Felix Leditzky

Quantum Channels II

DATA-PROCESSING, RECOVERY CHANNELS, AND QUANTUM MARKOV CHAINS

Compiled by Jacob L. Beckey

Last Update: June 2021

Contents

1

Relative Entropy and Data-Processing

1.1 Motivation: quantum state discrimination

We begin by considering the task of quantum state discrimination. Assume you are given one of two quantum state ρ_0, ρ_1 with equal probability and your goal is to decide which state you got. The strategy will be to perform a measurement[1] on the unknown state, the outcome of which will determine your guess at which state you were given.

[1] In this course, measurements will be described by a positive operator-valued measure (POVM): $\{\Lambda_i\}_{i=1}^k, \Lambda_i \geq 0, \sum_i \Lambda_i = \mathbb{1}_{\mathcal{H}}$. For a given state ρ, POVM gives outcome i with probability $p_i = \mathrm{tr}(\Lambda_i \rho)$.

In the problem of state discrimination, we use a POVM $\Lambda = \Lambda_0, \Lambda_1$ with $\Lambda_1 = \mathbb{1} - \Lambda_0$ and $\Lambda_0 \geq 0$. Thus, the outcome "0" corresponds to ρ_0 and outcome "1" corresponds to ρ_1 with probabilities given by $p_i = \mathrm{tr}(\Lambda_i \sigma)$ and $\sigma \in \{\rho_0, \rho_1\}$.

So, the success probability of correctly identifying the state is given by

$$\begin{aligned} p_{\text{succ}}(1) &= \frac{1}{2}\Pr(\rho_0|\rho_0) + \frac{1}{2}\Pr(\rho_1|\rho_1), && (1.1)\\ &= \frac{1}{2}(\mathrm{tr}\Lambda_0\rho_0 + \mathrm{tr}\Lambda_1\rho_1), && (1.2)\\ &= \frac{1}{2}(\mathrm{tr}\Lambda_0\rho_0 + 1 - \mathrm{tr}\Lambda_0\rho_1), && (1.3)\\ &= \frac{1}{2}\left(1 + \mathrm{tr}\left[\Lambda_0(\rho_0 - \rho_1)\right]\right). && (1.4) \end{aligned}$$

Of course, we want to maximize the success probability with respect to the POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$. That is we want to find

$$p^*_{\text{succ}} = \max_{0 \leq \Lambda_0 \leq \mathbb{1}} p_{\text{succ}}(1) = \frac{1}{2}\left(1 + \max_{0 \leq \Lambda_0 \leq \mathbb{1}} \mathrm{tr}\left[\Lambda_0(\rho_0 - \rho_1)\right]\right). \quad (1.5)$$

Recall that the trace norm $\|X\|_1 = \mathrm{tr}\sqrt{X^\dagger X} = \sum_i s_i(X)$ where the $s_i's$ are the singular values of X.[2] From this norm we can formulate the trace distance Also recall that the trace distance between two quantum states ρ_0, ρ_1 is given as

[2] That is, they are the eigenvalues of $|X| = \sqrt{X^\dagger X}$.

$$T(\rho_0, \rho_1) = \frac{1}{2}\|\rho_0 - \rho_1\|_1. \quad (1.6)$$

The fact that this distance can be formulated as a maximization over POVMs is the content of our first lemma.

Lemma 1. *Let ρ_0, ρ_1 be quantum states, then*

$$\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \max_{0\leq\Lambda\leq\mathbb{1}} tr\left[\Lambda(\rho_0 - \rho_1)\right] \tag{1.7}$$

Proof. $\rho_0 - \rho_1$ is Hermitian, so we can write is spectral decomposition as $\sum_i \lambda_i |i\rangle\langle i|$ where $\lambda_i \in \mathbb{R}$, $\langle i|j\rangle = \delta_{ij}$. Then, let us define

$$P = \sum_{i:\lambda_i\geq 0} \lambda_i |i\rangle\langle i| \geq 0, \tag{1.8}$$

$$Q = \sum_{i:\lambda_i< 0} (-\lambda_i) |i\rangle\langle i| \geq 0, \tag{1.9}$$

and $P - Q := \rho_0 - \rho_1$. Then we have[3]

$$\|\rho_0 - \rho_1\|_1 = \mathrm{tr}|\rho_0 - \rho_1| = \mathrm{tr}|P - Q| = \mathrm{tr}P + \mathrm{tr}Q = 2\mathrm{tr}P, \tag{1.10}$$

[3] Note that here we use $\mathrm{tr}(P-Q) = \mathrm{tr}(\rho_0 - \rho_1) = \mathrm{tr}\rho_0 - \mathrm{tr}\rho_1 = 0$ implies $\mathrm{tr}P = \mathrm{tr}Q$.

which implies $\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \mathrm{tr}P$. We now want to relate this result to $\mathrm{tr}\left[\Lambda(\rho_0 - \rho_1)\right]$ for arbitrary $0 \leq \Lambda \leq \mathbb{1}$. We write

$$\begin{aligned}
\mathrm{tr}\Lambda(\rho_0 - \rho_1) &= \mathrm{tr}\Lambda(P - Q) && (1.11)\\
&\leq \mathrm{tr}\Lambda P && (1.12)\\
&\leq \mathrm{tr}P && (1.13)\\
&= \frac{1}{2}\|\rho_0 - \rho_1\|_1 && (1.14)
\end{aligned}$$

This means that $\max_{0\leq\Lambda_0\leq\mathbb{1}} \mathrm{tr}\Lambda(\rho_0 - \rho_1) \leq \frac{1}{2}\|\rho_0 - \rho_1\|_1$. It remains to be shown that there exists a Λ that achieves this maximum. Set $\Lambda = \Pi_p = \sum_{i:\lambda_i\geq 0} |i\rangle\langle i|$ (the projector onto the support of P). Then, we have

$$\begin{aligned}
\mathrm{tr}\Pi_p(\rho_0 - \rho_1) &= \mathrm{tr}\Pi_p(P - Q) && (1.15)\\
&= \mathrm{tr}\Pi_p P - \mathrm{tr}\Pi_p Q && (1.16)\\
&= \mathrm{tr}P && (1.17)\\
&= \frac{1}{2}\|\rho_0 - \rho_1\|_1 && (1.18)
\end{aligned}$$

□

We see that the success probability of correctly identifying the state is

$$p_{\mathrm{succ}} = \frac{1}{2}(1 + \frac{1}{2}\|\rho_0 - \rho_1\|_1). \tag{1.19}$$

So, if the trace distance is one (zero) , we see success probability is one (one half). *Thus, the trace distance between two states is a measure of distinguishability.*

Proposition 2. *Let* $T(\rho_0, \rho_1) = \frac{1}{2}\|\rho_0 - \rho_1\|_1$. *For any two states* ρ_0, ρ_1 *and a quantum channel* $\mathcal{N}$, *we have*

$$T(\rho_0, \rho_1) \geq T(\mathcal{N}(\rho_0), \mathcal{N}(\rho_1)). \tag{1.20}$$

Proof. We know from Lemma (1) that

$$\frac{1}{2}\|\rho_0 - \rho_1\|_1 = \max_{0 \leq \Lambda \leq 1} \mathrm{tr}\Lambda(\rho_0 - \rho_1). \tag{1.21}$$

Let $\Lambda \geq 0$ with $\Lambda \leq \mathbb{1}$ be optimal for $\frac{1}{2}\|\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)\|_1$:

$$\begin{aligned}
\frac{1}{2}\|\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)\|_1 &= \mathrm{tr}\Lambda(\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)), && (1.22)\\
&= \mathrm{tr}\Lambda\mathcal{N}(\rho_0 - \rho_1), && (1.23)\\
&= \mathrm{tr}\mathcal{N}^\dagger(\Lambda)(\rho_0 - \rho_1). && (1.24)
\end{aligned}$$

[4]Now, we want to show 1) $\mathcal{N}^\dagger(\Lambda) \geq 0$ and 2) $\mathcal{N}^\dagger(\Lambda) \leq \mathbb{1}$.

[4] Where the last equality holds by definition of adjoint map.

1) The first is easy to see by recalling that the adjoint map is completely positive because all quantum channels are completely positive. Positivity is a weaker condition, so in particular we have

$$\Lambda \geq 0 \implies \mathcal{N}^\dagger(\Lambda) \geq 0. \tag{1.25}$$

2) Further, the adjoint map is unital[5], so we have

[5] Recall, unitality means $\mathcal{N}(\mathbb{1}) = \mathbb{1}$.

$$\begin{aligned}
\mathcal{N}^\dagger(\mathbb{1} - \Lambda) &\geq 0, && (1.26)\\
\mathcal{N}^\dagger(\mathbb{1}) - \mathcal{N}^\dagger(\Lambda) &\geq 0, && (1.27)\\
\mathbb{1} &\geq \mathcal{N}^\dagger(\Lambda). && (1.28)
\end{aligned}$$

So, from 1) and 2), we can conclude $\mathcal{N}^\dagger(\Lambda)$ is feasible[6] in

[6] A feasible region is the set of all possible points of an optimization problem that satisfy the problem's constraints.

$$\max_{0 \leq K \leq \mathbb{1}} \mathrm{tr}K(\rho_0 - \rho_1). \tag{1.29}$$

Now that we know the adjoint is achievable, we have

$$\begin{aligned}
\frac{1}{2}\|\rho_0 - \rho_1\|_1 &= \max_{0 \leq K \leq \mathbb{1}} \mathrm{tr}K(\rho_0 - \rho_1), && (1.30)\\
&\geq \mathrm{tr}\mathcal{N}^\dagger(\Lambda)(\rho_0 - \rho_1), && (1.31)\\
&= \frac{1}{2}\|\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)\|_1. && (1.32)
\end{aligned}$$

□

We have thus shown that the trace distance cannot increase when subject to the same noisy process. States will never become more distinguishable after processing. Mathematically,

$$\frac{1}{2}\|\rho_0 - \rho_1\|_1 \geq \frac{1}{2}\|\mathcal{N}(\rho_0) - \mathcal{N}(\rho_1)\|_1. \tag{1.33}$$

1.2 *Error analysis and hypothesis testing*

The probability of success can be expressed

$$p_{\text{succ}} = \frac{1}{2}\Pr(\rho_0|\rho_0) + \frac{1}{2}\Pr(\rho_1|\rho_1). \tag{1.34}$$

Then, the error probability will be given as

$$p_{\text{error}} = 1 - p_{\text{succ}} = \frac{1}{2}\left(\Pr(\rho_1|\rho_0) + \Pr(\rho_0|\rho_1)\right). \tag{1.35}$$

In hypothesis testing, there is a null hypothesis $H_0(\rho_0)$ and an alternative hypothesis $H_1(\rho_1)$. Then, we say a type-1 error[7] is committed when we infer ρ_1 when you actually have ρ_0 (false rejection or false negative). A type-2 error[8] is when you infer ρ_0 when you really have ρ_1 (false acceptance or false positive). There then exists a trade-off between these two errors.

[7] Often denoted by the Greek letter α.

[8] Often denoted by the Greek letter β.

- Symmetric hypothesis testing ("Bayesian"): try and minimize the sum of the two errors. This leads to the trace distance as discussed above.

- Asymmetric hypothesis testing: assume that type-1 error is constant and small. The question is then, how small can I make the type-2 error under this constraint.[9]

[9] In realistic settings we want to avoid false negatives at all costs!

Definition 3. *Let ρ, σ be two quantum states. Let ρ be the null hypothesis and let σ be the alternative hypothesis. Further, let Λ with $0 \leq \Lambda \leq \mathbb{1}$ be a test operator defining a two-element POVM $\{\Lambda, \mathbb{1} - \Lambda\}$. Then we have the following errors*

$$\alpha(\Lambda) = tr\rho(\mathbb{1} - \Lambda) \quad \textit{type-1 error} \tag{1.36}$$
$$\beta(\Lambda) = tr\sigma\Lambda \quad \textit{type-2 error} \tag{1.37}$$

In an information-theoretic setting, we wish to determine $\rho^{\otimes n}$ versus $\sigma^{\otimes n}$ as $n \to \infty$. In this case we have[10]

[10] Note that $\Lambda_n \in B(\mathcal{H}^{\otimes n})$, $\Lambda_n \geq 0$, and $\Lambda_n \leq \mathbb{1}_{\mathcal{H}}^{\otimes n}$

$$\alpha_n(\Lambda_n) = \mathrm{tr}\rho^{\otimes n}(\mathbb{1} - \Lambda_n), \tag{1.38}$$
$$\beta_n(\Lambda_n) = \mathrm{tr}\sigma^{\otimes}\Lambda_n. \tag{1.39}$$

Then, for $\epsilon > 0$, we define

$$\beta_n^*(\epsilon) = \min\{\beta_n(\Lambda_n) : 0 \leq \Lambda_n \leq \mathbb{1}, \alpha_n(\Lambda_n) \leq \epsilon\}. \tag{1.40}$$

The question is, how does $\beta_n^*(\epsilon)$ behave[11] as $n \to \infty$?

[11] For example, $\beta_n^* = f(n) \to 0$, what is f?

Definition 4. *1) For a linear operator $X \in B(\mathcal{H})$, the support of X is defined as*

$$\text{supp} X = (\ker X)^{\perp}. \tag{1.41}$$

If X is Hermitian with spectral decomposition $X = \sum_i \lambda_i |i\rangle \langle i|$, then $\text{supp} X = \text{span}\{|i\rangle : \lambda_i \neq 0\}$. The projection onto $\text{supp} X$ is given by

$$\sum_{i:\lambda_i \neq 0} |i\rangle \langle i| = \lim_{\alpha \to 0} X^{\alpha} = X^0 \tag{1.42}$$

2) Let $\rho \geq 0$, $\text{tr}\rho = 1$, $\sigma \geq 0$. Then, the relative entropy is defined as

$$D(\rho\|\sigma) = \begin{cases} \text{tr}(\rho \log \rho - \rho \log \sigma) & \text{if supp}\rho \subseteq \text{supp}\sigma, \\ \infty & \text{else.} \end{cases} \tag{1.43}$$

This definition brings us to the so-called *quantum Stein's lemma*.[12] The lemma says that for all $\epsilon > 0$,

$$\lim_{n \to \infty} \frac{1}{n} \log \beta_n^*(\epsilon) = -D(\rho\|\sigma). \tag{1.44}$$

Intuitively, we can see $\beta_n^* \approx \exp(-nD(\rho\|\sigma))$. Thus, we have a measure of distinguishability in the asymmetric setting. The larger $D(\rho\|\sigma)$, the better one can distinguish between ρ and σ (decay of β_n^*, optimal type-II error if type-I error is bounded and small). Note that because $D(\rho\|\sigma) \neq D(\sigma\|\rho)$, the relative entropy is not a metric in the rigorous sense. However, if ρ and σ are quantum states, then $D(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.

[12] This result was shown by Hiai and Petz and complemented nicely by Ogawa and Nagaoka.

1.3 *Properties of relative entropy*

Measures of distinguishability between quantum states should be monotonic under quantum operations to match the intuition that noise cannot make quantum states more distinguishable. This is summarized in the following theorem[13].

[13] This is the data-processing inequality for the quantum relative entropy.

Theorem 5. *Let ρ be a quantum state, $\sigma \geq 0$, and $\mathcal{N}$ a quantum channel. Then,*

$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{1.45}$$

We will defer the proof to later and first study the properties of the relative entropy.

Proposition 6. *1) Let ρ be a classical state, and $\sigma \geq 0$ be classical (diagonal with respect to the same basis). That is, for $\rho = \sum_x p_x |x\rangle\langle x|$ and $\sigma = \sum_x q_x |x\rangle\langle x|$, we have*

$$D(\rho\|\sigma) = \sum_x p_x \log \frac{p_x}{q_x} = D(p\|q), \tag{1.46}$$

where the final quantity is the classical Kullback-Leibler divergence.

2) Let ρ, σ be quantum states, then $D(\rho\|\sigma) \geq 0$, and $D(\rho\|\sigma) = 0$ iff $\rho = \sigma$.

3) $D(\rho\|\sigma) = D(V\rho V^\dagger \| V\sigma V^\dagger)$ for an isometry V.

4) For classical-quantum states $\rho_{xA} = \sum_x p_x |x\rangle\langle x|_x \otimes \rho_A^x$ and $\sigma_{xA} = \sum_x p_x |x\rangle\langle x|_X \otimes \sigma_A^x$, we have

$$D(\rho_{xA}\|\sigma_{xA}) = \sum_x p_x D(\rho_A^x\|\sigma_A^x). \tag{1.47}$$

5) Joint convexity: Let $\{\rho_x\}$ be states and $\{\sigma_x\}$ be positive semi-definite operators and $\{\lambda_x\}$ be a probability distribution. Then,

$$D(\sum_x \lambda_x\rho_x \| \sum_x \lambda_x\sigma_x) \leq \sum_x \lambda_x D(\rho_x\|\sigma_x). \tag{1.48}$$

6) Let ρ be a state and $\sigma, \sigma' \geq 0$ with $\sigma \leq \sigma'$, then

$$D(\rho\|\sigma) \geq D(\rho\sigma'). \tag{1.49}$$

Proof. 1) This follows directly from the definition of the relative entropy.

2) ρ, σ are states: use data-processing inequality with respect to $\mathcal{N} = \text{tr}$. We then have

$$\begin{aligned} D(\rho\|\sigma) &\geq D(\text{tr}\rho\|\text{tr}\sigma), & (1.50)\\ &= D(1\|1), & (1.51)\\ &= 0. & (1.52) \end{aligned}$$

We will prove that $D(\rho\|\sigma) = 0$ iff $\rho = \sigma$ later.

3) Now we wish to prove isometric[14] invariance. The proof will be done in two parts. First, note that $V \cdot V^\dagger$ can be viewed as a quantum channel. So, by DPI, we have

[14] (Keep in mind, an isometry is a linear map satisfying $V^\dagger V = I$).

$$D(\rho\|\sigma) \geq D(V\rho V^\dagger \| V\sigma V^\dagger). \tag{1.53}$$

Next, define $\Pi = VV^\dagger$ as the projection onto the image of V. Define $W : B(\mathcal{K}) \to B(\mathcal{H})$ act as

$$W(X) = V^\dagger XV + \mathrm{tr}\left((\mathbb{I} - \Pi)X\right)|0\rangle\langle 0|, \tag{1.54}$$

$$\mathrm{tr}W(X) = \mathrm{tr}V^\dagger XV + \mathrm{tr}((\mathbb{1} - \Pi)X), \tag{1.55}$$

$$= \mathrm{tr}X. \tag{1.56}$$

Further, note $W(VYV^\dagger) = Y$. This is as completely positive, trace preserving map (as one should check). Then the second step of our proof is

$$D(\rho\|\sigma) \geq D(V\rho V^\dagger\|V\sigma V^\dagger), \tag{1.57}$$

$$\geq D(W(V\rho V^\dagger)\|W(V\sigma V^\dagger)), \tag{1.58}$$

$$= D(\rho\|\sigma), \tag{1.59}$$

which completes the proof.

4) cq-states: $\rho_{xA} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_A^x = \bigoplus_x p_x\rho_A^x$. Then, $D(\rho\|\sigma) = \mathrm{tr}\left(\rho(\log\rho - \log\sigma)\right)$. **Insert diagram showing block diagonal form.** Then, we can have

$$\log\rho_{xA} - \log\sigma_{xA} = \sum_x |x\rangle\langle x|_x \otimes (\log\rho_A^x - \log\sigma_A^x). \tag{1.60}$$

But we know that

$$D(\rho_{xA}\|\sigma_{xA}) = \mathrm{tr}\rho_{xA}\left(\log\rho_{xA} - \log\sigma_{xA}\right), \tag{1.61}$$

$$= \mathrm{tr}[(\sum_x p_x |x\rangle\langle x| \otimes \rho_A^x) \tag{1.62}$$

$$\times (\sum_y |y\rangle\langle y| \otimes (\log\rho_A^y - \log\sigma_A^y))], \tag{1.63}$$

$$= \sum_x p_x \mathrm{tr}\left(|x\rangle\langle x| \otimes \rho_A^x(\log\rho_A^x - \log\sigma_A^x)\right), \tag{1.64}$$

$$= \sum_x p_x D(\rho_A^x\|\sigma_A^x). \tag{1.65}$$

This completes the proof.

5) We wish to show joint convexity. The idea is to use item 4) above with $\rho_{xA} = \sum_x \lambda_x |x\rangle\langle x| \otimes \rho_A^x, \sigma_{xA} = \sum_x \lambda_x |x\rangle\langle x| \otimes \sigma_A^x$. Then from DPI with respect to tr_X, we get the result we want.

6) Finally, we wish to show that if $\sigma \leq \sigma'$, then $D(\rho\|\sigma) \geq D(\rho\|\sigma')$. Using 4) we can write $D(\rho\|\sigma) = D(\rho \otimes |0\rangle\langle 0| \,\|\sigma \otimes |0\rangle\langle 0| + (\sigma' - \sigma) \otimes |1\rangle\langle 1|)$. Then by DPI, we have

$$D(\rho\|\sigma) \geq D(\rho\|\sigma + \sigma' - \sigma) = D(\rho\|\sigma') \tag{1.66}$$

$\square$

Note that the isometric invariance of the relative entropy can be proved directly, without DPI. Let V be an isometry. Then

$$
\begin{aligned}
D(V\rho V^\dagger \| V\sigma V^\dagger) &= \mathrm{tr} V\rho V^\dagger(\log V\rho V^\dagger - \log V\sigma V^\dagger) &(1.67)\\
&= \mathrm{tr} V\rho V^\dagger V(\log\rho - \log\sigma)V^\dagger &(1.68)\\
&= \mathrm{tr}\rho(\log\rho - \log\sigma) &(1.69)\\
&= D(\rho\|\sigma). &(1.70)
\end{aligned}
$$

Now although this proof works equally well, using DPI can actually be applied to any "divergence". It is in this way more fundamental. Than isometric invariance. Further, joint concavity can be used to prove DPI (this means they are equivalent). So, we want to show $D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_A\|\sigma_A)$.

In general, the following holds $D(\rho\otimes\omega\|\sigma\otimes\tau) = D(\rho\|\sigma) + D(\omega\|\tau)$.

Next goal is to prove the data-processing inequality we continue using. First we need some results from operator theory. They will be stated without proof.

Detour 1: Functions on Operators and Operator Convexity

Let $A \in B(\mathcal{H})$ be Hermitian with spectral decomposition given by

$$A = \sum_i \lambda_i |i\rangle\langle i| \tag{1.71}$$

with $\lambda_i \in \mathbb{R}$, $\langle i|j\rangle = \delta_{ij}$. Then, let $f : I \to \mathbb{R}$, $I \subseteq \mathbb{R}$, be such that $\mathrm{spec} A \subseteq I$, then

$$f(A) = \sum_i f(\lambda_i)|i\rangle\langle i|. \tag{1.72}$$

In words, $f(A)$ is a Hermitian operator with the same eigenbasis as A and spectrum $\{f(\lambda_i)\}_i$.[15]

[15] Note that this implies the often used property that $[A, f(A)] = 0$.

As an example, consider the matrix logarithm. Given a state with spectral decomposition $\rho = \sum_i \lambda_i |i\rangle\langle i| \implies \log\rho = \sum_i \log\lambda_i |i\rangle\langle i|$. Another important example is the entropy function[16] $\eta(t) = t\log t$.

[16] Note that because $\lim_{t\to 0}\eta(t) = 0$, we define $0\log 0 := 0$.

Letting $V : \mathcal{H} \to \mathcal{K}$ be an isometry, $V^\dagger V = \mathbb{I}$, then

$$f(VAV^\dagger) = Vf(A)V^\dagger. \tag{1.73}$$

We will study functions that are operator convex[17]:

[17] Recall the partial order "$\leq$" on Hermitian operators: $A \leq B :\leftrightarrow B - A \geq 0$.

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda)f(B), \tag{1.74}$$

for all $\lambda \in [0,1]$. An immediate result is that every operator convex function is convex as a real function[18]. However, the converse of this statement is *not* true[19]. Some examples of operator convex functions include

[18] Just take $\dim\mathcal{H} = 1$

[19] For example, $t \mapsto t^3$ is not operator convex.

1) $t \mapsto t^p$ for $-1 \leq p \leq 0$ and $1 \leq p \leq 2$,

2) $t \mapsto -t^p$ for $0 \leq p \leq 1$,

3) $t \mapsto -\log t$,

4) $t \mapsto \eta(t) = t \log t, \quad \eta(0) = 0$.

Finally, though we save the proof for later, we state an important result called the *operator Jensen*[20] *inequality*: Let $f : I \to \mathbb{R}$ be operator convex, $V : \mathcal{H} \to \mathcal{K}$ and isometry, A Hermitian with spec$A \subseteq I$. Then, for $A \in B(\mathcal{K})$:

$$f(V^\dagger AV) \leq V^\dagger f(A)V. \tag{1.75}$$

[20] Named after Johan Jensen (1859-1925).

Relative Modular Operator

Fix $A, B \in B(\mathcal{H})$ and define maps $L_A, R_B : B(\mathcal{H}) \to B(\mathcal{H})$:

$$L_A(X) = AX \quad R_B(X) = XB \tag{1.76}$$

Lemma 7. *1)* $[L_A, R_B] = 0$

2) If A is invertible, then L_A, R_A are invertible, and

$$L_A^{-1} = L_{A^{-1}}, \quad R_A^{-1} = R_{A^{-1}} \tag{1.77}$$

3) if A is Hermitian, then so are L_A, R_A with respect to the Hilbert-Schmidt inner product $\langle X, Y\rangle = Tr(X^\dagger Y)$

4) If A is Hermitian and $f : I \to \mathbb{R}$ with spec$A \subseteq I$, then $f(L_A), f(R_A)$ are well defined and

$$f(L_A) - L_{f(A)}, \quad f(R_A) = R_{f(A)}. \tag{1.78}$$

Proof. 1)

$$\begin{aligned}
[L_A, R_B]X &= L_A(R_B X) - R_B(L_A X), & (1.79)\\
&= L_A(XB) - R_B(AX), & (1.80)\\
&= AXB - AXB, & (1.81)\\
&= 0 & (1.82)
\end{aligned}$$

2) By definition

3) We have $\langle X, L_A(Y)\rangle = \langle L_A^\dagger(X), Y\rangle$ and

$$\begin{aligned}
\langle X, L_A(Y)\rangle &= \mathrm{Tr} X^\dagger AY, & (1.83)\\
&= \mathrm{Tr}((A^\dagger X)^\dagger Y), & (1.84)\\
&= \langle A^\dagger X, Y\rangle, & (1.85)
\end{aligned}$$

which implies $L_A^\dagger = L_{A^\dagger}$. Finally by the Hermiticity of A, we have

$$L_A^\dagger = L_A, \tag{1.86}$$

as desired. The analogous argument works for R_A.

4) Let $A = \sum_{i=1}^d a_i \left|i\right\rangle \left\langle i\right|$ where $d = \dim\mathcal{H}$. Then, it follows that $L_A(\left|i\right\rangle\left\langle j\right|) = A\left|i\right\rangle\left\langle j\right| = a_i\left|i\right\rangle\left\langle j\right|$. There are d eigenoperators $\left|i\right\rangle\left\langle j\right|$ with eigenvalue a_i (for $j = 1,\ldots,d$) which implies there are d^2 eigenvalues with orthogonal eigenoperators $\left|i\right\rangle\left\langle j\right|$:

$$\langle \left|i\right\rangle\left\langle j\right|, \left|k\right\rangle\left\langle l\right|\rangle = \delta_{ik}\delta_{jl}. \tag{1.87}$$

Further, $\dim B(\mathcal{H}) = d^2$. Thus, we can define $f(L_A)$ through $f(L_A)(\left|i\right\rangle\left\langle j\right|) = f(a_i)\left|i\right\rangle\left\langle j\right| = L_{f(A)}(\left|i\right\rangle\left\langle j\right|)$. Since $\{\left|i\right\rangle\left\langle j\right|\}_{i,j=1}^d$ is a basis, $f(L_A) = L_{f(A)}$. The same argument holds for R_A.

□

Definition 8. *Let $X, Y \in B(\mathcal{H})$ be Hermitian and Y invertible. The relative modular operator $\Delta = \Delta^{X,Y}$ is defined as*

$$\Delta^{X,Y} = L_X R_{Y^{-1}} : Z \mapsto XZY^{-1} \tag{1.88}$$

Lemma 9. *Let $X, Y \geq 0$, Y invertible, and $\eta(t) = t\log t$. Then $\eta(\Delta^{X,Y}) = \Delta^{X,Y}(L_{\log X} - R_{\log Y})$.*

Proof. Let X, Y have spectral decompositions[21]

$$X = \sum_i x_i \left|e_i\right\rangle\left\langle e_i\right|, \tag{1.89}$$

$$Y = \sum_i y_i \left|f_i\right\rangle\left\langle f_i\right|. \tag{1.90}$$

Then, we have

$$\begin{aligned}
\Delta^{X,Y}(\left|e_i\right\rangle\left\langle f_j\right|) &= L_X R_{Y^{-1}}(\left|e_i\right\rangle\left\langle f_j\right|), &(1.91)\\
&= X\left|e_i\right\rangle\left\langle f_j\right| Y^{-1}, &(1.92)\\
&= x_i\left|e_i\right\rangle\left\langle f_i\right| y_j^{-1}, &(1.93)\\
&= x_i y_j^{-1}\left|e_i\right\rangle\left\langle f_j\right|, &(1.94)
\end{aligned}$$

which implies that $\left|e_i\right\rangle\left\langle f_j\right|$ is an eigenmatrix of $\Delta^{X,Y}$ with eigenvalue $x_i y_j^{-1}$. Then, we have

$$\begin{aligned}
\eta(\Delta)(\left|e_i\right\rangle\left\langle f_j\right|) &= \eta(x_i y_j^{-1})\left|e_i\right\rangle\left\langle f_j\right|, &(1.95)\\
&= x_i y_j^{-1}(\log x_i - \log y_j)\left|e_i\right\rangle\left\langle f_j\right|. &(1.96)
\end{aligned}$$

[21] Note that Y being invertible impies $y_i > 0$.

Thus we have

$$\Delta^{X,Y}(L_{\log X} - R_{\log Y})(|e_i\rangle\langle f_j|) = x_i y_j^{-1}(\log x_i - \log y_j)\,|e_i\rangle\langle f_j|\,. \quad (1.97)$$

Because $\{|e_i\rangle\langle f_j|\}_{i,j=1}^{d}$ forms a basis for $B(\mathcal{H})$, we can conclude that

$$\eta(\Delta^{X,Y}) = \Delta^{X,Y}(L_{\log X} - R_{\log Y}), \quad (1.98)$$

as desired. □

We are now ready to prove Theorem 5, the data processing inequality for the quantum relative entropy.

Proof. First, recall that $\mathrm{supp}\rho \nsubseteq \mathrm{supp}\sigma \implies D(\rho\|\sigma) := \infty$. Next, note

1) Without loss of generality, we assume that $\mathrm{supp}\rho \subseteq \mathrm{supp}\sigma \implies \sigma$ can be taken to be invertible.[22]

2) Remember: for any quantum channel $\mathcal{N} : A \to B$, there exists and isometry $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ such that

$$\mathcal{N}(X_A) = \mathrm{tr}_E(V X_A V^\dagger) \quad (1.99)$$

Recall that $f(VAV^\dagger) = Vf(A)V^\dagger$ by definition. Applying this to the matrix logarithm yields $D(V\rho V^\dagger\|V\sigma V^\dagger) = D(\rho\|\sigma)$. Then, with $\mathcal{N}\mathrm{tr}_E V\cdot V^\dagger$, the claim follows from proving the data-processing inequality for $\mathcal{N} = \mathrm{tr}_B : AB \to A$.

3) To show[23]: $D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_A\|\sigma_A)$. Since[24] $\mathrm{supp}\sigma_{AB} \subseteq \mathrm{supp}\sigma_A \otimes \mathrm{supp}\sigma_B$, we can assume without loss of generality that both σ_{AB} and σ_A are invertible.

Define $\Delta_{AB} = L_{\rho_{AB}} R_{\sigma_{AB}^{-1}}$ and $\Delta_A = L_{\rho_A} R_{\sigma_A^{-1}}$. By definition of the quantum relative entropy, we have

$$D(\rho_{AB}\|\sigma_{AB}) = \mathrm{tr}\rho_{AB}(\log\rho_{AB} - \log\sigma_{AB}). \quad (1.100)$$

Then,[25]

$$\begin{aligned}
\eta(t) = t\log t \quad &= \quad \langle \sigma_{AB}^{1/2}, \eta(\Delta_{AB})(\sigma_{AB}^{1/2})\rangle, && (1.101)\\
&\overset{\text{Lemma 9}}{=} \quad \langle \sigma_{AB}^{1/2}, \Delta_{AB}(L_{\log\rho_{AB}} - R_{\log\sigma_{AB}})(\sigma_{AB}^{1/2})\rangle, && (1.102)\\
&= \quad \langle \sigma_{AB}^{1/2}, \rho_{AB}\log\rho_{AB}\sigma_{AB}^{1/2}\sigma_{AB}^{-1} - \rho_{AB}\sigma_{AB}^{1/2}\log\sigma_{AB}\sigma_{AB}^{-1}\rangle, && (1.103)\\
&= \quad \mathrm{tr}\rho_{AB}\log\rho_{AB} - \mathrm{tr}\rho_{AB}\log\sigma_{AB}, && (1.104)\\
D(\rho_A\|\sigma_A) = \quad & \quad \langle \sigma_A^{1/2}, \eta(\Delta_A)(\sigma_A^{1/2}).\rangle && (1.105)
\end{aligned}$$

Then, with this expression in mind, we will use the operator Jensen's inequality[26]

$$D(\rho_{AB}\|\sigma_{AB}) \geq D(\rho_A\|\sigma_A). \quad (1.106)$$

Let us list our goals before proceeding. We wish to find $V : B(\mathcal{H}_A) \to B(\mathcal{H}_{AB})$ such that

[22] In general, $\mathcal{H} = \ker\sigma \oplus \mathrm{supp}\sigma$; however, we can restrict $\mathcal{H}$ to $\mathrm{supp}\sigma$ by projecting.

[23] ρ_{AB} must be a state but σ_{AB} need only be positive semi-definite.

[24] Proved in Lemma B.4.1 of Renato Renner's PhD thesis.

[25] Remember, $\langle X, Y\rangle = \mathrm{tr}X^\dagger Y$ is the inner product on $B(\mathcal{H})$.

[26] For an isometry V and operator convex function f, $f(V^\dagger XV) \leq V^\dagger f(X)V$.

1) V is an isometry $V^\dagger V = \mathbb{I}_A$,

2) $V^\dagger \Delta_{AB} V = \Delta_A$,

3) $V(\sigma_A^{1/2}) = \sigma_{AB}^{1/2}$.

Now, suppose we have found such an isometry. How would we proceed? We could then write

$$\begin{aligned}
D(\rho_A \| \sigma_A) &= \langle \sigma_A^{1/2}, \eta(\Delta_A)(\sigma_A^{1/2}) \rangle, && \text{by Lemma 9} \quad (1.107)\\
&= \langle \sigma_A^{1/2}, \eta(V^\dagger \Delta_{AB} V)(\sigma_A^{1/2}) \rangle, && \text{by 2)} \quad (1.108)\\
&\leq \langle \sigma_A^{1/2}, V^\dagger \eta(\Delta_{AB}) V(\sigma_A^{1/2}) \rangle, && \text{operator Jensen's} \quad (1.109)\\
&= \langle \sigma_{AB}^{1/2}, \eta(\Delta_{AB})(\sigma_{AB}^{1/2}) \rangle, && \text{by 3)} \quad (1.110)\\
&= D(\rho_{AB} \| \sigma_{AB}), && \text{Lemma 9} \quad (1.111)
\end{aligned}$$

as desired. So, how do we choose the isometry V? Take

$$V : X_A \mapsto (X_A \sigma_A^{-1/2} \otimes \mathbb{I}_B) \sigma_{AB}^{1/2}, \tag{1.112}$$

where clearly $V : B(\mathcal{H}_A) \to B(\mathcal{H}_{AB})$.

1) The first item is easy to check. We have $V^\dagger(Y_{AB}) = \mathrm{tr}_B(Y_{AB} \sigma_{AB}^{1/2} (\sigma_A^{-1/2} \otimes \mathbb{I}_B))$, so

$$\begin{aligned}
V^\dagger V(X_A) &= V^\dagger(X_A \sigma_A^{-1/2} \sigma_{AB}^{1/2}), && (1.113)\\
&= \mathrm{tr}_B(X_A \sigma_A^{-1/2} \sigma_{AB}^{1/2} \sigma_{AB}^{1/2} \sigma_A^{-1/2}), && (1.114)\\
&= X_A \sigma_A^{-1/2} \sigma_A \sigma_A^{-1/2}, && (1.115)\\
&= X_A, && (1.116)
\end{aligned}$$

and because this holds for all X_A, we can conclude that $V^\dagger V = \mathbb{I}_A$ as desired.

2) Next, we wish to show $V^\dagger \Delta_{AB} V = \Delta_A$. We have

$$\begin{aligned}
V^\dagger \Delta_{AB} V(X_A) &= V^\dagger \Delta_{AB}(X_A \sigma_A^{-1/2} \sigma_{AB}^{1/2}), && (1.117)\\
&= V^\dagger(\rho_{AB} X_A \sigma_A^{-1/2} \sigma_{AB}^{1/2} \sigma_{AB}^{-1}), && (1.118)\\
&= \mathrm{tr}_B(\rho_{AB} X_A \sigma_A^{-1/2} \sigma_A^{-1/2} \sigma_A^{1/2} \sigma_A^{-1/2}), && (1.119)\\
&= \mathrm{tr}_B(\rho_{AB} X_A \sigma_A^{-1}), && (1.120)\\
&= \rho_A X_A \sigma_A^{-1}, && (1.121)\\
&= \Delta_A(X_A), && (1.122)
\end{aligned}$$

which holds for all X_A, and thus we conclude $V^\dagger \Delta_{AB} V = \Delta_A$ as desired.

3) $V(\sigma_A^{1/2}) = \sigma_{AB}^{1/2}$

□

This is an extremely important result but a natural question is: does it generalize? We have shown DPI for quantum channels but Denes Petz extended these methods to prove DPIs for trace-preserving 2-positive maps.[27] Moreover, a very recent result by Mueller-Hermes and Reeb shows that DPI holds for all trace-preserving, positive maps.[28]

[27] Φ i 2-positive if $\Phi \otimes \mathbb{I}_2$ is positive($X_{AB} \geq 0 \implies (\Phi \otimes \mathbb{I})(X_A) \geq 0$).

[28] Note that they use different proof methods based on complex interpolation.

2

Entropies and equality in data-processing

2.1 Entropic quantities

Entropies are fundamental quantities in information theory. Perhaps the most famous of which is the von Neumann entropy[1] defined as

$$S(\rho) = -\mathrm{tr}\rho\log\rho = -D(\rho\|\mathbb{I}). \tag{2.1}$$

[1] We often denote the entropy as $S(A)_\rho = S(\rho_A)$ where $\rho_A \in B(\mathcal{H}_A)$ is a quantum state.

The most useful entropic quantities also have operational interpretation. For the von Neumann entropy, there are two operational interpretations: one related to source/data compression and one to entanglement conversion of pure states. More on this later. First, let's meet some of the basic properties of the von Neumann entropy.

Proposition 11. *Let $S(A)_\rho$ be the von Neumann entropy of a quantum state $\rho_A \in B(\mathcal{H})$. Then, $S(A)_\rho$ has the following properties*

1) $0 \leq S(A)_\rho \leq \log|A|$, where the first equality is reached iff ρ_A is pure and the second equality is reached iff $\rho_A = \frac{1}{|A|}\mathbb{I}$

2) Concavity: $S(\sum_i \lambda_i\rho_i) \geq \sum_i \lambda_i S(\rho_i)$

3) Strong sub-additivity: $\forall \rho_{ABC} \in B(\mathcal{H}_\mathcal{A} \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$:

$$S(ABC) + S(C) \leq S(AC) + S(BC) \quad (A \leftrightarrow B \leftrightarrow C) \tag{2.2}$$

$$\Leftrightarrow S(A) + S(B) \leq S(AC) + S(BC), \quad \textit{(weak monotonicity)} \tag{2.3}$$

Note also, that taking $|C| = 1$ gives subadditivity, $(S(AB) \leq S(A) + S(B))$, from SSA.

4) Let $\mathcal{N}$ be a unital quantum channel ($\mathcal{N}(\mathbb{I}) = \mathbb{I}$). Then, $S(\rho) \leq S(\mathcal{N}(\rho))$ for all $\rho \in B(\mathcal{H})$. In particular, let $\{\Pi_i\}_{i=1}^k$ be a projective measurement, $(\Pi_i \geq 0, \Pi_i\Pi_j = \delta_{ij}\Pi_i, \sum_i \Pi_i = \mathbb{I})$ and $\mathcal{N}(X) = \sum_i \Pi_i X \Pi_i \implies S(\rho) \leq S(\mathcal{N}(\rho))$. In particular, $S(\rho) \leq S(diag\rho)$

Proof. 1) Let $\rho = \sum_i \lambda_i |i\rangle \langle i|$ be the spectral decomposition of ρ. Then the entropy can be expressed as $S(\rho) = -\sum_i \lambda_i \log \lambda_i$. Then, because $\lambda_i \in [0,1]$, we know $S(\rho) \geq 0$. The only way to have $S(\rho) = 0$, then, is if there is one eigenvalue equal to one and the rest zero. [2]

[2] That is, if ρ is pure $S(\rho) = 0$.

To see the upper bound on the entropy observe:

$$D(\rho_A \| \frac{1}{|A|}\mathbb{I}_A) = \mathrm{tr}\rho_A(\log \rho_A - \log \frac{1}{|A|}\mathbb{I}_A), \tag{2.4}$$
$$= -S(\rho_A) + \log |A| \tag{2.5}$$
$$\geq 0, \tag{2.6}$$

with equality iff ρ_A is maximally mixed[3].

[3] We will prove this later!

2) $S(\sum_i \lambda_i \rho_i) \geq \sum_i \lambda_i S(\rho_i)$:

$$S(\sum_i \lambda_i \rho_i) = -D(\sum_i \lambda_i \rho_i \| \mathbb{I}), \tag{2.7}$$
$$= -D(\sum_i \lambda_i \rho_i \| \sum_i \lambda_i), \tag{2.8}$$
$$\geq \sum_i \lambda_i (-D(\rho_i \| \mathbb{I})), \tag{2.9}$$
$$= \sum_i \lambda_i S(\rho_i) \tag{2.10}$$

where the last property follows from the joint convexity of relative entropy shown in the proof of Proposition 6.

3) To show: $S(ABC) + S(C) \leq S(AC) + S(BC)$. First note

$$D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) = \mathrm{tr}\rho_{ABC}(\log \rho_{ABC} - \log \rho_A \otimes \rho_{BC}). \tag{2.11}$$

Then we have[4]

[4] Here, we use the fact that when $[X,Y] = 0$, we have $\log XY = \log X + \log Y$. Specifically, we have $[\rho_A \otimes \mathbb{I}_B, \mathbb{I}_A \otimes \rho_{BC}] = 0$

$$\log \rho_A \otimes \rho_{BC} = \log(\rho_A \otimes \mathbb{I}_{BC}) + \log(\mathbb{I}_A \otimes \rho_{BC}), \tag{2.12}$$
$$= \log \rho_A \otimes \mathbb{I}_{BC} + \mathbb{I}_A \otimes \log \rho_{BC}, \tag{2.13}$$

which follows from the properties of logarithms of operators[5] Then, substituting back into the expression above we have

[5] If this step is not immediately obvious, this post shows how one can see $\log A \otimes \mathbb{I} = \log A \otimes \mathbb{I}$ for some diagonalizable operator A.

$$D(\rho_{ABC} \| \rho_A \otimes \rho_{BC}) = \mathrm{tr}\rho_{ABC} \log \rho_{ABC} \tag{2.14}$$
$$- \mathrm{tr}\rho_{ABC}(\log \rho_A \otimes \mathbb{I}_{BC}) \tag{2.15}$$
$$- \mathrm{tr}\rho_{ABC}(\mathbb{I}_A \otimes \rho_{BC}), \tag{2.16}$$
$$= -S(ABC) + S(A) + S(BC), \tag{2.17}$$

then by applying data-processing with respect to $\mathcal{N}(\cdot) = \mathrm{tr}(\cdot)$, we can conclude

$$S(ABC) + S(C) \leq S(AC) + S(BC), \tag{2.18}$$

as desired. Note that taking $|C| = 1$, we see[6] $S(AB) \leq S(A) + S(B)$. Next, we want to prove weak monotonocity:

$$S(A) + S(B) \leq S(AC) + S(BC). \tag{2.19}$$

To see this, let $|\rho\rangle_{ABCD}$ be a purification of ρ_{ABC}. By the Schmidt decomposition, ρ_B and ρ_{ACD} have the same spectrum[7], which implies $S(B) = S(ACD)$ and similarly $S(BC) = S(AD)$. applying these facts, we write

$$S(A) + S(ACD) \leq S(AC) + S(AD), \tag{2.20}$$
$$S(A) + S(B) \leq S(AC) + S(BC), \tag{2.21}$$

as desired.

4) Finally we need to show that if $\mathcal{N}$ is unital,[8] then $S(\rho) \leq S(\mathcal{N}(\rho))$. The proof is straighforward:

$$S(\rho) = -D(\rho\|\mathbb{I}), \tag{2.22}$$
$$\leq -D(\mathcal{N}(\rho)\|\mathbb{I}), \tag{2.23}$$
$$= (\mathcal{N}(\rho)), \tag{2.24}$$

as desired.

□

[6] Quantities that satisfy such an inequality are called sub-additive.

[7] This is a very useful fact that is often used in quantum information theory.

[8] That is, $\mathcal{N}(\mathbb{I}) = \mathbb{I}$.

Another ubiquitous quantity in quantum information theory is the conditional entropy

$$S(A|B)_\rho = S(AB)_\rho - S(B)_\rho, \tag{2.25}$$
$$= -D(\rho_{AB}\|\mathbb{I}_A \otimes \rho_B). \tag{2.26}$$

The conditional entropy has operational interpretations in terms of both the optimal rate of source compression with side information and state merging protocols in quantum Shannon theory.

Proposition 12. *1) Conditioning reduces entropy:*

$$S(A|B) \leq S(A). \tag{2.27}$$

2) Duality relation: let ρ_{AB} have a purification given as $|\rho\rangle_{ABC}$, then

$$S(A|B)_\rho = -S(A|C)_\rho. \tag{2.28}$$

3) Range of conditional entropy:

$$-\log|A| \leq S(A|B) \leq \log|A|, \tag{2.29}$$

where the first inequality is saturated for the maximally entangled state between systems A and B and where the second is saturated when one system is maximally mixed.

4) *Data-processing (strong sub-additivity):*

$$S(A|BC) \leq S(A|B). \tag{2.30}$$

5) *Weak monotonicity (monogomy of entanglement):*

$$S(A|B) + S(A|C) \geq 0. \tag{2.31}$$

6) *Classical conditioning: Let* $\rho_{AX} = \sum_x p_x |x\rangle \langle x|_A \otimes \rho_A^x$, *then*

$$S(A|X) = \sum_x p_x S(A)_{\rho^x}. \tag{2.32}$$

7) *Concavity: for* $\bar{\rho} = \sum_i \lambda_i \rho^i_{AB}$, *we have*

$$S(A|B)_{\bar{\rho}} \geq \sum_i \lambda_i S(A|B)_{\rho^i}. \tag{2.33}$$

Proof. 1) $S(A|B) \leq S(A) \Leftrightarrow S(AB) - S(B) \leq S(A)$

2) $S(A|B) = -S(A|C)$ for a state $|\rho\rangle_{ABC}$ purifying ρ_{AB}. By Schmidt decomposition, $S(AB) = S(C)$ and $S(B) = S(AC)$ so that

$$S(A|B) = S(AB) - S(B) = S(C) - S(AC) - S(A|C), \tag{2.34}$$

3) The lower bound on the range of the conditional entropy follows from the proof of 1) above and from Proposition 11:

$$S(A|B) = -S(A|C), \quad \text{(assume } C \text{ purifies } \rho_{AB}) \tag{2.35}$$
$$\geq -S(A), \tag{2.36}$$
$$\geq -\log |A|. \tag{2.37}$$

The upper bound is shown similarly $S(A|B) \leq S(A) \leq \log |A|$. The equality conditions are simple to verify. First, when ρ_{AB} is separable with the A system is in the maximally mixed state and the B system is in any state, we have

$$S(A|B) = S(AB) - S(B), \tag{2.38}$$
$$= S(A)_\Pi + S(B)_\omega - S(B)_\omega, \tag{2.39}$$
$$= \log |A|. \tag{2.40}$$

Lastly, we have $S(A|B) = -\log |A|$ when $\rho_{AB} = \Phi^+_{AB}$ with $d = |A| \leq |B|$. Because tracing out half of a maximally entangled state yields the maximally mixed state[9], we have

[9] That is, if $|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_i |i\rangle_A |i\rangle_B$, then $\mathrm{tr}_A \Phi^+_{AB} = \Pi_{B'}$, with $|B'| = d$.

$$S(A|B) = S(AB) - S(B), \tag{2.41}$$
$$= 0 - \log d, \tag{2.42}$$
$$= -\log |A|, \tag{2.43}$$

as desired.

4) Strong sub-additivity says: $S(ABC) + S(B) \leq S(AB) + S(CB)$, so

$$S(ABC) - S(BC) \leq S(AB) - S(B), \tag{2.44}$$
$$S(A|BC) \leq S(A|B). \tag{2.45}$$

5) Holds by weak monotonicity of von Neumann entropy[10]

[10] Proven in Proposition 11, part 3).

6) $\rho_{XA} = \sum_x p_x |x\rangle \langle x|_X \otimes \rho_A^x \implies \rho_x = \sum_x p_x |x\rangle \langle x|$. Then,

$$S(A|X) = -D(\rho_{XA} \| \mathbb{I}_A \otimes \rho_x), \tag{2.46}$$
$$= -D(\sum_x p_x |x\rangle \langle x| \otimes \rho_A^x \| \mathbb{I}_A \otimes \sum_x p_x |x\rangle \langle x|), \tag{2.47}$$
$$= -\sum_x p_x D(\rho_A^x \| \mathbb{I}_A), \tag{2.48}$$
$$= \sum_x p_x S(A)_{\rho^x}, \tag{2.49}$$

where the third equality holds by part 6) of Proposition 6.

7) Concavity is shown the same way.[11] We have

[11] Note that if $\bar{\rho}_{AB} = \sum_i \lambda_i \rho^i_{AB}$, then $\bar{\rho}_B = \sum_i \lambda_i \rho^i_B$.

$$S(A|B)_{\bar{\rho}} = -D(\sum_i \lambda_i \rho^i_{AB} \| \mathbb{I}_A \otimes \sum_i \rho^i_B), \tag{2.50}$$
$$\geq \sum_i \lambda_i (-D(\rho^i_{AB} \| \mathbb{I}_A \otimes \rho^i_B)), \tag{2.51}$$
$$= \sum_i \lambda_i S(A|B)_{\rho^i}. \tag{2.52}$$

□

Corollary 13.

$$\rho_{AB} \quad \textit{separable} \implies S(A|B)_\rho \geq 0 \tag{2.53}$$

Proof. If ρ_{AB} is separable, it can be expressed as $\rho_{AB} = \sum_i \lambda_i \omega^i_A \otimes \sigma^i_B$. Then, we have[12]

[12] Note: the first inequality holds by the concavity of conditional entropy that we showed in part 7) of Prop. 12.

$$S(A|B)_\rho \geq \sum_i \lambda_i S(A|B)_{\omega^i \otimes \sigma^i}, \tag{2.54}$$
$$\geq 0. \tag{2.55}$$
$$\tag{2.56}$$

Additionally, we have

$$S(A|B)_{\omega^i\otimes\sigma^i} = S(AB) - S(B), \tag{2.57}$$
$$= S(A) + S(B) - S(B), \tag{2.58}$$
$$= S(A), \tag{2.59}$$
$$\geq 0. \tag{2.60}$$

Combining these, we conclude

$$S(A|B)_\rho \geq 0, \tag{2.61}$$

as desired.[13] □

[13] The converse of this statement does not hold because there are *bound entangled states* for which $S(A|B)_\rho \geq 0$. A bound entangled state is an entangled state that is undistillable. See Quantum Channels I lecture notes.

Another well-studied entropic quantity is the *coherent information* which is defined as

$$I_c(A > B)_\rho = -S(A|B)_\rho, \tag{2.62}$$
$$= S(B) - S(AB), \tag{2.63}$$
$$= D(\rho_{AB}\|\mathbb{I}_A \otimes \rho_B) \tag{2.64}$$

The coherent information has multiple operational interpretations:

1) Entanglement distillation

2) Quantum information transmission

3) Quantum error correction

Of particular note is the so-called Hashing inequality[14]: ρ_{AB} with $I_c(A > B)_\rho > 0$ is distillable. We also note that

[14] See for example this paper which discusses the Hashing inequality for the coherent information.

$$S(A|B) + S(A|C) \geq 0 \tag{2.65}$$
$$\Leftrightarrow I_c(A > B) + I(A > C) \leq 0, \tag{2.66}$$

which can be interpreted as a statement of the no-cloning theorem.

Yet another crucially important entropic quantity is the *mutual information*[15]:

[15] A useful way of understanding the mutual information is the relative en tropy distance from being a product state.

$$I(A;B)_\rho = S(A) + S(B) - S(AB), \tag{2.67}$$
$$= S(A) - S(A|B), \tag{2.68}$$
$$= S(B) - S(B|A), \tag{2.69}$$
$$= D(\rho_{AB}\|\rho_A \otimes \rho_B). \tag{2.70}$$

It has the following operational interpretations:

1) Measure for total correlations (classical and quantum) in a bipartite state

2) entanglement-assisted classical communication

3) classical communication cost in state merging

The mutual information has many important features, several of which are summarized in the following proposition.

Proposition 14. *1)*

$$0 \leq I(A;B)_\rho \leq 2\log\min\{|A|,|B|\}, \tag{2.71}$$
$$0 \leq I(X;B)_\rho \leq \log\min\{|X|,|B|\} \tag{2.72}$$

2) $I(A;BC) \geq I(A;B), I(AB;C) \geq I(A;C)$

3) Holevo information: Let $\mathcal{E} = \{p_x, \rho_A^x\}$ *be a quantum state ensemble. The, the Holevo information is given as:*

$$\chi(\mathcal{E}) = S(\sum_x p_x\rho_A^x) - \sum_x p_x S(\rho_A^x), \tag{2.73}$$
$$= I(X;A)_\rho, \tag{2.74}$$

where $\rho_{XA} = \sum_x p_x |x\rangle\langle x|_x \otimes \rho_A^x$

4) Holevo bound: Let $x \sim p(x)$ *be a classical, discrete random variable,* $\{\rho_B^x\}$ *be a set of quantum states, and let* $E = \{E_B^y\}_y$ *be a POVM (*$E^y \geq 0, \sum_y E^y = \mathbb{I}_B$*). Denote* $p(y|x) = tr(E_\rho^y \rho_B^x)$ *the conditional probability distribution defining our random variable* Y, $p(x,y) = p(y|x)p(x)$*. Then, denote the accessible information as* $I_{acc}(\{p_x, \rho_B^x\}) = \max_{E\ POVM} I(X;Y)$*. Then,*

$$I_{acc}(\{p_x,\rho_B^x\}) \leq \chi(\{p_x,\rho_B^x\}) = I(X;B)_\rho, \tag{2.75}$$

where $\rho_{XB} = \sum_x p_x |x\rangle\langle x|_x \otimes \rho_B^x$

Proof. 1) $0 \leq I(A;B)_\rho \leq 2\log\min\{|A|,|B|\}$. By part 2) of Prop 6, we have

$$I(A;B)_\rho = D(\rho_{AB}\|\rho_A \otimes \rho_B) \geq 0. \tag{2.76}$$

We also have[16]

$$I(A;B) = S(A) = S(A|B) \leq S(A) + \log|A| \leq 2\log|A|, \tag{2.77}$$

which also holds for $S(B) - S(B|A)$. Recall that $S(A|B) = -\log|A|$ for Φ_{AB}^+ maximally entangled ($|A| \leq |B|$). So,

$$I(A;B)_{\Phi^+} = S(A) + S(B) - S(AB), \tag{2.78}$$
$$= \log|A| + \log|A| - 0, \tag{2.79}$$
$$= 2\log|A|. \tag{2.80}$$

When $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_A^x$, we have $I(X;A) \leq \log\min\{|X|,|A|\}$, so[17]

[16] Recall that from Prop 12 that $S(A|B) \geq -\log|A|$.

[17] Recall that by part 6) of Prop. 12, $\sum_x p_x S(A)_{\rho^x} \geq 0$.

$$I(X;A) = S(A) - S(A|X), \tag{2.81}$$
$$\leq S(A), \tag{2.82}$$
$$\leq \log|A|, \tag{2.83}$$

and finally[18]

[18] By Corollary 13 $S(X|A) \geq 0$ because ρ_{XA} is separable.

$$I(X;A) = S(X) - S(X|A), \tag{2.84}$$
$$\leq S(X), \tag{2.85}$$
$$\leq \log|X|. \tag{2.86}$$

2) To show that $I(AB;C) \geq I(A;C)$, we simply apply the data-processing inequality for the channel $\mathrm{tr}_B(\cdot)$:

$$D(\rho_{ABC}\|\rho_{AB}\otimes\rho_C) \geq D(\rho_{AC}\|\rho_A\otimes\rho_C) = I(A;C) \tag{2.87}$$

3) Note that $\rho_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_A^x$ implies $\rho_A = \sum_x p_x\rho_A^x$. Then,

$$I(X;A) = S(A) - S(A|X), \tag{2.88}$$
$$= S(\sum_x p_x\rho_A^x) - \sum_x p_x S(\rho_A^x), \tag{2.89}$$
$$= \chi(\{p_x,\rho_A^x\}) \tag{2.90}$$

4) $I_{\text{acc}}(\{p_x,\rho_B^x\}) = \max_{E\quad \text{POVM}} I(X;Y) \leq I(X;B)_\rho$. Let our POVM, $E = \{E_B^y\}_y$, be a measurement channel.[19] Then

[19] That is, $M(\rho) = \sum_y \mathrm{tr}(E^y\rho)\,|y\rangle\langle y|$.

$$I(X;B) = D(\rho_{XB}\|\rho_X\otimes\rho_B), \tag{2.91}$$
$$= D(\sum_x p_x |x\rangle\langle x| \otimes \rho_B^x \| \rho_X \otimes \sum_x p_x\rho_B^x), \tag{2.92}$$
$$\geq D(\sum_x p_x |x\rangle\langle x| \otimes \sum_y p(y|x)\,|y\rangle\langle y|\,\|\rho_x \otimes \sum_{x,y} p_x p(y|x)\,|y\rangle\langle y|), \tag{2.93}$$
$$= I(X;Y) \quad (p(y,x) = p(y|x)p(x)), \tag{2.94}$$

which finally implies

$$I_{\text{acc}}(\{p_x,\rho_B^x\}) \leq I(X;B), \tag{2.95}$$

as desired.

$\square$

Index