

# Entropy and Source coding

## I. DEFINITIONS OF CLASSICAL ENTROPIC FUNCTIONS

**Definition 1**  $u \in [0, \infty)^\Sigma$

$$\boxed{H(u) = - \sum_{a \in \Sigma} u(a) \log u(a)} \quad (1)$$

**Definition 2**  $u, v \in [0, \infty)^\Sigma$  and the support of  $u$  is contained in the support of  $v$  ( $\forall a \in \Sigma, u(a) > 0 \implies v(a) > 0$ ).

$$\boxed{D(u||v) = \sum_{a \in \Sigma} u(a) \log \frac{u(a)}{v(a)}} \quad (2)$$

**Remark 1** Note that we define

$$0 \log 0 = \lim_{x \rightarrow 0^+} x \log x = 0 \quad (3)$$

For all other choices of  $u$  and  $v$ , one defines  $D(u||v) = \infty$ .

**Definition 3** Scalar analogues of Shannon entropy and relative entropy

1.  $\eta : [0, \infty) \rightarrow \mathbb{R}$

$$\boxed{\eta(\alpha) = -\alpha \ln \alpha} \quad (4)$$

2.  $\theta : [0, \infty)^2 \rightarrow (-\infty, \infty]$  and  $\alpha > 0 \implies \beta > 0$

$$\boxed{\theta(\alpha, \beta) = \alpha \ln \frac{\alpha}{\beta}} \quad (5)$$

**Property 1** 1. The relationship between  $\eta$  and  $\theta$ :

$$\theta(\alpha, \beta) = -\beta \eta\left(\frac{\alpha}{\beta}\right) \quad (6)$$

$$\eta(\alpha) = -\theta(\alpha, 1) \quad (7)$$

2. The Shannon entropy function may be expressed in terms of the  $\eta$ -function as follows:

$$H(u) = \frac{1}{\ln 2} \sum_{a \in \Sigma} \eta(u(a)) \quad (8)$$

3. Expressed in terms of the  $\theta$ -function

$$D(u||v) = \frac{1}{\ln 2} \sum_{a \in \Sigma} \theta(u(a), v(a)) \quad (9)$$

4. The relationship between  $H$  and  $D$ :

$$H(u) = -D(u||1) \quad (10)$$

$$D(u||v) = - \sum_{a \in \Sigma} v(a) \eta\left(\frac{u(a)}{v(a)}\right) \quad (11)$$

5. Properties of  $\eta$ : concavity

$$\eta^{(n+1)}(\alpha) = \begin{cases} -(1 + \ln \alpha) & n = 0 \\ \frac{(-1)^n (n-1)!}{\alpha^n} & n \geq 1 \end{cases} \quad (12)$$

$$\eta(\lambda\alpha + (1-\lambda)\beta) \geq \lambda\eta(\alpha) + (1-\lambda)\eta(\beta) \quad (13)$$

6. Properties of  $\theta$ : subadditivity

$$\theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1) \geq \theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1) \quad (14)$$

The equality is achieved iff  $\alpha_0/\beta_0 = \alpha_1/\beta_1$ .

**Proof.**

$$\begin{aligned} \theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1) &= -(\beta_0 + \beta_1) \left[ \frac{\beta_0}{\beta_0 + \beta_1} \eta\left(\frac{\alpha_0}{\beta_0}\right) + \frac{\beta_1}{\beta_0 + \beta_1} \eta\left(\frac{\alpha_1}{\beta_1}\right) \right] \\ &\geq -(\beta_0 + \beta_1) \eta\left(\frac{\alpha_0 + \alpha_1}{\beta_0 + \beta_1}\right) \\ &= \theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1) \end{aligned}$$

■

7. Concavity of  $H$

$$H(\lambda u + (1-\lambda)v) \geq \lambda H(u) + (1-\lambda)H(v) \quad (15)$$

(Hint: concavity of  $\eta$ )

8. Subadditivity of  $D$

$$D(u_0 \| v_0) + D(u_1 \| v_1) \geq D(u_0 + u_1 \| v_0 + v_1) \quad (16)$$

The equality is achieved iff  $u_0(a)/v_0(a) = u_1(a)/v_1(a) \quad \forall a \in \Sigma$ . (Hint: subadditivity of  $\theta$ )

9. Joint convexity of  $D$

$$\lambda D(u_0 \| v_0) + (1-\lambda)D(u_1 \| v_1) \geq D(\lambda u_0 + (1-\lambda)u_1 \| \lambda v_0 + (1-\lambda)v_1) \quad (17)$$

10.  $\oplus$  and  $\otimes$

$$H(u \oplus v) = H(u) + H(v) \quad (18)$$

$$H(u \otimes v) = H(u) \sum_{b \in \Gamma} v(b) + H(v) \sum_{a \in \Sigma} u(a) \quad (19)$$

**Proof.**

$$\begin{aligned} H(u \otimes v) &= - \sum_{a \in \Sigma, b \in \Gamma} u(a)v(b) \log(u(a)v(b)) \\ &= - \sum_{a \in \Sigma, b \in \Gamma} u(a)v(b) \log(u(a)) - \sum_{a \in \Sigma, b \in \Gamma} u(a)v(b) \log(v(b)) \\ &= H(u) \sum_{b \in \Gamma} v(b) + H(v) \sum_{a \in \Sigma} u(a) \end{aligned}$$

■

11.  $p \in \mathcal{P}(\Sigma)$ ,  $\alpha > 0$

$$H(\alpha p) = \alpha H(p) - \alpha \log \alpha \quad (20)$$

12.  $p, q \in \mathcal{P}(\Sigma)$ ,  $\alpha, \beta > 0$

$$D(\alpha p \parallel \beta q) = \alpha D(p \parallel q) + \alpha \log \frac{\alpha}{\beta} \quad (21)$$

13. *Non-negativity of D*

$$\sum_{a \in \Sigma} u(a) \geq \sum_{a \in \Sigma} v(a) \implies D(u \parallel v) \geq 0 \quad (22)$$

*The equality is achieved iff  $u = v$ .*

**Proof.**

$$D(u \parallel v) = \frac{1}{\ln 2} \sum_{a \in \Sigma} \theta(u(a), v(a)) \geq \frac{1}{\ln 2} \theta\left(\sum_{a \in \Sigma} u(a), \sum_{a \in \Sigma} v(a)\right) \geq 0 \quad (23)$$

■

14. *Bounds of Shannon entropy of a single system*

$$\alpha = \sum_{a \in \Sigma} u(a) \quad (24)$$

$$0 \leq H(u) + \alpha \log \alpha \leq \alpha \log(|\Sigma|) \quad (25)$$

*The first equality is achieved iff  $u$  is pure. The second equality is achieved iff  $u$  is flat.*

*In particular,  $0 \leq H(p) \leq \log(|\Sigma|)$*

**Proof.**

$$D(p \parallel \frac{\mathbf{1}}{|\Sigma|}) = \log |\Sigma| - H(p) \geq 0 \implies H(p) \leq \log |\Sigma| \quad (26)$$

■

**Remark 2** *For convenience, we use  $H(X)$  to denote  $H(p)$ ,  $p \in \mathcal{P}(\Sigma)$ .*

**Definition 4** *Conditional Shannon entropy*

$$H(X|Y) = H(X, Y) - H(Y) = -D(p \parallel \mathbf{1} \otimes p[Y]) = \log |\Sigma| - D(p \parallel \frac{\mathbf{1}}{|\Sigma|} \otimes p[Y]) \quad (27)$$

*where  $\Sigma$  is the alphabet of register  $X$ .*

**Definition 5** *Mutual information*

$$I(X, Y) = H(X) + H(Y) - H(X, Y) = D(p \parallel p[X] \otimes p[Y]) \quad (28)$$

**Theorem 1** *Let  $X$  and  $Y$  be classical registers. With respect to an arbitrary probabilistic state of these registers, it holds that*

1. *A compound register cannot be greater than the total uncertainty one has about its individual registers*

$$H(X, Y) \leq H(X) + H(Y) \quad (29)$$

*The equality is saturated iff  $X$  and  $Y$  are uncorrelated.*

*(Hint:  $I(X : Y) = H(X) + H(Y) - H(X, Y) = D(p \parallel p[X] \otimes p[Y])$ )*

2. A pair of classical registers cannot be less than the Shannon entropy of either of the registers viewed in isolation.

$$H(X) \leq H(X, Y) \quad (30)$$

The equality is saturated iff the value of  $Y$  is totally dependent on the value of  $X$ . It should be noted that this property does not carry over to the von Neumann entropy of quantum states.

**Proof.**

$$H(X, Y) = - \sum_{ab} p(a, b) \log p(a, b) \geq - \sum_a \left( \sum_b p(a, b) \right) \log \left( \sum_b p(a, b) \right) = H(X) \quad (31)$$

■

3. (Strong subadditivity)  $H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z)$  with equality iff  $Z \rightarrow Y \rightarrow X$  forms a Markov chain. That is

$$\forall a, b, c \quad p[X = a | Y = b, Z = c] = p[X = a | Y = b] \quad (32)$$

And strong subadditivity is equivalent to

$$\begin{aligned} H(X|Y, Z) &\leq H(X|Y) \\ I(X : Y, Z) &\geq I(X : Y) \end{aligned}$$

**Theorem 2**  $p_0, p_1 \in \mathcal{P}(\Sigma)$ ,  $|\Sigma| \geq 2$  and  $\lambda = \frac{1}{2} \|p_0 - p_1\|_1$

$$|H(p_0) - H(p_1)| \leq \lambda \log(|\Sigma| - 1) + H(\lambda, 1 - \lambda) \quad (33)$$

The condition that the equality is saturated can be specified by the following proof.

**Proof.** Define

$$u_0(a) = \begin{cases} p_0(a) - p_1(a) & p_0(a) > p_1(a) \\ 0 & \text{otherwise} \end{cases} \quad (34)$$

$$u_1(a) = \begin{cases} p_1(a) - p_0(a) & p_0(a) < p_1(a) \\ 0 & \text{otherwise} \end{cases} \quad (35)$$

$$w(a) = \min\{p_0(a), p_1(a)\} \quad (36)$$

Then we have the following properties

1.  $\lambda = \sum_{a \in \Sigma} u_0(a) = \sum_{a \in \Sigma} u_1(a) = 1 - \sum_{a \in \Sigma} w(a)$
2.  $p_0 = u_0 + w$ ,  $p_1 = u_1 + w$

By concavity of Shannon entropy and notice that

$$-\alpha \log \alpha - \beta \log \beta + (\alpha + \beta) \log(\alpha + \beta) = (\alpha + \beta) H\left(\frac{\alpha}{\alpha + \beta}, \frac{\beta}{\alpha + \beta}\right) \quad (37)$$

$$0 \leq H(u_0) + H(w) - H(p_0) = \sum_{a \in \Sigma} p_0(a) H\left(\frac{u_0(a)}{p_0(a)}, \frac{w(a)}{p_0(a)}\right) \leq H(\lambda, 1 - \lambda)$$

$$0 \leq H(u_1) + H(w) - H(p_1) = \sum_{a \in \Sigma} p_1(a) H\left(\frac{u_1(a)}{p_1(a)}, \frac{w(a)}{p_1(a)}\right) \leq H(\lambda, 1 - \lambda)$$

$$\implies |(H(u_0) - H(u_1)) - (H(p_0) - H(p_1))| \leq H(\lambda, 1 - \lambda)$$

$$\implies |H(p_0) - H(p_1)| - |H(u_0) - H(u_1)| \leq H(\lambda, 1 - \lambda)$$

Then we need to prove that

$$|H(u_0) - H(u_1)| \leq \lambda \log(|\Sigma| - 1)$$

Just let  $\Gamma$  be a proper subset of  $\Sigma$ .

$$\sum_{b \in \Gamma} v(b) = \lambda \implies 0 \leq H(v) + \lambda \log \lambda \leq \lambda \log(|\Gamma|) \quad (38)$$

Thus

$$|H(u_0) - H(u_1)| \leq \lambda \log(|\Gamma|) \leq \lambda \log(|\Sigma| - 1) \quad (39)$$

■

**Theorem 3** *Let  $p_0, p_1 \in \mathcal{P}(\Sigma)$  be probability vectors, for  $\Sigma$  being an alphabet. It holds that*

$$D(p_0 \| p_1) \geq \frac{1}{2 \ln 2} \|p_0 - p_1\|_1^2 \quad (40)$$

The equality is saturated iff  $p_0 = p_1$ .

**Proof.** The following formula is useful.

$$\theta(\alpha, \beta) + \theta(1 - \alpha, 1 - \beta) \geq 2(\alpha - \beta)^2 \quad (41)$$

Define

$$\Sigma_0 = \{a \in \Sigma : p_0(a) > p_1(a)\}$$

$$\alpha = \sum_{a \in \Sigma_0} p_0(a) \quad \beta = \sum_{a \in \Sigma_0} p_1(a) \quad (42)$$

We have

$$\alpha - \beta = \sum_{a \in \Sigma} u_0(a) = \lambda \quad (43)$$

And

$$\begin{aligned} D(p_0 \| p_1) &= \frac{1}{\ln 2} \sum_{a \in \Sigma} \theta(p_0(a), p_1(a)) \\ &\geq \frac{1}{\ln 2} (\theta(\alpha, \beta) + \theta(1 - \alpha, 1 - \beta)) \\ &\geq \frac{2}{\ln 2} \lambda^2 = \frac{1}{2 \ln 2} \|p_0 - p_1\|_1^2 \end{aligned}$$

The bound is saturated iff  $p_0 = p_1$ . ■

**Definition 6** *Rényi function is defined as*

$$H_\alpha(u) = \frac{1}{1 - \alpha} \log \frac{\sum_{a \in \Sigma} u(a)^\alpha}{\sum_{a \in \Sigma} u(a)} \quad (44)$$

**Property 2** *Special cases:*

1.  $\alpha \rightarrow 0$ : Zero entropy

$$H_0(u) = \log \frac{|\Sigma|}{\sum_{a \in \Sigma} u(a)} \quad (45)$$

2.  $\alpha = \frac{1}{2}$ : *Max-entropy*

$$H_{\max}(u) = 2 \log \frac{\sum_{a \in \Sigma} \sqrt{u(a)}}{\sum_{a \in \Sigma} u(a)} \quad (46)$$

3.  $\alpha \rightarrow 1$ : *Shannon entropy*

$$\begin{aligned} \lim_{\alpha \rightarrow 1} H_{\alpha}(u) &= \lim_{\alpha \rightarrow 1} \frac{1}{1 - \alpha} \log \frac{\sum_{a \in \Sigma} u(a)^{\alpha}}{\sum_{a \in \Sigma} u(a)} \\ &= \lim_{\alpha \rightarrow 1} - \frac{\sum_{a \in \Sigma} u(a)}{\sum_{a \in \Sigma} u(a)^{\alpha}} \sum_{a \in \Sigma} u(a)^{\alpha} \log u(a) \\ &= - \sum_{a \in \Sigma} u(a) \log u(a) \end{aligned}$$

4.  $\alpha = 2$ : *Collision entropy*

5.  $\alpha \rightarrow \infty$ : *Min-entropy*

$$H_{\min}(u) = - \log \frac{\max_{a \in \Sigma} u(a)}{\sum_{a \in \Sigma} u(a)} \quad (47)$$

**Definition 7**

$$D_{\alpha}(u||v) = \frac{1}{\alpha - 1} \log \frac{\sum_{a \in \Sigma} u(a)^{\alpha} v(a)^{1-\alpha}}{\sum_{a \in \Sigma} u(a)} \quad (48)$$

It is easy to see that

$$H_{\alpha}(u) = - D_{\alpha}(u||\mathbf{1}) \quad (49)$$

We can define conditional entropy and mutual information this way:

$$H_{\alpha}(X|Y) = - \min_{q \in \mathcal{P}(\Gamma)} D_{\alpha}(p||\mathbf{1} \otimes q) \quad (50)$$

$$I_{\alpha}(X : Y) = \min_{q \in \mathcal{P}(\Gamma)} D_{\alpha}(p||p[X] \otimes q) \quad (51)$$

## II. DEFINITIONS OF QUANTUM ENTROPIC FUNCTIONS

**Definition 8**  $P \in \text{Pos}(\mathcal{X})$ . The von Neumann entropy of  $P$  is defined as

$$\boxed{H(P) = H(\lambda(P)) = - H(P \log P)} \quad (52)$$

for  $\lambda(P)$  being the vector of eigenvalues of  $P$ .

**Remark 3** Similar to the Shannon entropy usually being considered for probability vectors, it is most common that one considers the von Neumann entropy function on density operator inputs.

**Definition 9**  $P, Q \in \text{Pos}(\mathcal{X})$ ,  $\text{im}(P) \subset \text{im}(Q)$ . The quantum relative entropy of  $P$  with respect to  $Q$  is defined as

$$\boxed{D(P||Q) = \text{Tr}(P \log P) - \text{Tr}(P \log Q)} \quad (53)$$

**Definition 10** The conditional von Neumann entropy and quantum mutual information are defined in an analogous manner to the conditional Shannon entropy and mutual information.

**Property 3** 1.  $P, Q \in \mathbb{C}^\Sigma$  with spectrum decompositions

$$P = \sum_{a \in \Sigma} \lambda_a x_a x_a^* \quad Q = \sum_{b \in \Sigma} \mu_b y_b y_b^* \quad (54)$$

Then

$$\boxed{D(P\|Q) = \frac{1}{\ln 2} \sum_{a,b \in \Sigma} \theta(|x_a^* y_b|^2 \lambda_a, |x_a^* y_b|^2 \mu_b)} \quad (55)$$

Hint

$$D(P\|Q) = \sum_{a \in \Sigma} \lambda_a \log \lambda_a - \sum_{a,b \in \Sigma} |x_a^* y_b|^2 \lambda_a \log \mu_b = \sum_{a,b \in \Sigma} |x_a^* y_b|^2 \lambda_a \log \frac{\lambda_a}{\mu_b} \quad (56)$$

2.  $P, Q \in \text{Pos}(\mathcal{X})$  and  $V \in \mathcal{U}(\mathcal{X}, \mathcal{Y})$

$$H(VPV^*) = H(P) \quad D(VPV^* \| VQV^*) = D(P\|Q) \quad (57)$$

3.  $P, Q \in \text{Pos}(\mathcal{X})$

$$H\left(\begin{bmatrix} P & \\ & Q \end{bmatrix}\right) = H(P) + H(Q) \quad (58)$$

$$H(P \otimes Q) = \text{Tr}(Q) H(P) + \text{Tr}(P) H(Q) \quad (59)$$

$$D(P_0 \otimes P_1 \| Q_0 \otimes Q_1) = \text{Tr}(P_1) D(P_0 \| Q_0) + \text{Tr}(P_0) D(P_1 \| Q_1) \quad (60)$$

$$H(\alpha\rho) = \alpha H(\rho) - \alpha \log \alpha \quad (61)$$

$$D(\alpha\rho \| \beta\sigma) = \alpha D(\rho \| \sigma) + \alpha \log \frac{\alpha}{\beta} \quad (62)$$

4. non-negativity of quantum relative entropy

$$\text{Tr}(P) \geq \text{Tr}(Q) \implies D(P\|Q) \geq 0 \quad (63)$$

with equality saturated iff  $P = Q$ . This is equivalent to

$$\begin{aligned} H(P) &\leq -\text{Tr}(P \log Q) \quad \text{Tr}(Q) \leq \text{Tr}(P) \\ H(P) &= \min_{\text{Tr}(Q) \leq \text{Tr}(P)} -\text{Tr}(P \log Q) \end{aligned}$$

$$\begin{aligned} D(P\|Q) &= \frac{1}{\ln 2} \sum_{a,b \in \Sigma} \theta(|x_a^* y_b|^2 \lambda_a, |x_a^* y_b|^2 \mu_b) \\ &\geq \frac{1}{\ln 2} \theta\left(\sum_{a,b \in \Sigma} |x_a^* y_b|^2 \lambda_a, \sum_{a,b \in \Sigma} |x_a^* y_b|^2 \mu_b\right) \\ &= \frac{1}{\ln 2} \theta(\text{Tr}(P), \text{Tr}(Q)) \end{aligned}$$

5. Concavity

$$H(\lambda P + (1 - \lambda)Q) \geq \lambda H(P) + (1 - \lambda) H(Q) \quad (64)$$

**Proof.** Since  $H$  is a continuous, the following mid-point convexity is enough.

$$H\left(\left[\begin{array}{cc} P & \\ & Q \end{array}\right] \parallel \left[\begin{array}{cc} \frac{P+Q}{2} & \\ & \frac{P+Q}{2} \end{array}\right]\right) = 2H\left(\frac{P+Q}{2}\right) - H(P) - H(Q) \geq 0$$

Another proof. Let  $f(x) = H(xP + (1-x)Q)$ . The convexity of  $H$  is equivalent to

$$f(\lambda) \geq \lambda f(1) + (1-\lambda)f(0) \quad (65)$$

It suffices to prove  $f$  is convex over  $[0, 1]$ .

$$\begin{aligned} f'(x) &= \lim_{\Delta x \rightarrow 0} \frac{\Delta f(x)}{\Delta x} \\ &= -\text{Tr}[(P-Q)\log(xP + (1-x)Q) + (P-Q)] \\ f''(x) &= \lim_{\Delta x \rightarrow 0} \frac{\Delta f'(x)}{\Delta x} \\ &= -\text{Tr}[(P-Q)(xP + (1-x)Q)^{-1}(P-Q)] \\ &\leq 0 \end{aligned}$$

A third proof. Consider a classical-quantum state  $\sigma \in D(\mathcal{X} \otimes \mathcal{Y})$ :

$$\sigma = \sum_{a \in \Sigma} p(a) E_{a,a} \otimes \rho_a \quad (66)$$

$$\begin{cases} H(X) = H(p) \\ H(Y) = H(\sum_{a \in \Sigma} p(a) \rho_a) \\ H(X, Y) = H(p) + \sum_{a \in \Sigma} p(a) H(\rho_a) \\ H(X) + H(Y) \geq H(X, Y) \end{cases} \implies H\left(\sum_{a \in \Sigma} p(a) \rho_a\right) \geq \sum_{a \in \Sigma} p(a) H(\rho_a) \quad (67)$$

## 6. Subadditivity

$$H(X, Y) \leq H(X) + H(Y) \quad (68)$$

with equality iff  $X$  and  $Y$  are uncorrelated.

$$D(\rho \parallel \rho[X] \otimes \rho[Y]) = -H(\rho) + H(\rho[X]) + H(\rho[Y]) \geq 0$$

The following formula is useful.

$$\log(P \otimes Q) = \log P \otimes \mathbb{1} + \mathbb{1} \otimes \log Q \quad (69)$$

7. Assume the compound register  $(X, Y)$  is in a pure state  $u u^*$ . The Schmidt decomposition of  $u$  is

$$u = \sum_{a \in \Sigma} \sqrt{p(a)} x_a \otimes y_a \quad (70)$$

Then

$$H(X) = H(Y) = H(p) \quad (71)$$

8.  $H(X) \leq H(Y) + H(X, Y)$  The equality is achieved iff  $Y$  is uncorrelated with the purifying system. Or specifically, Let  $\rho \in D(\mathcal{X}, \otimes \mathcal{Y})$  with the following spectrum decomposition:

$$\rho = \sum_{a \in \Sigma} p(a) u_a u_a^* \quad (72)$$

The equality is saturated iff  $\{\text{Tr}_Y u_a u_a : a \in \Sigma\}$  have a common eigenbasis and  $\{\text{Tr}_X u_a u_a : a \in \Sigma\}$  are orthogonal.

To prove the result, consider a purification  $u \in D(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$  of  $\rho \in D(\mathcal{X}, \mathcal{Y})$ . Then

$$\begin{aligned} H(X) &= H(Y, Z) \\ H(X, Y) &= H(Z) \\ H(Y, Z) &\leq H(Y) + H(Z) \end{aligned}$$



9. Combining 6 and 8, we have

$$|\mathbf{H}(\mathbf{X}) - \mathbf{H}(\mathbf{Y})| \leq \mathbf{H}(\mathbf{X}, \mathbf{Y}) \leq \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}) \quad (73)$$

**Theorem 4** (*Fannes-Audenaert inequality*) Let  $\rho_0, \rho_1 \in \mathbf{D}(\mathcal{X})$  be density operators, for  $\mathcal{X}$  a complex Euclidean space of dimension  $n \geq 2$ , and let

$$\delta = \frac{1}{2} \|\rho_0 - \rho_1\|_1 \quad (74)$$

we have

$$|\mathbf{H}(\rho_0) - \mathbf{H}(\rho_1)| \leq \delta \log(n-1) + \mathbf{H}(\delta, 1-\delta) \quad (75)$$

The condition that the equality is saturated can be specified by the following proof.

**Proof.** The following bound is useful.

$$\sum_{k=1}^n |\lambda_k(X) - \lambda_k(Y)| \leq \|X - Y\|_1 \leq \sum_{k=1}^n |\lambda_k(X) - \lambda_{n-k+1}(Y)| \quad (76)$$

Define

$$\delta_0 = \frac{1}{2} \sum_{k=1}^n |\lambda_k(\rho_0) - \lambda_k(\rho_1)| \quad (77)$$

$$\delta_1 = \frac{1}{2} \sum_{k=1}^n |\lambda_k(\rho_0) - \lambda_{n-k+1}(\rho_1)| \quad (78)$$

$$(79)$$

Use the above bound, we have  $\delta_0 \leq \delta \leq \delta_1$ . Let

$$\delta = \alpha \delta_0 + (1-\alpha) \delta_1 \quad (80)$$

$$|\mathbf{H}(\rho_0) - \mathbf{H}(\rho_1)| \leq \delta_0 \log(n-1) + \mathbf{H}(\delta_0, 1-\delta_0)$$

$$|\mathbf{H}(\rho_0) - \mathbf{H}(\rho_1)| \leq \delta_1 \log(n-1) + \mathbf{H}(\delta_1, 1-\delta_1)$$

Thus

$$\begin{aligned} |\mathbf{H}(\rho_0) - \mathbf{H}(\rho_1)| &\leq \alpha(\delta_0 \log(n-1) + \mathbf{H}(\delta_0, 1-\delta_0)) + (1-\alpha)(\delta_1 \log(n-1) + \mathbf{H}(\delta_1, 1-\delta_1)) \\ &\leq \delta \log(n-1) + \mathbf{H}(\delta, 1-\delta) \end{aligned}$$

■

**Theorem 5** (*Projective measurement increases entropy*)  $Q \in \mathbf{Pos}(\mathcal{X})$ . Let  $\{P_a : a \in \Sigma\} \in \mathbf{Proj}(\mathcal{X})$  be a complete set of projectors.

$$Q' = \sum_{a \in \Sigma} P_a Q P_a \quad (81)$$

$$\mathbf{H}(Q') \geq \mathbf{H}(Q) \quad (82)$$

**Proof.**

$$\begin{aligned} -\mathrm{Tr}(Q \log Q') &= -\mathrm{Tr} \left( \sum_{a \in \Sigma} P_a Q \log Q' \right) \\ &= -\sum_{a \in \Sigma} \mathrm{Tr}(P_a Q \log Q' P_a) \\ &= -\sum_{a \in \Sigma} \mathrm{Tr}(P_a Q P_a \log Q') \\ &= \mathbf{H}(Q') \end{aligned}$$

Then

$$H(Q') - H(Q) = -\text{Tr}(Q \log Q') + \text{Tr}(Q \log Q) = D(Q \| Q') \geq 0 \quad (83)$$

■

**Theorem 6** (*Entropy of classical-quantum state*)  $p \in \mathcal{P}(\Sigma)$ ,  $\{\rho_a : a \in \Sigma\} \subset \mathcal{D}(\mathcal{Y})$ . Define a classical-quantum state  $\sigma \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$

$$\sigma = \sum_{a \in \Sigma} p(a) E_{a,a} \otimes \rho_a \quad (84)$$

Then

$$\begin{aligned} H(X) &= H(p) \\ H(Y) &= H\left(\sum_{a \in \Sigma} p(a) \rho_a\right) \\ H(X, Y) &= H(p) + \sum_{a \in \Sigma} p(a) H(\rho_a) \\ H(Y) &\leq H(X, Y) \\ H(X) &\leq H(X, Y) \end{aligned}$$

$H(Y) = H(X, Y)$  holds iff  $\{\rho(a)\}$  are orthogonal.  $H(X) = H(X, Y)$  holds iff  $\{\rho(a)\}$  are pure.

**Proof.** Suppose  $\rho_a = u_a u_a^*$ , consider a purification of  $\rho = \sum_{a \in \Sigma} p(a) \rho_a$

$$w = \sum_{a \in \Sigma} \sqrt{p(a)} u_a \otimes e_a \quad (85)$$

Then

$$H(\rho) = H(X) = H(Y) \quad (86)$$

Perform a projective measurement on  $w[Y]$ . We get

$$H(Y) \leq H(p) \quad (87)$$

Thus, we have proved that when the states  $\rho_a$  are pure,

$$H\left(\sum_{a \in \Sigma} p(a) \rho_a\right) \leq H(p) \quad (88)$$

If not pure, we can decompose

$$\rho_a = \sum_{b \in \Gamma} q(a, b) v_b v_b^* \quad (89)$$

Then

$$\begin{aligned} H\left(\sum_{a \in \Sigma} p(a) \rho_a\right) &= H\left(\sum_{a \in \Sigma, b \in \Gamma} p(a) q(a, b) v_b v_b^*\right) \\ &\leq H(p(a) q(a, b)) \\ &= H(p) + \sum_{a \in \Sigma} p(a) H(\rho_a) \end{aligned}$$

■

**Theorem 7** *Combining concavity and the entropy inequality of classical-quantum state, we have*

$$\sum_{a \in \Sigma} p(a) H(\rho_a) \leq H\left(\sum_{a \in \Sigma} p(a) \rho_a\right) \leq \sum_{a \in \Sigma} p(a) H(\rho_a) + H(p) \quad (90)$$

**Theorem 8**

$$D(P||Q) = \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \log \frac{\text{Tr}(P)}{\langle P^{1-\varepsilon}, Q^\varepsilon \rangle} = \frac{1}{\ln 2} \lim_{\varepsilon \rightarrow 0^+} \frac{\text{Tr}(P) - \langle P^{1-\varepsilon}, Q^\varepsilon \rangle}{\varepsilon} \quad (91)$$

**Lemma 1** *Let  $P, Q \in \text{Pos}(\mathcal{X})$  be positive semidefinite operators such that  $[P, Q] = 0$ , and let  $H \in \text{Herm}(\mathcal{X})$  be a Hermitian operator for which*

$$\begin{pmatrix} P & H \\ H & Q \end{pmatrix} \in \text{Pos}(\mathcal{X} \oplus \mathcal{X}) \quad (92)$$

*It holds that  $H \leq \sqrt{P}\sqrt{Q}$ .*

**Proof.**

$$\begin{aligned} & \begin{pmatrix} P & H \\ H & Q \end{pmatrix} \in \text{Pos}(\mathcal{X} \oplus \mathcal{X}) \\ \implies & \|P^{-\frac{1}{2}} H Q^{-\frac{1}{2}}\| \leq 1 \\ \implies & \|P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}}\| \leq 1 \\ \implies & H \leq \sqrt{P}\sqrt{Q} \end{aligned}$$

■

**Lemma 2** *(Liebs concavity theorem)  $A_0, A_1 \in \text{Pos}(\mathcal{X})$ ,  $B_0, B_1 \in \text{Pos}(\mathcal{Y})$ ,  $\alpha \in [0, 1]$*

$$(A_0 + A_1)^\alpha \otimes (B_0 + B_1)^{1-\alpha} \geq A_0^\alpha \otimes B_0^{1-\alpha} + A_1^\alpha \otimes B_1^{1-\alpha} \quad (93)$$

**Proof.** Define

$$\begin{aligned} X(\alpha) &= A_0^\alpha \otimes B_0^{1-\alpha} \\ Y(\alpha) &= A_1^\alpha \otimes B_1^{1-\alpha} \\ Z(\alpha) &= (A_0 + A_1)^\alpha \otimes (B_0 + B_1)^{1-\alpha} \\ f(\alpha) &= Z(\alpha) - X(\alpha) - Y(\alpha) \end{aligned}$$

Our goal is to prove that  $f(\alpha) \geq 0$  for every  $\alpha \in [0, 1]$ .  $f(0) \geq 0$  and  $f(1) \geq 0$  are trivial.  $f$  is continuous on the interval  $[0, 1]$ , and therefore the preimage of the closed set  $\text{Pos}(\mathcal{X} \otimes \mathcal{Y})$  under this function is closed. It therefore suffices to prove that the set  $\{\alpha \in [0, 1] : f(\alpha) \geq 0\}$  is dense in  $[0, 1]$ .

Suppose  $f(\alpha) > 0$  and  $f(\beta) > 0$ .

$$\begin{aligned} & \begin{bmatrix} Z(\alpha) & X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) \\ X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) & Z(\alpha) \end{bmatrix} \\ &= \begin{bmatrix} \sqrt{X(\alpha)} \\ \sqrt{X(\beta)} \end{bmatrix} [\sqrt{X(\alpha)} \quad \sqrt{X(\beta)}] + \begin{bmatrix} \sqrt{Y(\alpha)} \\ \sqrt{Y(\beta)} \end{bmatrix} [\sqrt{Y(\alpha)} \quad \sqrt{Y(\beta)}] \\ &\in \text{Pos}(X \oplus \mathcal{X}) \\ \implies & X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) \leq \sqrt{Z(\alpha)}\sqrt{Z(\beta)} = Z\left(\frac{\alpha+\beta}{2}\right) \end{aligned}$$

Then we keep dividing a interval into two halves. We get  $f(\alpha) \geq 0$  for the set  $\{\alpha = k/2^n : k, n \in \mathbb{N}^*, k \leq 2^n\}$ . This completes the proof. ■

**Corollary 1**  $P_0, P_1, Q_0, Q_1 \in \text{Pos}(\mathcal{X})$

$$\langle (P_0 + P_1)^\alpha, (Q_0 + Q_1)^{1-\alpha} \rangle \geq \langle P_0^\alpha, Q_0^{1-\alpha} \rangle + \langle P_1^\alpha, Q_1^{1-\alpha} \rangle \quad (94)$$

**Theorem 9**  $P_0, P_1, Q_0, Q_1 \in \text{Pos}(\mathcal{X})$

$$D(P_0 + P_1 \| Q_0 + Q_1) \leq D(P_0 \| Q_0) + D(P_1 \| Q_1) \quad (95)$$

**Proof.**

$$\begin{aligned} & D(P_0 + P_1 \| Q_0 + Q_1) \\ &= \frac{1}{\ln 2} \lim_{\varepsilon \rightarrow 0} \frac{\text{Tr}(P_0 + P_1) - \langle (P_0 + P_1)^{1-\varepsilon}, (Q_0 + Q_1)^\varepsilon \rangle}{\varepsilon} \\ &\leq \frac{1}{\ln 2} \left( \lim_{\varepsilon \rightarrow 0} \frac{\text{Tr} P_0 - \langle P_0^{1-\varepsilon}, Q_0^\varepsilon \rangle}{\varepsilon} + \lim_{\varepsilon \rightarrow 0} \frac{\text{Tr} P_1 - \langle P_1^{1-\varepsilon}, Q_1^\varepsilon \rangle}{\varepsilon} \right) \\ &= D(P_0 \| Q_0) + D(P_1 \| Q_1) \end{aligned}$$

■

**Corollary 2** (*Joint convexity of quantum relative entropy*)  $P_0, P_1, Q_0, Q_1 \in \text{Pos}(\mathcal{X})$ ,  $\lambda \in [0, 1]$ . It holds that

$$D(\lambda P_0 + (1 - \lambda) P_1 \| \lambda Q_0 + (1 - \lambda) Q_1) \leq \lambda D(P_0 \| Q_0) + (1 - \lambda) D(P_1 \| Q_1) \quad (96)$$

**Theorem 10** (*Monotonicity of quantum relative entropy*)  $P, Q \in \text{Pos}(\mathcal{X})$ ,  $\Phi \in \text{Chan}(\mathcal{X}, \mathcal{Y})$ .

$$D(\Phi(P) \| \Phi(Q)) \leq D(P \| Q) \quad (97)$$

**Proof.** First use joint convexity to prove that it holds for mixed unitary channels. Then

$$\begin{aligned} D(\Phi(P) \| \Phi(Q)) &= D(\Phi(P) \otimes \omega \| \Phi(Q) \otimes \omega) \\ &= D((\mathbb{1}_{\text{Lin}(\mathcal{Y})} \otimes \Omega)(APA^*) \| (\mathbb{1}_{\text{Lin}(\mathcal{Y})} \otimes \Omega)(AQA^*)) \\ &\leq D(APA^* \| AQA^*) \\ &= D(P \| Q) \end{aligned}$$

■

**Theorem 11** (*Strong subadditivity of von Neumann entropy*)

$$H(X, Y, Z) + H(Z) \leq H(X, Z) + H(Y, Z) \quad (98)$$

This is equivalent to

$$\begin{aligned} H(X|Y, Z) &\leq H(X|Z) \\ I(X : Y, Z) &\geq I(X : Z) \end{aligned}$$

**Proof.**

$$\begin{aligned} & D(\rho[X, Y, Z] \| \rho[X] \otimes \rho[Y, Z]) \geq D(\rho[X, Z] \| \rho[X] \otimes \rho[Z]) \\ \implies & H(X) + H(Y, Z) - H(X, Y, Z) \geq H(X) + H(Z) - H(X, Z) \end{aligned}$$

■

**Theorem 12** (*Quantum Pinsker inequality*) Let  $\rho_0, \rho_1 \in D(\mathcal{X})$  be density operators, for  $\mathcal{X}$  a complex Euclidean space. It holds that

$$D(\rho_0 \| \rho_1) \geq \frac{1}{2 \ln 2} \|\rho_0 - \rho_1\|_1 \quad (99)$$

The equality is saturated iff  $\rho_0 = \rho_1$ .

**Proof.** Perform an optimal measurement which discriminate the two states, then we get two classical state  $p_0$  and  $p_1$  such that

$$\|p_0 - p_1\|_1 = \|\rho_0 - \rho_1\|_1 \quad (100)$$

By monotonicity of quantum relative entropy, we get

$$D(\rho_0\|\rho_1) \geq D(p_0\|p_1) \geq \frac{1}{2\ln 2} \|p_0 - p_1\|_1 = \frac{1}{2\ln 2} \|\rho_0 - \rho_1\|_1 \quad (101)$$

■

**Definition 11** *Rényi entropies*

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \frac{\text{Tr}(\rho^\alpha)}{\text{Tr}(\rho)} \quad (102)$$

**Definition 12** *Quantum divergences have multiple definitions*

$$\begin{aligned} \tilde{D}_\alpha(P\|Q) &= \frac{1}{\alpha-1} \log \frac{1}{\text{Tr}(P)} \text{Tr}[(Q^{\frac{1}{2\alpha}-\frac{1}{2}} P Q^{\frac{1}{2\alpha}-\frac{1}{2}})^\alpha] = \frac{\alpha}{\alpha-1} \log \frac{1}{\text{Tr}(P)^{1/\alpha}} \left\| Q^{\frac{1}{2\alpha}-\frac{1}{2}} P Q^{\frac{1}{2\alpha}-\frac{1}{2}} \right\|_\alpha \\ D_\alpha(P\|Q) &= \frac{1}{\alpha-1} \log \frac{\text{Tr}(P^\alpha Q^{1-\alpha})}{\text{Tr}(P)} \end{aligned}$$

*It is easy to see that*

$$H_\alpha(P) = -D_\alpha(P\|\mathbb{1}) = \tilde{H}_\alpha(P) = -\tilde{D}_\alpha(P\|\mathbb{1}) \quad (103)$$

*In particular, we define*

$$H_0(P) = -D_0(P\|\mathbb{1}) = \log \frac{\text{Tr}(\Pi_P)}{\text{Tr}(P)} \quad (104)$$

**Remark 4** *The range of  $\alpha$ :*

- $\tilde{D}_\alpha: [\frac{1}{2}, \infty)$
- $D_\alpha: [0, 2]$

**Theorem 13** *If  $[P, Q] = 0$ , then  $\tilde{D}_\alpha(P\|Q) = D_\alpha(P\|Q)$ .*

**Theorem 14**  $\tilde{D}_\alpha(P\|Q) \leq D_\alpha(P\|Q)$

**Property 4** *Special cases*

1.  $\alpha = 0$

$$D_0(P\|Q) = \log \text{Tr}(P) - \log \text{Tr}(\Pi_{\text{im}(P)} Q) \quad (105)$$

2.  $\alpha = \frac{1}{2}$

$$\tilde{D}_{\frac{1}{2}}(P\|Q) = -2 \log \frac{\text{Tr}[(Q^{1/2} P Q^{1/2})^{1/2}]}{\text{Tr}(P)} = -2 \log \frac{F(P, Q)}{\text{Tr}(P)} \quad (106)$$

3.  $\alpha = 1$ : *the usual relative entropy.*

$$\tilde{D}_1(P\|Q) = D_1(P\|Q) = \text{Tr}(P \log P) - \text{Tr}(P \log Q) \quad (107)$$

4.  $\alpha = 2$ : collision entropy

5.  $\alpha \rightarrow \infty$

$$\tilde{D}_\infty(P\|Q) = \min\{\lambda : P \leq 2^\lambda Q\} = \log \|Q^{-1/2} P Q^{-1/2}\| \quad (108)$$

**Definition 13** *Conditional entropy*

$$H_\alpha(X|Y) = - \min_{\sigma \in \mathcal{D}(\mathcal{Y})} D_\alpha(\rho \| \mathbb{1}_X \otimes \sigma) \quad (109)$$

$$\tilde{H}_\alpha(X|Y) = - \min_{\sigma \in \mathcal{D}(\mathcal{Y})} \tilde{D}_\alpha(\rho \| \mathbb{1}_X \otimes \sigma) \quad (110)$$

**Property 5** *Special cases*

1.  $\alpha \rightarrow 0$ : Zero-entropy

$$\begin{aligned} H_0(X|Y) &= - \min_{\sigma \in \mathcal{D}(\mathcal{Y})} D_0(\rho \| \mathbb{1}_X \otimes \sigma) \\ &= \max_{\sigma \in \mathcal{D}(\mathcal{Y})} \log \text{Tr}(\Pi_{\text{im}(\rho)}(\mathbb{1}_X \otimes \sigma)) \\ &= \log \|\text{Tr}_X \Pi_{\text{im}(\rho)}\| \end{aligned}$$

2.  $\alpha = \frac{1}{2}$ : Max-entropy

$$\begin{aligned} H_{\max}(X|Y) &= \tilde{H}_{\frac{1}{2}}(X|Y) \\ &= - \min_{\sigma \in \mathcal{D}(\mathcal{Y})} \tilde{D}_{\frac{1}{2}}(\rho \| \mathbb{1}_X \otimes \sigma) \\ &= \max_{\sigma \in \mathcal{D}(\mathcal{Y})} 2 \log F(\rho, \mathbb{1}_X \otimes \sigma) \end{aligned}$$

3.  $\alpha \rightarrow \infty$ : Min-entropy

$$\begin{aligned} H_{\min}(X|Y) &= \tilde{H}_\infty(X|Y) \\ &= - \min_{\sigma \in \mathcal{D}(\mathcal{Y})} \tilde{D}_\infty(\rho \| \mathbb{1}_X \otimes \sigma) \end{aligned}$$

The min-entropy and the max-entropy, as defined in the previous section, are discontinuous in the sense that a slight modification of the system's state might have a large impact on its entropy.

We will see later that the zero-entropy  $H_0(p_X)$  can be interpreted as the minimum number of bits needed to encode  $X$  in such a way that its value can be recovered from the encoding without errors. Indeed, while we need at least  $\log n$  bits to store a value  $X$  distributed according to  $p_X$ , one single bit is sufficient to store a value distributed according to  $\bar{p}_X$ . However, for most applications, we allow some small error probability. For example, we might want to encode  $X$  in such a way that its value can be recovered with probability  $1 - \varepsilon$ .

**Definition 14** Let  $\hat{\rho}, \rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$  and  $\hat{\rho}$  is  $\varepsilon$ -close to  $\rho$ . One definition of distance is the fidelity distance  $F^2(\hat{\rho}, \rho) \geq 1 - \varepsilon^2$ .

$$H_{\min}^\varepsilon(X|Y)_\rho = \sup_{\hat{\rho}} H_{\min}(X|Y)_{\hat{\rho}} \quad (111)$$

$$H_{\max}^\varepsilon(X|Y)_\rho = \inf_{\hat{\rho}} H_{\max}(X|Y)_{\hat{\rho}} \quad (112)$$

### III. CLASSICAL SOURCE CODING

**Definition 15**  $\Gamma = \{0, 1\}$ ,  $n \in \mathbb{N}^*$ ,  $\alpha > 0$ ,  $\delta \in (0, 1)$ ,  $m = \lfloor \alpha n \rfloor$

$$f : \Sigma^n \rightarrow \Gamma^m \quad g : \Gamma^m \rightarrow \Sigma^n \quad (113)$$

is said to be an  $(n, \alpha, \delta)$ -coding scheme for  $p \in \mathcal{P}(\Sigma)$  if it holds that

$$G = \{a_1 \cdots a_n \in \Sigma^n : g(f(a_1 \cdots a_n)) = a_1 \cdots a_n\} \quad (114)$$

$$\sum_{a_1, \dots, a_n \in G} p(a_1) \cdots p(a_n) > 1 - \delta \quad (115)$$

Alice encode  $a_1 \cdots a_n$  to a string with  $m = \lfloor \alpha n \rfloor$  bits and send it to Bob. If it is the case that the pair  $(f, g)$  is an  $(n, \alpha, \delta)$ -coding scheme for  $p$ , then the number  $\delta$  is an upper bound on the probability that the coding scheme fails to be correct, so that Bob does not recover the string Alice obtained from the source, while  $\alpha$  represents the average number of bits (as the value of  $n$  increases) needed to encode each symbol.

**Definition 16** *Typical strings.  $\varepsilon$ -typical*

$$\left| \frac{1}{n} \sum_{i=1}^n \log \frac{1}{p(a_i)} - H(p) \right| < \varepsilon$$

$$2^{-n(H(p)+\varepsilon)} < p(a_1) \cdots p(a_n) < 2^{-n(H(p)-\varepsilon)}$$

The set of  $\varepsilon$ -strings is denoted  $T_{n,\varepsilon}(p)$ .

**Theorem 15**

$$\lim_{n \rightarrow \infty} \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) = 1 \quad (116)$$

(Hint: use the weak law of large numbers)

**Theorem 16**

$$|T_{n,\varepsilon}(p)| < 2^{n(H(p)+\varepsilon)} \quad (117)$$

**Theorem 17** *Shannon's source coding theorem. Let  $\text{Scheme}(p, n, \alpha, \delta)$  denote the set of  $(n, \alpha, \delta)$ -coding schemes for  $p$ . Fix  $\alpha$ ,  $p$  and  $\delta$*

1. If  $\alpha > H(p)$ , then

$$\exists N \in \mathbb{N} \ (n \geq N \implies \text{Scheme}(p, n, \alpha, \delta) \neq \emptyset) \quad (118)$$

2. If  $\alpha < H(p)$ , then

$$\forall N \in \mathbb{N} \ (\exists n \geq N \ \text{Scheme}(p, n, \alpha, \delta) = \emptyset) \quad (119)$$

**Proof.**

1. Assume  $\alpha > H(p)$  and choose  $\varepsilon > 0$  such that  $\alpha > H(p) + 2\varepsilon$ . A coding theorem will be defined for  $n > 1/\varepsilon$

$$m = \lfloor \alpha n \rfloor > n(H(p) + \varepsilon)$$

Then

$$|T_{n,\varepsilon}| < 2^{n(H(p)+\varepsilon)} < 2^m$$

one may therefore define a function  $f_n : \Sigma^n \rightarrow \Gamma^m$  that is injective when restricted to  $T_{n,\varepsilon}$ , together with a function  $g_n : \Gamma^m \rightarrow \Sigma^n$  that is chosen so that

$$g_n(f_n(a_1, \dots, a_n)) = a_1 \cdots a_n \quad (120)$$

Thus it holds that  $T_{n,\varepsilon} \subset G_n$  and therefore

$$\sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \geq \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n) \quad (121)$$

It follows that the quantity on the right-hand side is greater than  $1 - \delta$  for sufficiently large values of  $n$ .

2. Assume  $\alpha < H(p)$ .

$$|G_n| \leq 2^m = 2^{\lfloor \alpha n \rfloor} \quad (122)$$

Then we prove

$$\lim_{n \rightarrow \infty} \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) = 0 \quad (123)$$

Since

$$G_n \subset (\Sigma^n \setminus T_{n,\varepsilon}) \cup (G_n \cap T_{n,\varepsilon}) \quad (124)$$

$$\sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \leq \left( 1 - \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n) \right) + 2^{-n(H(p)-\varepsilon)} |G_n| \quad (125)$$

Choosing  $\varepsilon > 0$  so that  $\alpha < H(p) - \varepsilon$ , one has

$$\lim_{n \rightarrow \infty} 2^{-n(H(p)-\varepsilon)} = 0 \quad (126)$$

■

#### IV. QUANTUM SOURCE CODING

**Definition 17**  $\Gamma = \{0, 1\}$ ,  $n \in \mathbb{N}^*$ ,  $\alpha > 0$ ,  $\delta \in (0, 1)$ ,  $m = \lfloor \alpha n \rfloor$

$$\Phi \in \text{Chan}(\mathcal{X}^n, \mathcal{Y}^m) \quad \Psi \in \text{Chan}(\mathcal{Y}^m, \mathcal{X}^n) \quad (127)$$

is said to be an  $(n, \alpha, \delta)$ -coding scheme for  $\rho \in \mathcal{D}(\mathcal{X})$  if it holds that

$$F(\Psi\Phi, \rho^{\otimes n}) > 1 - \delta \quad (128)$$

**Theorem 18** *Shumacher's source coding theorem.* Let  $\text{Scheme}(\rho, n, \alpha, \delta)$  denote the set of  $(n, \alpha, \delta)$ -coding schemes for  $\rho$ . Fix  $\alpha$ ,  $\rho$  and  $\delta$

1. If  $\alpha > H(p)$ , then

$$\exists N \in \mathbb{N} \ (n \geq N \implies \text{Scheme}(p, n, \alpha, \delta) \neq \emptyset) \quad (129)$$

2. If  $\alpha < H(p)$ , then

$$\forall N \in \mathbb{N} \ (\exists n \geq N \ \text{Scheme}(p, n, \alpha, \delta) = \emptyset) \quad (130)$$

**Proof.** Spectrum decomposition of  $\rho$

$$\rho = \sum_{a \in \Sigma} p(a) u_a u_a^* \quad (131)$$

1. Assume  $\alpha > H(p)$ . For a given choice of  $n > 1/\varepsilon$ , the quantum coding scheme  $(\Phi_n, \Psi_n)$  is defined as follows. First, consider the set of  $\varepsilon$ -typical strings associated with the probability vector  $p$ , and define a projection operator

$$\Pi_{n,\varepsilon} = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}(p)} u_{a_1} u_{a_1}^* \otimes \cdots \otimes u_{a_n} u_{a_n}^* \quad (132)$$



The subspace upon which this operator projects is the  $\varepsilon$ -typical subspace of  $\mathcal{X}^{\otimes n}$  with respect to  $\rho$ . Notice that

$$\langle \Pi_{n,\varepsilon}, \rho^{\otimes n} \rangle = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \quad (133)$$

Now, by Shannon's source coding theorem, there exists a classical coding scheme  $(f_n, g_n)$  for  $p$  that satisfies

$$g_n(f_n(a_1 \cdots a_n)) = a_1 \cdots a_n \quad a_1 \cdots a_n \in T_{n,\varepsilon}(p) \quad (134)$$

Define a linear operator of the form  $A_n \in \text{Lin}(\mathcal{X}^{\otimes n}, \mathcal{Y}^{\otimes m})$  as

$$A_n = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}(p)} e_{f_n(a_1, \dots, a_n)}(u_{a_1} \otimes \cdots \otimes u_{a_n})^* \quad (135)$$

Finally, define channels  $\Phi_n$  and  $\Psi_n$  of the form as

$$\begin{aligned} \Phi_n(X) &= A_n X A_n^* + \langle \mathbb{1} - A_n^* A_n, X \rangle \sigma \\ \Psi_n(X) &= A_n^* X A_n + \langle \mathbb{1} - A_n A_n^*, Y \rangle \xi \end{aligned}$$

where  $\sigma \in \text{D}(\mathcal{Y}^{\otimes m})$  and  $\xi \in \text{D}(\mathcal{X}^{\otimes n})$  chosen arbitrarily.

It holds that

$$F(\Psi_n \Phi_n, \rho^{\otimes n}) \geq \langle \rho^{\otimes n}, A_n^* A_n \rangle = \langle \rho^{\otimes n}, \Pi_{n,\varepsilon} \rangle \quad (136)$$

It follows that the quantity on the right-hand side is greater than  $1 - \delta$  for sufficiently large values of  $n$ .

2. Suppose the Kraus representation of  $\Psi\Phi$  is

$$(\Psi_n \Phi_n)(X) = \sum_{jk} (B_k A_j) X (B_k A_j)^* \quad (137)$$

Notice that

$$\text{rank}(B_k A_j) \leq \dim(\mathcal{Y}^{\otimes m}) = 2^m \quad (138)$$

one may choose a projection operator  $\Pi_k \in \text{Proj}(\mathcal{X}^{\otimes n})$  with  $\text{rank}(\Pi_k) \leq 2^m$  such that  $\Pi_k B_k = B_k$ . Therefore

$$\begin{aligned} F(\Psi_n \Phi_n, \rho^{\otimes n})^2 &= \sum_{jk} |\langle B_k A_j, \rho^{\otimes n} \rangle|^2 \\ &= \sum_{jk} |\langle \Pi_k B_k A_j, \rho^{\otimes n} \rangle|^2 \\ &= \sum_{jk} |\langle B_k A_j \sqrt{\rho^{\otimes n}}, \Pi_k \sqrt{\rho^{\otimes n}} \rangle|^2 \\ &\leq \sum_{jk} \text{Tr}(B_k A_j \rho^{\otimes n} A_j^* B_k^*) \langle \Pi_k, \rho^{\otimes n} \rangle \end{aligned}$$

Then the following completes the proof

$$\sum_{jk} \text{Tr}(B_k A_j \rho^{\otimes n} A_j^* B_k^*) = 1 \quad (139)$$

$$\langle \Pi_k, \rho^{\otimes n} \rangle \leq \sum_{i=1}^{2^m} \lambda_i(\rho^{\otimes n}) = \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \quad (140)$$

for some subset  $G_n \subset \Sigma^n$  having size at most  $2^m$ . ■

## V. ACCESSIBLE INFORMATION

**Definition 18** *Classical communications.* Let  $p \in \mathcal{P}(\Sigma)$  be the classical information source. The classical communication channel is characterized by conditional probabilities.

$$\{p(Y = b | X = a) : a \in \Sigma, b \in \Gamma\} \quad (141)$$

The channel capacity of the classical channel is given by

$$C = \max_{p[X]} I(X : Y) \quad (142)$$

**Definition 19** *Encoding classical information into quantum states.* Let  $X$  and  $Z$  be classical registers having classical state sets  $\Sigma$  and  $\Gamma$ , respectively, and let  $Y$  be a register. Also let  $p \in \mathcal{P}(\Sigma)$  be a probability vector, let

$$\{\rho_a : a \in \Sigma\} \subset \mathcal{D}(\mathcal{Y}) \quad (143)$$

$$\mu : \Gamma \rightarrow \text{Pos}(\mathcal{Y}) \quad (144)$$

Alice obtains an element  $a \in \Sigma$ , stored in the register  $X$ , that has been randomly generated by a source according to the probability vector  $p$ . She prepares  $Y$  in the state  $\rho_a$  and sends  $Y$  to Bob. Bob measures  $Y$  with respect to the measurement  $\mu$ , and stores the outcome of this measurement in the classical register  $Z$ . This measurement outcome represents information that Bob has obtained regarding the classical state of  $X$ . Then the pair  $(X, Z)$  will be left in the probabilistic state  $q \in \mathcal{P}(\Sigma \times \Gamma)$  defined by

$$q(a, b) = p(a) \langle \mu(b), \rho_a \rangle \quad (145)$$

Define the ensemble

$$\eta(a) = p(a) \rho_a \quad (146)$$

The probability vector  $q$  may be expressed as

$$q(a, b) = \langle \mu(b), \eta(a) \rangle \quad (147)$$

The notation  $I_\mu(\eta)$  will denote the mutual information between  $X$  and  $Z$ , with respect to a probabilistic state defined in this way, so that

$$I_\mu(\eta) = H(q[X]) + H(q[Z]) - H(q) = D(q \| q[X] \otimes q[Z]) \quad (148)$$

The accessible information  $I_{\text{acc}}(\eta)$  of the ensemble  $\eta$  is defined as the supremum value, ranging over all possible choices of a measurement  $\mu$ , that may be obtained in this way.

$$I_{\text{acc}}(\eta) = \sup_{\mu} I_\mu(\eta) \quad (149)$$

**Lemma 3**  $\eta : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  is an ensemble of states.  $\mu_0, \mu_1 : \Gamma \rightarrow \text{Pos}(\mathcal{Y})$  be measurements,  $\lambda \in [0, 1]$

$$I_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) \leq \lambda I_{\mu_0}(\eta) + (1-\lambda) I_{\mu_1}(\eta) \quad (150)$$

**Proof.** Define

$$\begin{aligned} p(a) &= \text{Tr}(\eta(a)) \\ q_1(a, b) &= \langle \mu_1(b), \eta(a) \rangle \\ q_2(a, b) &= \langle \mu_2(b), \eta(a) \rangle \end{aligned}$$

$$\begin{aligned} I_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) &= D(\lambda q_0 + (1-\lambda)q_1 \| p \otimes \lambda q_0[Z] + (1-\lambda)q_1[Z]) \\ &\leq \lambda D(q_0 \| p \otimes q_0[Z]) + (1-\lambda) D(q_1 \| p \otimes q_1[Z]) \\ &= \lambda I_{\mu_0}(\eta) + (1-\lambda) I_{\mu_1}(\eta) \end{aligned}$$

■

**Theorem 19**  $\eta : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  is an ensemble of states.  $\exists(\mu : \Gamma \rightarrow \text{Pos}(\mathcal{Y})) \begin{cases} |\Gamma| \leq \dim(\mathcal{Y})^2 \\ I_\mu(\eta) = I_{acc}(\eta) \end{cases}$

**Proof.** Let  $\nu : \Lambda \rightarrow \text{Pos}(\mathcal{Y})$  be an measurement. Since  $I_\mu(\eta)$  is convex on the set of measurement of the form  $\mu : \Lambda \rightarrow \text{Pos}(\mathcal{Y})$ . There exists an extreme measurement  $\mu : \Lambda \rightarrow \text{Pos}(\mathcal{Y})$  satisfying  $I_\mu(\eta) \geq I_\nu(\eta)$ .  
 $\mu$  is extremal implies

$$|\{a \in \Lambda : \mu(a) \neq 0\}| \leq \dim(\mathcal{Y})^2 \quad (151)$$

It follows that  $I_{acc}(\eta)$  is equal to the supremum value of  $I_\mu(\eta)$ , ranging over all measurements  $\mu$  having  $\dim(\mathcal{Y})^2$  measurement outcomes. So the supremum is taken over an compact set. This complete the proof. ■

**Definition 20** The Holevo information. Let  $\eta : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  be an ensemble.  $\sigma \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$  is a classical-quantum state

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a) \quad (152)$$

Holevo  $\chi$ -quantity is defined as

$$\chi(\eta) = I(\mathbf{X} : \mathbf{Y}) = H\left(\sum_{a \in \Sigma} \eta(a)\right) - \sum_{a \in \Sigma, \eta(a) \neq 0} \text{Tr}(\eta(a)) H\left(\frac{\eta(a)}{\text{Tr}(\eta(a))}\right) \quad (153)$$

**Theorem 20** (Convexity) Let  $\eta_0 : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  and  $\eta_1 : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  be ensembles of states. Suppose further that at least one of the following two conditions is satisfied:

1. The ensembles  $\eta_0$  and  $\eta_1$  have the same average state:

$$\sum_{a \in \Sigma} \eta_0(a) = \sum_{a \in \Sigma} \eta_1(a) = \rho \quad (154)$$

2. The ensembles  $\eta_0$  and  $\eta_1$  correspond to the same probability distribution, over possibly different states:

$$\text{Tr}(\eta_0(a)) = \text{Tr}(\eta_1(a)) = p(a) \quad (155)$$

Then for  $\lambda \in [0, 1]$ , it holds that

$$\chi(\lambda\eta_0 + (1 - \lambda)\eta_1) \leq \lambda\chi(\eta_0) + (1 - \lambda)\chi(\eta_1) \quad (156)$$

**Proof.** For condition 1

$$\sigma_0 = \sum_{a \in \Sigma} E_{a,a} \otimes \eta_0(a) \quad \sigma_1 = \sum_{a \in \Sigma} E_{a,a} \otimes \eta_1(a) \quad \sigma = \sum_{a \in \Sigma} E_{a,a} \otimes (\lambda\eta_0 + (1 - \lambda)\eta_1) \quad (157)$$

Then

$$\chi(\eta_0) = D(\sigma_0 \| \sigma_0[\mathbf{X}] \otimes \rho) \quad \chi(\eta_1) = D(\sigma_1 \| \sigma_1[\mathbf{X}] \otimes \rho) \quad (158)$$

$$\begin{aligned} \chi(\lambda\eta_0 + (1 - \lambda)\eta_1) &= D(\sigma \| \sigma[\mathbf{X}] \otimes \rho) \\ &= D(\lambda\sigma_0 + (1 - \lambda)\sigma_1 \| (\lambda\sigma_0[\mathbf{X}] + (1 - \lambda)\sigma_1[\mathbf{Y}]) \otimes \rho) \\ &\leq \lambda\chi(\eta_0) + (1 - \lambda)\chi(\eta_1) \end{aligned}$$

For condition 2

$$\sigma_0 = \sum_{a \in \Sigma} p(a) E_{a,a} \otimes \rho_{1,a} \quad \sigma_1 = \sum_{a \in \Sigma} p(a) E_{a,a} \otimes \rho_{2,a} \quad \sigma = \sum_{a \in \Sigma} p(a) E_{a,a} \otimes (\lambda\rho_{1,a} + (1 - \lambda)\rho_{2,a}) \quad (159)$$

Then

$$\chi(\eta_0) = D(\sigma_0 \| \text{diag}(p) \otimes \rho_0) \quad \chi(\eta_1) = D(\sigma_1 \| \text{diag}(p) \otimes \rho_1) \quad (160)$$

$$\begin{aligned}
\chi(\lambda\eta_0 + (1-\lambda)\eta_1) &= D(\sigma\|\sigma[\mathbf{X}] \otimes \rho) \\
&= D(\lambda\sigma_0 + (1-\lambda)\sigma_1\|\text{diag}(p) \otimes (\lambda\rho_0 + (1-\lambda)\rho_1)) \\
&\leq \lambda\chi(\eta_0) + (1-\lambda)\chi(\eta_1)
\end{aligned}$$

■

**Theorem 21** (Concavity) Let  $\eta_0 : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  and  $\eta_1 : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  be ensembles of states. Suppose further that

$$\frac{\eta_0(a)}{\text{Tr}(\eta_0(a))} = \frac{\eta_1(a)}{\text{Tr}(\eta_1(a))} \quad (161)$$

Then

$$\lambda\chi(\eta_0) + (1-\lambda)\chi(\eta_1) \leq \chi(\lambda\eta_0 + (1-\lambda)\eta_1) \quad (162)$$

**Theorem 22** (Holevo's theorem) Let  $\eta : \Sigma \rightarrow \text{Pos}(\mathcal{Y})$  be an ensemble of states. It holds that

$$I_{\text{acc}}(\eta) \leq \chi(\eta) \quad (163)$$

**Proof.**

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a) \quad (164)$$

$$\chi(\eta) = D(\sigma\|\sigma[\mathbf{X}] \otimes \sigma[\mathbf{Y}]) \quad (165)$$

let  $\mu : \Gamma \rightarrow \text{Pos}(\mathcal{Y})$  be a measurement. The corresponding quantum-to-classical channel is  $\Phi \in \text{Chan}(\mathcal{Y}, \mathcal{Z})$

$$\Phi(Y) = \sum_{b \in \Gamma} \langle \mu(b), Y \rangle E_{b,b} \quad (166)$$

Then

$$I_\mu(\eta) = D((\mathbb{1}_{\text{Lin}(\mathcal{X})} \otimes \Phi)(\sigma) \| (\mathbb{1}_{\text{Lin}(\mathcal{X})} \otimes \Phi)(\sigma[\mathbf{X}] \otimes \sigma[\mathbf{Y}])) \quad (167)$$

As the quantum relative entropy does not increase under the action of a channel, it follows

$$I_{\text{acc}}(\eta) \leq \chi(\eta) \quad (168)$$

■

**Definition 21** Holevo information of a quantum channel. Let  $\eta : \Sigma \rightarrow \text{Pos}(\mathcal{Y}_1)$  be an ensemble.  $\sigma \in D(\mathcal{X} \otimes \mathcal{Y}_1)$ ,  $\omega \in D(\mathcal{X} \otimes \mathcal{Y}_2)$  are classical-quantum states.  $\Phi \in \text{Chan}(\mathcal{Y}_1, \mathcal{Y}_2)$

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a) \quad (169)$$

$$\omega = (\mathbb{1}_{\text{Lin}(\mathcal{X})} \otimes \Phi)(\sigma) = \sum_{a \in \Sigma} E_{a,a} \otimes \Phi(\eta) \quad (170)$$

The Holevo information  $\chi(\Phi)$  of a channel  $\Phi$  is a measure of the classical correlations that Alice can establish with Bob

$$\chi(\Phi) = \max_{\eta} I(\mathbf{X} : \mathbf{Y}_2) \quad (171)$$

**Theorem 23** It is sufficient to maximize the Holevo information with respect to pure states  $\frac{\eta(a)}{\text{Tr}(\eta(a))} : a \in \Sigma$

$$\chi(\Phi) = \max_{\eta} I(\mathbf{X} : \mathbf{Y}_2) \quad (172)$$

**Proof.** Let  $\eta(a) = p(a)\rho_a$  be an ensemble. Introduce another classical register  $Z$ .  $\sigma \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Z} \otimes \mathcal{Y}_2)$

$$\sigma = \sum_{a \in \Sigma, b \in \Gamma} p(a, b) E_{a,a} \otimes E_{b,b} \otimes \Phi(\xi_{ab}) \quad (173)$$

where  $\xi_{a,b}$  are the eigenvectors of  $\rho_a$ . That is, the spectrum decomposition of  $\rho_a$  is

$$\rho_a = \sum_{b \in \Gamma} \frac{p(a, b)}{p(a)} \xi_{ab} \quad (174)$$

Thus we get a pure ensemble  $\eta' : \Sigma \times \Gamma \rightarrow \text{Pos}(\mathcal{X})$

$$\chi(\eta) = I(\sigma[X] : \sigma[Y_2]) \leq I(\sigma[X, Z] : \sigma[Y_2])$$

■

#### References

<https://arxiv.org/abs/1306.3142>

<https://warwick.ac.uk/fac/sci/math/research/events/2013-2014/statmech/su/Nilanjana-slides.pdf>

<https://arxiv.org/abs/quant-ph/0512258>