = \mathcal{E}^{(n)}[|\psi^{(n)}_\alpha\rangle\langle\psi^{(n)}_\alpha|]$

Otherwise, $\boxed{\mathcal{E}^{(n)}}$ and $\boxed{\mathcal{D}^{(n)}}$ are arbitrary.

Notice $\boxed{\mathrm{Tr}\left(P^{(n)}\right) = d_{c(n)} = 2^{nR}}$

$$F_n = \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] | \psi_\alpha^{(n)} \rangle$$

$$= \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{P}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \tilde{P}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle \qquad (I)$$

$$+ \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{Q}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \tilde{Q}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle \qquad (II)$$

$$+ \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{Q}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \tilde{P}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle \qquad (III)$$

$$+ \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{P}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \tilde{Q}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle \qquad (IV)$$

where $\boxed{\tilde{Q}_{\varepsilon,\eta}^{(n)} := 1 - \tilde{P}_{\varepsilon,\eta}^{(n)}}$

(I): $\tilde{\rho}_\alpha^{(n)} \leq p^{(n)}$ (Prove it) $\Rightarrow \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \leq \mathcal{D}^{(n)}[p^{(n)}]$

$$(I) \leq \mathrm{Tr}\left(\rho^{(n)} \tilde{P}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[p^{(n)}] \tilde{P}_{\varepsilon,\eta}^{(n)} \right) = \sum_{i^{(n)} : z_{i^{(n)}} \in I_{\varepsilon,\eta}^{(n)}} \pi_{i^{(n)}} \langle z_{i^{(n)}} | \mathcal{D}^{(n)}[p^{(n)}] | z_{i^{(n)}} \rangle$$

$$\leq 2^{-n(S(\rho)-\varepsilon)} \sum_{i^{(n)} : z_{i^{(n)}} \in T_{\varepsilon,\eta}^{(n)}} \langle z_{i^{(n)}} | \mathcal{D}^{(n)}[p^{(n)}] | z_{i^{(n)}} \rangle \leq 2^{-n(S(\rho)-\varepsilon)} \boxed{\mathrm{Tr}(p^{(n)})}$$

$$\leq 2^{-n(S(\rho)-\varepsilon-R)} \xrightarrow[n]{} 0$$

(II): $\mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \leq 1$ (prove it) $\Rightarrow$ (II) $\leq \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{Q}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle$

$$= \mathrm{Tr}\left(\rho^{(n)} \tilde{Q}_{\varepsilon,\eta}^{(n)}\right)$$

$$= 1 - \mathrm{Tr}\left(\rho^{(n)} \tilde{P}_{\varepsilon,\eta}^{(n)}\right) \leq \eta$$

(III): use the Cauchy-Schwartz inequality for the Hilbert-Schmidt scalar product, with $Y \geq 0$,

$$\left|\mathrm{Tr}(\rho X^\dagger Y Z)\right|^2 = \left|\mathrm{Tr}(\sqrt{\rho} X^\dagger \sqrt{Y}\sqrt{Y} Z \sqrt{\rho})\right|^2$$

$$= \left|\mathrm{Tr}\left((\sqrt{Y} X \sqrt{\rho})^\dagger \sqrt{Y} Z \sqrt{\rho}\right)\right|^2$$

$$\leq \mathrm{Tr}\left((\sqrt{Y} X \sqrt{\rho})^\dagger (\sqrt{Y} X \sqrt{\rho})\right) \mathrm{Tr}\left((\sqrt{Y} Z \sqrt{\rho})^\dagger \sqrt{Y} Z \sqrt{\rho}\right)$$

$$= \mathrm{Tr}\left(\rho X^\dagger Y X\right) \mathrm{Tr}\left(\rho Z^\dagger Y Z\right)$$

$$\left|(\mathrm{III})\right|^2 \leq \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{Q}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \tilde{Q}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle$$

$$\times \sum_\alpha d_\alpha^{(n)} \langle \psi_\alpha^{(n)} | \tilde{P}_{\varepsilon,\eta}^{(n)} \mathcal{D}^{(n)}[\tilde{\rho}_\alpha^{(n)}] \tilde{P}_{\varepsilon,\eta}^{(n)} | \psi_\alpha^{(n)} \rangle$$

$$\mathcal{D}^{(n)}\left[\vec{\rho}_{\alpha}^{(n)}\right] \leq 1 \Rightarrow |(III)|^2 \leq \mathrm{Tr}\left(\rho^{(n)} \tilde{Q}_{\varepsilon,\eta}^{(n)}\right) \mathrm{Tr}\left(\rho^{(n)} \tilde{P}_{\varepsilon,\eta}^{(n)}\right)$$

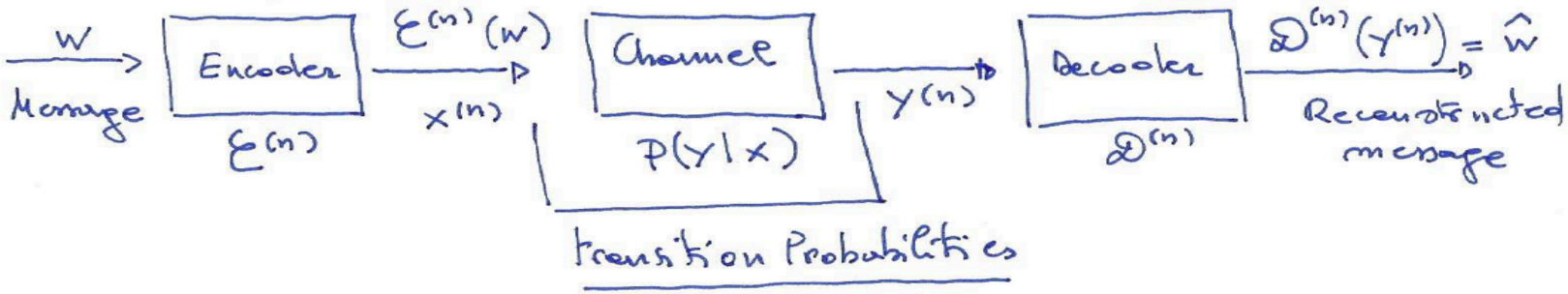
$$\leq \left(1 - \mathrm{Prob}\left(T_{\varepsilon,\eta}^{(n)}\right)\right) \leq \eta$$

(IV) : IV is the hermitian conjugate of III.

Remark : There exists a quantum extension of the Shannon - Mc Millan - Breiman theorem to ergodic quantum sources, where the optimum compression rate is the entropy rate of the source (Bjelaković, Krüger, Seigmund-Schultze, Szkoła).

Entropy rate : $s = \lim_n \frac{1}{n} S(\rho^{(n)})$ (Prove the existence).

2.3 Shannon noisy coding theorem : Channel Capacity.



Definition 2.3.1

A discrete <u>memoryless channel</u> denoted by $(\mathcal{X}, p(y,x), \mathcal{Y})$ consists of an input alphabet $\mathcal{X}$ of finite cardinality $\#(\mathcal{X})$ and an output alphabet $\mathcal{Y}$ also of finite cardinality $\#(\mathcal{Y})$ together with a collection of transition probabilities

$$p(y|x) \geq 0 \quad x \in \mathcal{X}, y \in \mathcal{Y} \text{ such that } \sum_{y \in \mathcal{Y}} p(y|x) = 1$$

$$p(y^{(n)} | x^{(n)}) = \prod_{i=1}^{n} p(y_i | x_i) \; ; \; y^{(n)} \ni y^{(n)} = y_1 y_2 \cdots y_n \, , \quad x^{(n)} = x_1 x_2 \cdots x_n \in \mathcal{X}^{(n)}$$

Remark : - the transition probabilities $p(y|x)$ refer to the possibility of getting y as output upon sending x as input to the channel.

- The extension of the transition probabilities to words of length n might in general be of the form

$$p(y^{(n)} | x^{(n)}; y^{(n-1)}) = \prod_{j=1}^{n} p(y_j | x_j; y^{(j-1)})$$

where the set of outputs $y^{(j-1)}$ prior to y_j could be used for feedback effects.

Definition 2.3.2 Given two random variables X and Y with values $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, their mutual information is defined by

$$I(X;Y) := H(X) + H(Y) - H(X,Y)$$

Remark : given the joint probability $\pi_{XY} = \{P_{XY}(x,y)\}_{x,y \in \mathcal{X} \times \mathcal{Y}}$ and the marginal distributions

$$\pi_X = \left\{ P_X(x) = \sum_{y \in \mathcal{Y}} P_{XY}(x,y) \right\}_{x \in \mathcal{X}}$$

$$\pi_Y = \left\{ P_Y(y) = \sum_{x \in \mathcal{X}} P_{XY}(x,y) \right\}_{y \in \mathcal{Y}}$$

$$I(X;Y) := \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)\, P_Y(y)}$$

Remark : by introducing the conditional probabilities

$$\pi_{X|Y=y} = \left\{ P(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)} \right\}_{x \in \mathcal{X}} ; \quad \pi_{Y|X=x} = \left\{ P(y|x) = \frac{P_{XY}(x,y)}{P_X(x)} \right\}_{y \in \mathcal{Y}}$$

one gets

$$I(X;Y) = \sum_{x \in \mathcal{X}} P_X(x) \sum_{y \in \mathcal{Y}} P(y|x) \log \frac{P(y|x)}{P_Y(y)} = H(Y) - H(Y|X)$$

$$= \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \mathcal{X}} P(x|y) \log \frac{P(x|y)}{P_X(x)} = H(X) - H(X|Y)$$

Remark : $I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y)$

These expressions allow us to interprete the mutual information as the residual ignorance about one random variable when all the ignorance about another one has been eliminated.

Exercise 2.3.1

Show that $I(X;Y) = 0$ if X and Y are statistically independent and that $I(X;Y) = H(X)$ if $Y = f(X)$.

Definition 2.3.3 The conditional mutual information of random variables X and Y given Z is defined by

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

Exercise 2.3.2 Write $I(X;Y|Z)$ explicitly in terms of probabilities.

$\Pi_{XYZ} = \{P(x,y,z)\}_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}\\ z\in\mathcal{Z}}}$: joint probability

marginals

$$\Pi_{XY} = \{P(x,y) := \sum_{z\in\mathcal{Z}} P(x,y,z)\}_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}}; \quad \Pi_{XZ} = \{P(x,z) := \sum_{y\in\mathcal{Y}} P(x,y,z)\}_{\substack{x\in\mathcal{X}\\ z\in\mathcal{Z}}}; \quad \Pi_{YZ} = \{P(y,z) := \sum_{x\in\mathcal{X}} P(x,y,z)\}_{\substack{y\in\mathcal{Y}\\ z\in\mathcal{Z}}}$$

$$\Pi_X = \{P(x) := \sum_{y\in\mathcal{Y},\, z\in\mathcal{Z}} P(x,y,z)\}_{x\in\mathcal{X}}; \quad \Pi_Y = \{P(y) := \sum_{x\in\mathcal{X},\, z\in\mathcal{Z}} P(x,y,z)\}_{y\in\mathcal{Y}}; \quad \Pi_Z = \{P(z) := \sum_{x\in\mathcal{X},\, y\in\mathcal{Y}} P(x,y,z)\}_{z\in\mathcal{Z}}$$

$$\Pi_{X|Z=z} = \left\{P(x|z) = \frac{P(x,z)}{P(z)}\right\}_{x\in\mathcal{X}}; \quad \Pi_{X|Y=y,Z=z} = \left\{P(x|y,z) := \frac{P(x,y,z)}{P(y,z)}\right\}_{x\in\mathcal{X}}$$

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = \sum_{z\in\mathcal{Z}} P(z)\, H(X|Z=z) - \sum_{y\in\mathcal{Y},z\in\mathcal{Z}} P(y,z)\, H(X|Y=y,Z=z)$$

$$= \sum_{x\in\mathcal{X},\, y\in\mathcal{Y},\, z\in\mathcal{Z}} P(x,y,z) \log \frac{P(x,y,z)\,P(z)}{P(y,z)\,P(x,z)}$$

Definition 2.3.4 Random variables X, Y, Z are said to form a Markov chain in that order (denoted by $X \to Y \to Z$) if the conditional probability of Z depends only on Y:

$$p(x,y,z) = p(x)\,p(y|x)\,p(z|y) \iff p(z|x,y) = p(z|y)$$

Example 2.3.1

- $X \to Y \to Z$ iff X and Z are conditionally independent given Y.

$$p(x,z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)\,p(y|x)\,p(z|y)}{p(y)} = \frac{p(x,y)}{p(y)}\,p(z|y) = p(x|y)\,p(z|y)$$

- $X \to Y \to Z \implies Z \to Y \to X$

$$p(x|z,y) = \frac{p(x,y,z)}{p(y,z)} = \frac{p(x)\,p(y|x)\,p(z|y)}{p(y,z)} = \frac{p(x,y)}{p(y)} = p(x|y)$$

- $z = f(y) \to X \to Y \to Z$: $p(z|x,y) = \frac{p(x,y,f(y))}{p(x,y)} = \frac{p(x,y)}{p(x,y)} = 1 = p(z|y) = \frac{p(y,f(y))}{p(y)} = 1$

Proposition 2.3.1 Data Processing Inequality.

If $X \to Y \to Z$ then $I(X;Y) \geq I(X;Z)$.

Proof $I(X;Y,Z) = H(X) - H(X|Y,Z)$

$= H(X) - H(X|Z) + H(X|Z) - H(X|Y,Z)$

$= I(X;Z) + I(X;Y|Z)$

$I(X;Y,Z) = H(X) - H(X|Y,Z)$

$= H(X) - H(X|Y) + H(X|Y) - H(X|Y,Z)$

$= I(X;Y) + I(X;Z|Y)$

Since X and Z are conditionally independent given Y : $\boxed{I(X;Z|Y) = 0}$

Since $I(X;Y|Z) = H(X|Z) - H(X|Y,Z) \geq 0$: $\boxed{I(X;Y) = I(X;Z) + I(X;Y|Z) \geq 0}$

Exercise 2.3.3 Prove that $I(X;Y|Z) \geq 0$

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = -\sum_{z\in\mathcal{Z},\, x\in\mathcal{X}} p(x,z) \log \frac{p(x,z)}{p(z)}$$

$$+ \sum_{x\in\mathcal{X},\, y\in\mathcal{Y},\, z\in\mathcal{Z}} p(x,y,z) \log \frac{p(x,y,z)}{p(y,z)}$$

$$= \sum_{x,y,z} p(x,y,z) \log \frac{p(x,y,z)\,p(z)}{p(x,z)\,p(y,z)} \geq \sum_{x,y,z} \left(p(x,y,z) - \frac{p(x,z)\,p(y,z)}{p(z)} \right) = 0$$

Remark: $I(X;Y) \geq I(X;Z)$ when $X \to Y \to Z$ means that no processing of Y can increase the information relative to X.

Corollary 2.3.1 If $Z = g(Y)$ then $I(X;Y) \geq I(X;g(Y))$

Proof: $X \mapsto Y \to Z = g(Y)$

Lemma 2.3.1 Let $X^{(n)} = (X_1, X_2, \cdots, X_n)$ and $Y^{(n)} = (Y_1, Y_2, \cdots, Y_n)$ be joint random variables, then

$$I(X^{(n)}; Y^{(n)}) = H(Y^{(n)}) - \sum_{j=1}^{n} H(Y_j \mid X^{(n)}, Y^{(j-1)})$$

where $H(Y_1 \mid X^{(n)}, Y^{(0)}) = H(Y_1 \mid X^{(n)})$.

Proof.

$$I(X^{(n)}; Y^{(n)}) = H(Y^{(n)}) - H(Y^{(n)} \mid X^{(n)})$$

$$H(Y^{(n)} \mid X^{(n)}) = H(X^{(n)}, Y^{(n)}) - H(X^{(n)})$$

$$= H(Y_n \mid X^{(n)}, Y^{(n-1)}) + H(X^{(n)}, Y^{(n-1)}) - H(X^{(n)})$$

$$= H(Y_n \mid X^{(n)}, Y^{(n-1)}) + H(Y_{n-1} \mid X^{(n)}, Y^{(n-2)})$$

$$+ H(X^{(n)}, Y^{(n-2)}) - H(X^{(n)})$$

$$= \sum_{j=2}^{n} H(Y_j \mid X^{(n)}, Y^{(j-1)}) + \underbrace{H(X^{(n)}, Y_1) - H(X^{(n)})}_{H(Y_1 \mid X^{(n)})}$$