Entropy and Source coding

I. DEFINITIONS OF CLASSICAL ENTROPIC FUNCTIONS

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

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

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

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

Remark 1 *Note that we define*

$$0\log 0 = \lim_{x\to 0^+} x\log x = 0 \tag{3}$$

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

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

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

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

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

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

Property 1

1. *The relationship between* η *and* θ*:*

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

$$\eta(\alpha) = -\theta(\alpha, 1) \tag{7}$$

2. *The Shannon entropy function may be expressed in terms of the* η*-function as follows:*

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

3. *Expressed in terms of the* θ*-function*

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

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

$$\mathrm{H}(u) = -\,\mathrm{D}(u\|\mathbb{1}) \tag{10}$$

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

5. *Properties of η: concavity*

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

$$\eta(\lambda\alpha + (1-\lambda)\beta) \geq \lambda\eta(\alpha) + (1-\lambda)\eta(\beta) \tag{13}$$

6. *Properties of θ: subadditivity*

$$\theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1) \geq \theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1) \tag{14}$$

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

$$\mathrm{H}(\lambda u + (1-\lambda)v) \geq \lambda\,\mathrm{H}(u) + (1-\lambda)\,\mathrm{H}(v) \tag{15}$$

(Hint: concavity of η)

8. *Subadditivity of* D

$$\mathrm{D}(u_0\|v_0) + \mathrm{D}(u_1\|v_1) \geq \mathrm{D}(u_0+u_1\|v_0+v_1) \tag{16}$$

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 θ)

9. *Joint convexity of* D

$$\lambda\,D(u_0\|v_0) + (1-\lambda)D(u_1\|v_1) \geq \mathrm{D}(\lambda u_0 + (1-\lambda)u_1\|\lambda v_0 + (1-\lambda)v_1) \tag{17}$$

10. *$\oplus$ and $\otimes$*

$$\mathrm{H}(u \oplus v) = \mathrm{H}(u) + \mathrm{H}(v) \tag{18}$$

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

Proof.

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

■

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

$$\mathrm{H}(\alpha p) = \alpha\,\mathrm{H}(p) - \alpha\log\alpha \tag{20}$$

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

$$\mathrm{D}(\alpha p \| \beta q) = \alpha\, \mathrm{D}(p\|q) + \alpha \log \frac{\alpha}{\beta} \tag{21}$$

13. *Non-negativity of* D

$$\sum_{a\in\Sigma} u(a) \geq \sum_{a\in\Sigma} v(a) \implies \mathrm{D}(u\|v) \geq 0 \tag{22}$$

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

$$\mathrm{D}(u\|v) = \frac{1}{\ln 2} \sum_{a\in\Sigma} \theta(u(a), v(a)) \geq \frac{1}{\ln 2} \theta(\sum_{a\in\Sigma} u(a), \sum_{a\in\Sigma} v(a)) \geq 0 \tag{23}$$

■

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

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

$$0 \leq \mathrm{H}(u) + \alpha \log \alpha \leq \alpha \log(|\Sigma|) \tag{25}$$

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

In particular, $0 \leq \mathrm{H}(p) \leq \log(|\Sigma|)$

Proof.

$$\mathrm{D}(p\| \frac{\mathbb{1}}{|\Sigma|}) = \log|\Sigma| - \mathrm{H}(p) \geq 0 \implies \mathrm{H}(p) \leq \log|\Sigma| \tag{26}$$

■

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

Definition 4 *Conditional Shannon entropy*

$$\mathrm{H}(\mathsf{X}|\mathsf{Y}) = \mathrm{H}(\mathsf{X},\mathsf{Y}) - \mathrm{H}(\mathsf{Y}) = -\,\mathrm{D}(p\|\mathbb{1} \otimes p[\mathsf{Y}]) = \log|\Sigma| - \mathrm{D}(p\|\frac{\mathbb{1}}{|\Sigma|} \otimes p[\mathsf{Y}]) \tag{27}$$

where is the alphabet of register X.

Definition 5 *Mutual information*

$$\mathrm{I}(\mathsf{X},\mathsf{Y}) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) = \mathrm{D}(p\|p[\mathsf{X}] \otimes p[\mathsf{Y}]) \tag{28}$$

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*

$$\mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) \tag{29}$$

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

(Hint: $\mathrm{I}(\mathsf{X}:\mathsf{Y}) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) = \mathrm{D}(p\|p[\mathsf{X}] \otimes p[\mathsf{Y}])$*)*

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

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{X}, \mathsf{Y}) \tag{30}$$

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.

$$\mathrm{H}(\mathsf{X}, \mathsf{Y}) = -\sum_{ab} p(a,b) \log p(a,b) \geq -\sum_{a} (\sum_{b} p(a,b)) \log(\sum_{b} p(a,b)) = \mathrm{H}(\mathsf{X}) \tag{31}$$

■

3. *(Strong subadditivity)* $\mathrm{H}(\mathsf{X}, \mathsf{Y}, \mathsf{Z}) + \mathrm{H}(\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}, \mathsf{Y}) + \mathrm{H}(\mathsf{Y}, \mathsf{Z})$ *with equality iff* $\mathsf{Z} \to \mathsf{Y} \to \mathsf{X}$ *forms a Markov chain. That is*

$$\forall a, b, c \quad p[\mathsf{X} = a | \mathsf{Y} = b, \mathsf{Z} = c] = p[\mathsf{X} = a | \mathsf{Y} = b] \tag{32}$$

And strong subadditivity is equivalent to

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

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

$$|\,\mathrm{H}(p_0) - \mathrm{H}(p_1)| \leq \lambda \log(|\Sigma| - 1) + \mathrm{H}(\lambda, 1 - \lambda) \tag{33}$$

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

Proof. Define

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

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

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

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

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

$$\begin{aligned} 0 \leq \mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) &= \sum_{a \in \Sigma} p_0(a)\,\mathrm{H}\left(\frac{u_0(a)}{p_0(a)}, \frac{w(a)}{p_0(a)}\right) \leq \mathrm{H}(\lambda, 1 - \lambda) \\ 0 \leq \mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) &= \sum_{a \in \Sigma} p_1(a)\,\mathrm{H}\left(\frac{u_1(a)}{p_1(a)}, \frac{w(a)}{p_1(a)}\right) \leq \mathrm{H}(\lambda, 1 - \lambda) \end{aligned}$$

$$\begin{aligned} &\Longrightarrow |(\mathrm{H}(u_0) - \mathrm{H}(u_1)) - (\mathrm{H}(p_0) - \mathrm{H}(p_1))| \leq \mathrm{H}(\lambda, 1 - \lambda) \\ &\Longrightarrow |\,\mathrm{H}(p_0) - \mathrm{H}(p_1)| - |\,\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \mathrm{H}(\lambda, 1 - \lambda) \end{aligned}$$

Then we need to prove that

$$|\,\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \lambda \log(|\Sigma| - 1)$$

Just let Γ be a proper subset of Σ.

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

Thus

$$|\,\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \lambda \log(|\Gamma|) \leq \lambda \log(|\Sigma| - 1) \tag{39}$$

■

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

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

The equality is saturated iff $p_0 = p_1$.

Proof. The following formula is useful.

$$\theta(\alpha, \beta) + \theta(1 - \alpha, 1 - \beta) \geq 2(\alpha - \beta)^2 \tag{41}$$

Define

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

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

We have

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

And

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

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

Definition 6 *Rényi function is defined as*

$$\mathrm{H}_\alpha(u) = \frac{1}{1 - \alpha} \log \frac{\sum_{a \in \Sigma} u(a)^\alpha}{\sum_{a \in \Sigma} u(a)} \tag{44}$$

Property 2 *Special cases:*

1. *$\alpha \to 0$: Zero entropy*

$$\mathrm{H}_0(u) = \log \frac{|\Sigma|}{\sum_{a \in \Sigma} u(a)} \tag{45}$$

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

$$\mathrm{H}_{\max}(u) = 2\log\frac{\sum_{a\in\Sigma}\sqrt{u(a)}}{\sum_{a\in\Sigma}u(a)} \tag{46}$$

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

$$\begin{aligned}\lim_{\alpha\to 1}\mathrm{H}_\alpha(u) &= \lim_{\alpha\to 1}\frac{1}{1-\alpha}\log\frac{\sum_{a\in\Sigma}u(a)^\alpha}{\sum_{a\in\Sigma}u(a)}\\ &= \lim_{\alpha\to 1} -\frac{\sum_{a\in\Sigma}u(a)}{\sum_{a\in\Sigma}u(a)^\alpha}\sum_{a\in\Sigma}u(a)^\alpha\log u(a)\\ &= -\sum_{a\in\Sigma}u(a)\log u(a)\end{aligned}$$

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

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

$$\mathrm{H}_{\min}(u) = -\log\frac{\max_{a\in\Sigma}u(a)}{\sum_{a\in\Sigma}u(a)} \tag{47}$$

Definition 7

$$\mathrm{D}_\alpha(u\|v) = \frac{1}{\alpha-1}\log\frac{\sum_{a\in\Sigma}u(a)^\alpha v(a)^{1-\alpha}}{\sum_{a\in\Sigma}u(a)} \tag{48}$$

It is easy to see that

$$\mathrm{H}_\alpha(u) = -\,\mathrm{D}_\alpha(u\|\mathbb{1}) \tag{49}$$

We can define conditional entropy and mutual information this way:

$$\mathrm{H}_\alpha(\mathsf{X}|\mathsf{Y}) = -\min_{q\in\mathcal{P}(\Gamma)}\mathrm{D}_\alpha(p\|\mathbb{1}\otimes q) \tag{50}$$

$$\mathrm{I}_\alpha(\mathsf{X}:\mathsf{Y}) = \min_{q\in\mathcal{P}(\Gamma)}\mathrm{D}_\alpha(p\|p[\mathsf{X}]\otimes q) \tag{51}$$

II. DEFINITIONS OF QUANTUM ENTROPIC FUNCTIONS

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

$$\boxed{\mathrm{H}(P) = \mathrm{H}(\lambda(P)) = -\,\mathrm{H}(P\log P)} \tag{52}$$

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

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

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*

$$P = \sum_{a\in\Sigma} \lambda_a x_a x_a^* \qquad Q = \sum_{b\in\Sigma} \mu_b y_b y_b^* \tag{54}$$

Then

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

Hint

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

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

$$\mathrm{H}(VPV^*) = \mathrm{H}(P) \qquad \mathrm{D}(VPV^*\|VQV^*) = \mathrm{D}(P\|Q) \tag{57}$$

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

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

$$\mathrm{H}(P \otimes Q) = \mathrm{Tr}(Q)\,\mathrm{H}(P) + \mathrm{Tr}(P)\,\mathrm{H}(Q) \tag{59}$$

$$\mathrm{D}(P_0 \otimes P_1 \| Q_0 \otimes Q_1) = \mathrm{Tr}(P_1)\,\mathrm{D}(P_0\|Q_0) + \mathrm{Tr}(P_0)\,\mathrm{D}(P_1\|Q_1) \tag{60}$$

$$\mathrm{H}(\alpha\rho) = \alpha\,\mathrm{H}(\rho) - \alpha \log \alpha \tag{61}$$

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

4. non-negativity of quantum relative entropy

$$\mathrm{Tr}(P) \geq \mathrm{Tr}(Q) \implies \mathrm{D}(P\|Q) \geq 0 \tag{63}$$

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

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

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

5. Concavity

$$\mathrm{H}(\lambda P + (1-\lambda) Q) \geq \lambda\,\mathrm{H}(P) + (1-\lambda)\,\mathrm{H}(Q) \tag{64}$$

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

$$\mathrm{H}\left(\begin{bmatrix} P & \\ & Q \end{bmatrix} \middle\| \begin{bmatrix} \frac{P+Q}{2} & \\ & \frac{P+Q}{2} \end{bmatrix}\right) = 2\,\mathrm{H}\left(\frac{P+Q}{2}\right) - \mathrm{H}(P) - \mathrm{H}(Q) \geq 0$$

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

$$f(\lambda) \geq \lambda f(1) + (1-\lambda) f(0) \tag{65}$$

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

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

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

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

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

6. *Subadditivity*

$$\mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) \tag{68}$$

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

$$\mathrm{D}(\rho \| \rho[\mathsf{X}] \otimes \rho[\mathsf{Y}]) = -\mathrm{H}(\rho) + \mathrm{H}(\rho[\mathsf{X}]) + \mathrm{H}(\rho[\mathsf{Y}]) \geq 0$$

The following formula is useful.

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

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

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

Then

$$\mathrm{H}(\mathsf{X}) = \mathrm{H}(\mathsf{Y}) = \mathrm{H}(p) \tag{71}$$

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

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

The equality is saturated iff $\{\operatorname{Tr}_{\mathcal{Y}} u_a u_a : a \in \Sigma\}$ *have a common eigenbasis and* $\{\operatorname{Tr}_{\mathcal{X}} u_a u_a : a \in \Sigma\}$ *are orthogonal.*

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

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

9. Combining 6 and 8, we have

$$|\,\mathrm{H}(\mathsf{X}) - \mathrm{H}(\mathsf{Y})| \le \mathrm{H}(\mathsf{X}, \mathsf{Y}) \le \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) \tag{73}$$

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

$$\delta = \frac{1}{2}\|\rho_0 - \rho_1\|_1 \tag{74}$$

we have

$$|\,\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \le \delta \log(n-1) + \mathrm{H}(\delta, 1-\delta) \tag{75}$$

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

Proof. The following bound is useful.

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

Define

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

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

$$\tag{79}$$

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

$$\delta = \alpha\delta_0 + (1-\alpha)\delta_1 \tag{80}$$

$$\begin{aligned} |\,\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| &\le \delta_0 \log(n-1) + \mathrm{H}(\delta_0, 1-\delta_0) \\ |\,\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| &\le \delta_1 \log(n-1) + \mathrm{H}(\delta_1, 1-\delta_1) \end{aligned}$$

Thus

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

■

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

$$Q' = \sum_{a\in\Sigma} P_a Q P_a \tag{81}$$

$$\mathrm{H}(Q') \ge \mathrm{H}(Q) \tag{82}$$

Proof.

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

Then

$$\mathrm{H}(Q') - \mathrm{H}(Q) = -\operatorname{Tr}(Q \log Q') + \operatorname{Tr}(Q \log Q) = \mathrm{D}(Q \| Q') \geq 0 \tag{83}$$

■

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

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

Then

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

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

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

$$w = \sum_{a \in \Sigma} \sqrt{p(a)} u_a \otimes e_a \in \mathsf{D}(\mathcal{Y}, \mathcal{Z}) \tag{85}$$

Then

$$\mathrm{H}(\rho) = \mathrm{H}(\mathsf{Y}) = \mathrm{H}(\mathsf{Z}) \tag{86}$$

If perform a projective measurement $\{e_a^* e_a\}$ on $w[\mathsf{Z}]$, the entropy of Z becomes $\mathrm{H}(p)$. Because projection increases entropy, we get

$$\mathrm{H}(\mathsf{Z}) \leq \mathrm{H}(p) \tag{87}$$

Thus, we have proved that when the states ρ_a are pure,

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

If not pure, we can decompose

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

Then

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

■

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

$$\sum_{a\in\Sigma} p(a)\,\mathrm{H}(\rho_a) \le \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) \le \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\rho_a) + \mathrm{H}(p) \tag{90}$$

Theorem 8

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

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

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

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

Proof.

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

■

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

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

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 $f(\alpha) \ge 0$ for every $\alpha \in [0,1]$. $f(0) \ge 0$ and $f(1) \ge 0$ are trivial.

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

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

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

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

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

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

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

$$\mathrm{D}(P_0 + P_1 \| Q_0 + Q_1) \leq \mathrm{D}(P_0 \| Q_0) + \mathrm{D}(P_1 \| Q_1) \tag{95}$$

Proof.

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

■

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

$$\mathrm{D}(\lambda P_0 + (1-\lambda) P_1 \| \lambda Q_0 + (1-\lambda) Q_1) \leq \lambda\, \mathrm{D}(P_0 \| Q_0) + (1-\lambda)\, \mathrm{D}(P_1 \| Q_1) \tag{96}$$

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

$$\mathrm{D}(\Phi(P) \| \Phi(Q)) \leq \mathrm{D}(P \| Q) \tag{97}$$

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

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

■

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

$$\mathrm{H}(\mathsf{X}, \mathsf{Y}, \mathsf{Z}) + \mathrm{H}(\mathsf{Z}) \leq \mathrm{H}(\mathsf{X}, \mathsf{Z}) + \mathrm{H}(\mathsf{Y}, \mathsf{Z}) \tag{98}$$

This is equivalent to

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

Proof.

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

■

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

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

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

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

$$\|p_0 - p_1\|_1 = \|\rho_0 - \rho_1\|_1 \tag{100}$$

By monotonicity of quantum relative entropy, we get

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

■

Definition 11 *Rényi entropies*

$$\mathrm{H}_\alpha(P) = \frac{1}{1-\alpha}\log\frac{\mathrm{Tr}(\rho^\alpha)}{\mathrm{Tr}(\rho)} \tag{102}$$

Definition 12 *Quantum divergences have multiple definitions*

$$\widetilde{\mathrm{D}}_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\frac{1}{\mathrm{Tr}(P)}\mathrm{Tr}[(Q^{\frac{1}{2\alpha}-\frac{1}{2}}PQ^{\frac{1}{2\alpha}-\frac{1}{2}})^\alpha] = \frac{\alpha}{\alpha-1}\log\frac{1}{\mathrm{Tr}(P)^{1/\alpha}}\left\|Q^{\frac{1}{2\alpha}-\frac{1}{2}}PQ^{\frac{1}{2\alpha}-\frac{1}{2}}\right\|_\alpha$$
$$\mathrm{D}_\alpha(P\|Q) = \frac{1}{\alpha-1}\log\frac{\mathrm{Tr}(P^\alpha Q^{1-\alpha})}{\mathrm{Tr}(P)}$$

It is easy to see that

$$\mathrm{H}_\alpha(P) = -\mathrm{D}_\alpha(P\|\mathbb{1}) = \widetilde{\mathrm{H}}_\alpha(P) = -\widetilde{\mathrm{D}}_\alpha(P\|\mathbb{1}) \tag{103}$$

In particular, we define

$$\mathrm{H}_0(P) = -\mathrm{D}_0(P\|\mathbb{1}) = \log\frac{\mathrm{Tr}(\Pi_P)}{\mathrm{Tr}(P)} \tag{104}$$

Remark 4 *The range of α:*

- $\widetilde{\mathrm{D}}_\alpha$*:* $[\frac{1}{2},\infty)$
- D_α *:* $[0,2]$

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

Theorem 14 $\widetilde{\mathrm{D}}_\alpha(P\|Q) \leq \mathrm{D}_\alpha(P\|Q)$

Property 4 *Special cases*

1. $\alpha = 0$

$$\mathrm{D}_0(P\|Q) = \log\mathrm{Tr}(P) - \log\mathrm{Tr}(\Pi_{\mathsf{im}(P)}Q) \tag{105}$$

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

$$\widetilde{\mathrm{D}}_{\frac{1}{2}}(P\|Q) = -2\log\frac{\mathrm{Tr}[(Q^{1/2}PQ^{1/2})^{1/2}]}{\mathrm{Tr}(P)} = -2\log\frac{\mathrm{F}(P,Q)}{\mathrm{Tr}(P)} \tag{106}$$

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

$$\widetilde{\mathrm{D}}_1(P\|Q) = \mathrm{D}_1(P\|Q) = \mathrm{Tr}(P\log P) - \mathrm{Tr}(P\log Q) \tag{107}$$

4. $\alpha = 2$: collision entropy

5. $\alpha \to \infty$

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

Definition 13 *Conditional entropy*

$$\mathrm{H}_\alpha(\mathsf{X}|\mathsf{Y}) = -\min_{\sigma \in \mathsf{D}(\mathcal{Y})} \mathrm{D}_\alpha(\rho\|\mathbb{1}_\mathcal{X} \otimes \sigma) \tag{109}$$

$$\widetilde{\mathrm{H}}_\alpha(\mathsf{X}|\mathsf{Y}) = -\min_{\sigma \in \mathsf{D}(\mathcal{Y})} \widetilde{\mathrm{D}}_\alpha(\rho\|\mathbb{1}_\mathcal{X} \otimes \sigma) \tag{110}$$

Property 5 *Special cases*

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

$$\begin{aligned}\mathrm{H}_0(\mathsf{X}|\mathsf{Y}) &= -\min_{\sigma \in \mathsf{D}(\mathcal{Y})} \mathrm{D}_0(\rho\|\mathbb{1}_\mathcal{X} \otimes \sigma) \\ &= \max_{\sigma \in \mathsf{D}(\mathcal{Y})} \log \mathrm{Tr}(\Pi_{\mathsf{im}(\rho)}(\mathbb{1}_\mathcal{X} \otimes \sigma)) \\ &= \log \| \mathrm{Tr}_\mathcal{X} \Pi_{\mathsf{im}(\rho)} \|\end{aligned}$$

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

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

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

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

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 $\mathrm{H}_0(p_X)$ 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 $\overline{p}_X$. 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 $\hat{\rho}, \rho \in \mathsf{D}(\mathcal{X} \otimes \mathcal{Y})$ and $\hat{\rho}$ is ε-close to ρ. One definition of distance is the fidelity distance* $\mathrm{F}^2(\hat{\rho}, \rho) \geq 1 - \varepsilon^2$.

$$\mathrm{H}^\varepsilon_{\min}(\mathsf{X}|\mathsf{Y})_\rho = \sup_{\hat{\rho}} \mathrm{H}_{\min}(\mathsf{X}|\mathsf{Y})_{\hat{\rho}} \tag{111}$$

$$\mathrm{H}^\varepsilon_{\max}(\mathsf{X}|\mathsf{Y})_\rho = \inf_{\hat{\rho}} \mathrm{H}_{\max}(\mathsf{X}|\mathsf{Y})_{\hat{\rho}} \tag{112}$$

III. CLASSICAL SOURCE CODING

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

$$f : \Sigma^n \to \Gamma^m \qquad g : \Gamma^m \to \Sigma^n \tag{113}$$

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

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

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

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, α, δ)-coding scheme for p, then the number δ 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 α represents the average number of bits (as the value of n increases) needed to encode each symbol.

Definition 16 *Typical strings. ε-typical*

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

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

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

Theorem 15

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

(Hint: use the weak law of large numbers)

Theorem 16

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

Theorem 17 *Shannon's source coding theorem. Let* Scheme(p, n, α, δ) *denote the set of (n, α, δ)-coding schemes for p. Fix α, p and δ*

1. *If $\alpha > \mathrm{H}(p)$, then*

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

2. *If $\alpha < \mathrm{H}(p)$, then*

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

Proof.

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

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

Then

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

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

$$g_n(f_n(a_1, \cdots, a_n)) = a_1 \cdots a_n \tag{120}$$

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

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

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

2. Assume $\alpha < \mathrm{H}(p)$.

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

Then we prove

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

Since

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

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

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

$$\lim_{n\to\infty} 2^{-n(\mathrm{H}(p)-\varepsilon)} = 0 \tag{126}$$

■

IV. QUANTUM SOURCE CODING

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

$$\Phi \in \mathsf{Chan}(\mathcal{X}^n, \mathcal{Y}^m) \qquad \Psi \in \mathsf{Chan}(\mathcal{Y}^m, \mathcal{X}^n) \tag{127}$$

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

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

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

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

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

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

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

Proof. Spectrum decomposition of ρ

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

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

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

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

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

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

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

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

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

Finally, define channels Φ_n and Ψ_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 $\sigma \in \mathrm{D}(\mathcal{Y}^{\otimes m})$ and $\xi \in \mathrm{D}(\mathcal{X}^{\otimes n})$ chosen arbitrarily.
It holds that

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

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

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

Notice that

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

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

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

Then the following completes the proof

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

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

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.*

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

The channel capacity of the classical channel is given by

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

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

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

$$\mu : \Gamma \to \mathsf{Pos}(\mathcal{Y}) \tag{144}$$

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 ρ_a and sends Y to Bob. Bob measures Y with respect to the measurement μ, 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

$$q(a, b) = p(a)\langle \mu(b), \rho_a \rangle \tag{145}$$

Define the ensemble

$$\eta(a) = p(a)\rho_a \tag{146}$$

The probability vector q may be expressed as

$$q(a, b) = \langle \mu(b), \eta(a) \rangle \tag{147}$$

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

$$\mathrm{I}_\mu(\eta) = \mathrm{H}(q[\mathsf{X}]) + \mathrm{H}(q[\mathsf{Z}]) - \mathrm{H}(q) = \mathrm{D}(q \| q[\mathsf{X}] \otimes q[\mathsf{Z}]) \tag{148}$$

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

$$\mathrm{I}_{acc}(\eta) = \sup_\mu \mathrm{I}_\mu(\eta) \tag{149}$$

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

$$\mathrm{I}_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) \leq \lambda\, \mathrm{I}_{\mu_0}(\eta) + (1 - \lambda)\, \mathrm{I}_{\mu_1}(\eta) \tag{150}$$

Proof. Define

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

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

■

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

Proof. Let $\nu : \Lambda \to \mathsf{Pos}(\mathcal{Y})$ be an measurement. Since $\mathrm{I}_\mu(\eta)$ is convex on the set of measurement of the form $\mu : \Lambda \to \mathsf{Pos}(\mathcal{Y})$. There exists an extreme measurement $\mu : \Lambda \to \mathsf{Pos}(\mathcal{Y})$ satisfying $\mathrm{I}_\mu(\eta) \geq \mathrm{I}_\nu(\eta)$.

μ is extremal implies

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

It follows that $\mathrm{I}_{\mathrm{acc}}(\eta)$ is equal to the supremum value of $\mathrm{I}_\mu(\eta)$, ranging over all measurements μ 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 $\eta : \Sigma \to \mathsf{Pos}(\mathcal{Y})$ be an ensemble. $\sigma \in \mathsf{D}(\mathcal{X} \otimes \mathcal{Y})$ is a classical-quantum state*

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

Holevo χ-quantity is defined as

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

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

1. *The ensembles η_0 and η_1 have the same average state:*

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

2. *The ensembles η_0 and η_1 correspond to the same probability distribution, over possibly different states:*

$$\mathrm{Tr}(\eta_0(a)) = \mathrm{Tr}(\eta_1(a)) = p(a) \tag{155}$$

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

$$\chi(\lambda\eta_0 + (1-\lambda)\eta_1) \leq \lambda\chi(\eta_0) + (1-\lambda)\chi(\eta_1) \tag{156}$$

Proof. For condition 1

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

Then

$$\chi(\eta_0) = \mathrm{D}(\sigma_0 \| \sigma_0[\mathsf{X}] \otimes \rho) \qquad \chi(\eta_1) = \mathrm{D}(\sigma_1 \| \sigma_1[\mathsf{X}] \otimes \rho) \tag{158}$$

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

For condition 2

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

Then

$$\chi(\eta_0) = \mathrm{D}(\sigma_0 \| \mathsf{diag}(p) \otimes \rho_0) \qquad \chi(\eta_1) = \mathrm{D}(\sigma_1 \| \mathsf{diag}(p) \otimes \rho_1) \tag{160}$$

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

■

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

$$\frac{\eta_0(a)}{\mathrm{Tr}(\eta_0(a))} = \frac{\eta_1(a)}{\mathrm{Tr}(\eta_1(a))} \tag{161}$$

Then

$$\lambda\chi(\eta_0) + (1-\lambda)\chi(\eta_1) \leq \chi(\lambda\eta_0 + (1-\lambda)\eta_1) \tag{162}$$

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

$$\mathrm{I}_{acc}(\eta) \leq \chi(\eta) \tag{163}$$

Proof.

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

$$\chi(\eta) = \mathrm{D}(\sigma \| \sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}]) \tag{165}$$

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

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

Then

$$\mathrm{I}_{\mu}(\eta) = \mathrm{D}((\mathbb{1}_{\mathsf{Lin}(\mathcal{X})} \otimes \Phi)(\sigma) \| (\mathbb{1}_{\mathsf{Lin}(\mathcal{X})} \otimes \Phi)(\sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}])) \tag{167}$$

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

$$\mathrm{I}_{\mathrm{acc}}(\eta) \leq \chi(\eta) \tag{168}$$

■

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

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

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

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

$$\chi(\Phi) = \max_{\eta} \mathrm{I}(\mathsf{X} : \mathsf{Y}_2) \tag{171}$$

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

$$\chi(\Phi) = \max_{\eta} \mathrm{I}(\mathsf{X} : \mathsf{Y}_2) \tag{172}$$

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

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

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

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

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

$$\chi(\eta) = \mathrm{I}(\sigma[\mathsf{X}] : \sigma[\mathsf{Y}_2]) \leq \mathrm{I}(\sigma[\mathsf{X}, \mathsf{Z}] : \sigma[\mathsf{Y}_2])$$

■

References
https://arxiv.org/abs/1306.3142
https://warwick.ac.uk/fac/sci/maths/research/events/2013-2014/statmech/su/Nilanjana-slides.pdf
https://arxiv.org/abs/quant-ph/0512258