Math 266Y:
Generalizing and strengthening the Data Processing Inequality

Abhishek Anand & Juspreet Singh Sandhu

April 25, 2019

1 Overview

The data processing inequality (DPI) is a fundamental information theoretic notion that the distinguishability between two quantum states can *not* be increased via any physically permitted operation. In other words, there is no information (regarding how different two states are) to be gained by conducting any post-processing.

Here, we first introduce the information theory and quantum mechanics prerequisites to understand DPI. Then, we review a special case of DPI where distinguishability is measured via the Umegaki relative entropy and the operation considered is the partial trace. We also state the elementary strong subadditivity result of Von-Neumann entropy and its relationship to a weak classical version in the concentration inequalities literature. Following that, we discuss the generalized DPI both across a family of quantum quasi-relative entropies and across completely positive trace preserving operations (quantum channels). Lastly, we define the Petz recovery map, its relation to saturating the DPI, and a strengthened DPI based on the performance of the Petz recovery map.

2 Introduction

2.1 Basics of Information Theory

We define the notions of Shannon entropy, KL-Divergence and mutual information to lay down the classical setting for the DPI, and help give motivation for defining the quantum DPI.

Definition 1 (Shannon entropy)**.** Given a distribution $p \sim X$, its Shannon entropy is defined as

$$H(p) = \int_{x \in X} p(x) \log_2 \left(\frac{1}{p(x)} \right) dx$$

The Shannon entropy measures the optimal encoding scheme for a given distribution (into bits). That is, when elements are picked *i.i.d.* from X using the distribution p, the Shannon entropy gives the expected length of the optimally encoded sequence.

Definition 2 (KL-divergence)**.** Given two distributions $p \sim X$ and $q \sim X$, their KL-divergence is defined as:

$$D_{KL}(p \parallel q) = \int_{x \in X} p(x) \log_2 (\frac{p(x)}{q(x)}) dx$$

An intuitive way to interpret KL-divergence is the difference between the optimal encoding of a distribution p (as its Shannon entropy) and an encoding based on a 'guessed' distribution q:

$$D_{KL}(p \parallel q) = \int_{x \in X} p(x) \log_2(\frac{1}{q(x)}) - H(p)$$

One can use the Log-sum inequality to show that $D_{KL}(p \parallel q) \geq 0$. This hints at the notion of calling D_{KL} a distance measure between distributions. However, D_{KL} is not *symmetric* and can be *unbounded* when the support of the two distributions do not intersect.

Definition 3 (Mutual information)**.** Given 2 random variables X and Y, the mutual information between them is defined as

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

This definition measures the information between two random variables as the reduction of uncertainty in one conditioned on knowing the other. Mutual information is symmetric.

Theorem 4 (Classical DPI)**.** *Given Random Variables $X \to Y \to \hat{X}$ forming a Markov Chain,*

$$I(X;Y) \geq I(X;\hat{X})$$

The theorem asserts that given some $\hat{X}$ conditionally independent of X, it can give no more information about X than Y can. More informally, it says that information is never gained through a noisy channel.

2.2 Quantum Mechanics prerequisites

We introduce the notions of pure and mixed quantum states, the Umegaki relative entropy and the partial trace operation, building towards the quantum DPI.

Definition 5 (Quantum state)**.** A matrix ρ of size $n \times n$ is a quantum state if and only if:

- ρ is a positive semi-definite Hermitian operator ($\rho^\dagger = \rho$ and $\rho \geq 0$)
- $\mathrm{Tr}(\rho) = 1$

Quantum state as a probability measure: A quantum state is a matrix-valued generalization of a probability measure. Intuitively, as $\mathrm{Tr}(\rho) = 1$, the terms across the diagonal seem to encode some probabilities, while the non-diagonal terms encode some correlations. To understand what these probabilities and correlations mean, we must define pure and mixed states.

Definition 6 (Pure state)**.** A quantum state ρ is a pure state if it is a rank-1 projection.

Pure and mixed states: A state which is not pure, is mixed. From a physics perspective, a pure state can be expressed as a *ket* while a mixed state can not. Let us use the spectral theorem on a quantum state ρ:

$$\rho = \sum_{i=1}^{n} \lambda_i \rho_i$$

where $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = \mathrm{Tr}(\rho) = 1$. Here, ρ_i are rank-1 projectors and hence, pure states. Therefore, a quantum state is a statistical (classical) mixture of pure states $\{\rho_i\}$ with probability weights $\{\lambda_i\}$.

Definition 7 (Umegaki relative entropy)**.** The Umegaki relative entropy [1] for quantum states ρ and σ is defined as

$$D_{\log}(\rho||\sigma) = \mathrm{Tr}(\rho\log\rho - \rho\log\sigma)$$

Operational significance of relative entropy: Imagine a setting where you are given a quantum state that may be either ρ or σ (this setting would is also useful to understand the significance of DPI). The Umegaki relative entropy encodes the amount of distinguishability between two quantum states. This is because the number of measurements required to distinguish the two states is proportional to $e^{-\alpha D_{\log}(\rho||\sigma)}$ where α is positive constant [2]. Hence, higher the relative entropy, lower the number of measurements required In other words, if $D_{\log}(\rho||\sigma)$ is high, ρ and σ are hard to distinguishable.

Reduction to KL-Divergence: We can see that when the states are classical, that is, they commute with each other ($\rho\sigma = \sigma\rho$), the definition of the Umegaki relative entropy reduces to that of KL-Divergence.

Klein's identity: Notice that $D_{\log}(\rho||\rho) = \mathrm{Tr}(\rho\log\rho - \rho\log\rho) = \mathrm{Tr}\ 0 = 0$. Hence, it is hardest to distinguish a quantum state from itself. Hence, it is desirable for a relative entropy to be always greater than or equal to 0. This can be shown via the Klein's inequality. We state Klein's inequaity without proof.

Theorem 8 (Klein's inequality)**.** *Let $A, B \in \mathbb{H}_n$ and $f : \mathbb{R} \to \mathbb{R}$ be a convex and differentiable function. Then,*

$$\mathrm{Tr}\left[f(A) - f(B) - (A - B)f'(B)\right] \geq 0$$

Corollary 9. Umegaki relative entropy is non-negative.
Proof: Consider $f(x) = x\log x$.

$$f'(x) = 1 + \log x$$

$$f''(x) = \frac{1}{x}$$

Hence, $f''(x) \geq 0$ for $x \geq 0$. Hence, f is convex on $\mathbb{R}_+$. Note that we only work on $\mathbb{R}_+$ as we are dealing with positive operators (quantum states). Applying Klein's inequality on quantum states ρ and σ, we get:

$$\mathrm{Tr}\left[\rho\log\rho - \sigma\log\sigma - (\rho - \sigma)(1 + \log\sigma)\right] \geq 0$$

$$\implies D_{\log}(\rho||\sigma) - \mathrm{Tr}\rho + \mathrm{Tr}\sigma = D_{\log}(\rho||\sigma) \geq 0$$

where the last step follows as $\mathrm{Tr}\rho = \mathrm{Tr} = 1$ (as they are quantum states).

Note that to receive the desirable result of the relative entropy being non-negative for all ρ and σ, it was crucial for f to be convex. We will return to this constraint in Section 4.1.

Definition 10 (Partial trace)**.** For density matrix ρ on $\mathbb{C}^m \otimes \mathbb{C}^n$ space, the partial trace with respect to the second system, $\mathrm{Tr}_2\rho$ is defined as the unique quantum state that satisfies, for arbitrary operation M acting on the first system:

$$\mathrm{Tr}(M \cdot \mathrm{Tr}_2\rho) = \mathrm{Tr}(M \otimes I_n \cdot \rho).$$

Physical intuition: Let us assume that we have a composite quantum state made of two systems. When we trace out the second system, we get the quantum state describing the first system if we were looking at it in isolation. In other words, we are discarding the degrees of freedom of the second system. If we started with a pure composite state, we would end up with a mixed state on the first system. This is because when we are looking at the first system in isolation, it would appear as simply a classical statistical mixture (depending on some external variable) over some pure states. Furthermore, note that we *lose* information about the state by applying the partial trace (as some degrees of freedom are discarded).

3 DPI & Strong-subadditivity of entropy

We begin by stating the statement proved in lecture for DPI, briefly mentioning the proof along with the significance of the statement. We then state the statement about the Strong-subadditivity of entropy, give a proof-sketch and state the classical (weak) version of the inequality in the concentration inequalities literature and its significance.

3.1 Umegaki DPI

Theorem 11 (Umegaki DPI)**.** *Given quantum states* $\rho, \sigma \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$*:*

$$D_{log}(\rho \| \sigma) \geq D_{log}(tr_2\rho \| tr_2\sigma)$$

Proof (sketch): We briefly recall the argument from class, where we used two facts:

- For all locally compact groups G, there exists a unique left-invariant Haar-measure μ, such that, for a subset D of G: $\mu(D) = \mu(\rho D), \forall \rho \in G$
- The average unitary action on one sub-system (the other system being constant) "averages" the sub-system to the maximally mixed state:

$$\frac{1}{dim(H_2)}\left(\rho_1 \otimes I_{\mathcal{H}_2}\right) = \int_{U_{\mathcal{H}_2}} \left(I_{H_1} \otimes U^*\right) \rho \left(I_{\mathcal{H}_1} \otimes U\right) \ d\mu$$

  where $U_{\mathcal{H}_2}$ is the set of all unitaries on $\mathcal{H}_2$

We rewrite the relative entropy as:

$$D_{log}\left(\frac{1}{dim(\mathcal{H}_2)}(\rho_1 \otimes I_{H_2}) \| \frac{1}{dim(\mathcal{H}_2)}(\sigma_1 \otimes I_{H_2})\right) = D_{log}\left((\rho_1) \| (\sigma_1)\right)$$

Applying Jensen's inequality to the above and then using the above-mentioned facts allows us to re-write the representation and then average over $\mathcal{H}_\infty$ to conclude the theorem.

The statement above simply says that averaging over a part of the composite space (uniformly at random) does not allow us to gain information. More informally, it says that the partial information recovered by the partial trace can never tell us more than the complete picture.

Significance: Recall the operational picture of quantum relative entropy: the number of measurements required to distinguish the two states is proportional to $e^{-\alpha D_{\log}(\rho||\sigma)}$ where α is some positive constant [2]. The standard DPI hence, states that if you wish to determine whether a composite system is ρ or σ, looking at just parts of the system (partial traces), will not reduce the number of measurements you require: you will gain no more information.

3.2 Strong-subadditivity of entropy: quantum & classical

We begin by stating the strong-subadditvity of the Von-Neumann entropy result, and then give its significance. We then state an analogous classical result in the concentration inequalities literature to present some structural similarity in the classical information theory & quantum information theory notions.

Theorem 12 (Strong subadditivity of Von-Neumann entropy)**.** *For all quantum states $\rho, \rho_1, \rho_2 \in H$, the Von-Neumann entropy:*

$$S(\rho) \leq S(\rho_1) + S(\rho_2)$$

with equality iff $\rho = \rho_1 \otimes \rho_2$

Proof (sketch): There are three main steps:

- We use Klein's inequality on $D(\rho \parallel \rho_1 \otimes \rho_2)$ to conclude that the relative entropy is positive.
- Then, we decompose the the term $\log(\rho_1 \otimes \rho_2)$ using tensor algebra as:

$$\log(\rho_1 \otimes \rho_2) = \log(\rho_1) \otimes I_{H_2} - I_{H_1} \otimes \log(\rho_2)$$

- We use the above to rewrite the relative entropy as:

$$D(\rho \| \rho_1 \otimes \rho_2) = -S(\rho) - \mathrm{Tr}(\rho(\log \rho_1 \otimes I_{H_2})) - \mathrm{Tr}(\rho(I_{H_1} \otimes \log \rho_2)) = -S(\rho) + S(\rho_1) + S(\rho_2)$$

The RHS of the equation above is positive (by Klein's inequality) and so, we can rearrange the equation above to conclude the theorem.

Significance: The lemma above tells us that the "joint" entropy or the collective entropy of a system is always highest when it can be viewed as a composition (or product) of independent sub-systems.

Theorem 13 (Weak classical subadditivity of entropy [3])**.** *Given $\Phi(x) = x\log(x), \forall x > 0$, along with $X_1, .., X_n \sim$ i.i.d. random variables taking values in a countable set X, a function $f : X^n \to [0, \infty)$ and a random variable $Z = f(X_1, .., X_n)$ we can assert that:*

$$E\,\Phi(Z) - \Phi(E\,Z) \leq \sum_{i=1}^{n} E[E^{(i)}\,\Phi(Z) - \Phi(E^{(i)}\,Z)]$$

where $E^{(i)}$ denotes the marginal expectation conditioned on a fixed $i \in [n]$

Parsing the theorem, one realizes that by scaling the Z, such that, $\mathbb{E}[Z] = 1$, we get the LHS to be $\mathbb{E}[Z\log(Z)]$, which can be interpreted as the relative entropy between $q(x) = f(x)p(x)$ and $p(x)$ ($\forall x \in \mathbb{X}^n$), where $p(x)$ is the product distribution across each X_i. We can then view the statement as saying that the relative entropy is upper bounded by the sum of the conditional entropies. For $n = 2$, this gives an analogous statement to the strong subadditivity of Von-Neumann entropy.

4 Petz generalization

4.1 Motivation for extension

4.1.1 Across operator monotone decreasing functions

Notice that while proving Umegaki relative entropy is non-negative, we relied on $f(x) = x\log x$ to be convex. It was convex as $f''(x) = \frac{1}{x} \geq 0$ for $x \geq 0$. In some sense, this is closely linked to log being a concave function. Furthermore, while proving the standard statement of DPI in Section 3.1, we applied the Jensen's inequality for concave functions. Hence, it was critical for *log* to be operator monotone or equivalently, for $-log$ to be operator monotone decreasing.

Hence, if we replace $-log$ with another operator monotone decreasing function, we should be able to retain the desirable properties of relative entropy: non-negativeness and DPI. The function f that we use ($f(x) = -\log x$ for the Umegaki relative entropy), would simply determine our distance metric between ρ and σ.

4.1.2 Across quantum channels

For the standard DPI, in Section 3.1, we were performing a partial trace on our states, and gaining no information. The partial trace belongs to a larger family of maps: completely positive trace preserving maps. Given that through evolution, the trace of a quantum state must remain 1, these are the "allowed" evolution maps from a physics perspective. They generalize the classical information-theoretic notion of noisy channels. We wish to extend DPI to this entire family of maps. Why should the retain DPI?

Let us consider a special member of this family, that unlike the partial trace, that does not reduce degrees of freedom or dimensions, the unitary evolution operator U, for which $U^\top U = I$. Notice that $U^\top$ is a completely positive trace preserving operator as well. For two quantum states, ρ and σ,

$$D_{\log}(\rho||\sigma) = \mathrm{Tr}(\rho \log \rho - \rho \log \sigma)$$

Consider the spectral decomposition of ρ and σ:

$$\rho = \sum_{i=1}^{m} \lambda_i \rho_i$$

$$\sigma = \sum_{i=1}^{n} \mu_i \sigma_i$$

where ρ_i and σ_i are rank-1 projectors. Then, the relative entropy is given by:

$$D_{\log}(\rho||\sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma) = \sum_{i=1}^{m} \lambda_i \log \lambda_i - \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \log \mu_i \mathrm{Tr}(\rho_i \sigma_j)$$

as $\mathrm{Tr}\rho_i = 1$, given that they are rank-1 projectors. Let us compute the Umegaki relative entropy after the operator U is applied to both quantum states via the map ϕ:

$$\phi(\rho) = U \rho U^\top$$

$$D_{\log}(U\rho U^\top || U\sigma U^\top) = \mathrm{Tr}(U\rho U^\top \log U\rho U^\top - U\rho U^\top \log U\sigma U^\top)$$

Given that unitary evolution will not change the eigenvalues of ρ or σ, only the projectors will change:

$$= \sum_{i=1}^{m} \lambda_i \log \lambda_i \mathrm{Tr}(U\rho_i U^\top) - \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \log \mu_i \mathrm{Tr}(U\rho_i U^\top U \sigma_j U^\top) = \sum_{i=1}^{m} \lambda_i \log \lambda_i - \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i \log \mu_i \mathrm{Tr}(\rho_i \sigma_j)$$

using the cyclicity of trace and $U^\top U = I$.

Hence, we can see that a unitary map does not change the relative entropy between two states. Intuitively, this is expected. Information should not gain or lost by performing reversible transformations as we can simply "undo" them and our definition of relative entropy takes that into account. Furthermore, any other completely positive trace preserving transformation will lead to loss of degrees of freedom and hence, loss of information and hence, can only decrease the relative entropy. To formalize this, note that any quantum channel can be represented as a unitary evolution in a higher-dimension Hilbert space and then applying a partial trace on certain dimensions [4]. We have shown that the unitary evolution will not increase entropy and the standard DPI states that partial trace will not increase entropy as well. Using these two cases, we can expect that for all quantum channels, DPI should hold.

4.2 Generalization

To introduce and briefly sketch the proof for the DPI statement on operator monotone functions, we introduce the preliminary notions of a super-operator:

Definition 14 (Relative-modular operator)**.** A relative-modular operator $\Delta_{\rho,\sigma}$ is a map from a Hilbert space to itself ($\mathcal{H} \to \mathcal{H}$) that is a conjugation action:

$$\Delta_{\rho,\sigma}(X) = \rho X \sigma^{-1}$$

We notice that a relative-modular Operator with the restriction that it is trace preserving and element-wise monotone becomes a quantum channel. With the notions of a relative-Mmdular operator and convexity, we can define f-relative quasi entropy as:

Definition 15 (Quasi-relative entropy)**.** Given a finite dimensional Hilbert Space $\mathcal{H}$, quantum states ρ, σ, a linear map $\mathcal{K} : \mathcal{H} \to \mathcal{H}$, and an operator convex function $f : (0, \infty) \to \mathbb{R}$, we define the Quasi-Relative Entropy to be:

$$D_f^K(\rho\|\sigma) = Tr(\rho^{\frac{1}{2}} K^* f(\Delta_{\rho,\sigma}) K \rho^{\frac{1}{2}})$$

This definition is a generalization of the Umegaki relative entropy to any convex function. Setting $f(x) = -\log(x)$, we have:

$$f(\Delta_{\rho,\sigma})\rho^{\frac{1}{2}} = -(\log(\Delta_{\rho,\sigma})\rho^{\frac{1}{2}}) = \rho^{\frac{1}{2}} \log(\sigma) - \log(\rho)\rho^{\frac{1}{2}}$$

Substituting the above equality into the definition of quasi-relative entropy (with K being the identity map) yields:

$$D^{id}_{-log(x)}(\rho\|\sigma) = Tr(\rho^{\frac{1}{2}}(-\log(\Delta_{\rho,\sigma})\rho^{\frac{1}{2}}))$$

We then use our evaluation of $-(\log(\Delta_{\rho,\sigma})\rho^{\frac{1}{2}})$ along with the cyclicity of the trace to conclude that:

$$D^{id}_{-log(x)}(\rho\|\sigma) = D_{log}(\rho\|\sigma) = \mathrm{Tr}(\rho \log \rho - \rho \log \sigma) = \text{Umegaki entropy}$$

Definition 16 (Completely positive, trace preserving maps)**.** A map $\phi : A \to B$ is completely positive and trace preserving if $\mathrm{Tr}(\rho) = \mathrm{Tr}(A(\rho))$ and if its action is element-wise monotone on any $B \in \mathbb{C}^{k \times k}$, where $B_{ij} = A_{ij}$ for $k \leq n$

Essentially, a trace-preserving completely-positive map is the equivalent of a general (noisy) quantum channel. We now state the general DPI, framed in the context of quasi-relative entropy.

Theorem 17 (CPTP f-relative quasi entropy DPI [5])**.** $\forall f : (0, \infty) \to \mathbb{R}$ *operator convex functions and* $\forall \mathcal{N}$ *CPTP maps with* $K = I$*:*

$$D_f^{id}(\rho\|\sigma) \geq D_f^{id}((\mathcal{N}\rho)\|(\mathcal{N}\sigma))$$

Proof (sketch): We begin by defining two operators:

- **Tracial projection operator** - This operator $\mathcal{E}_\tau$ simply maps a given matrix X to its unit-norm projection in a restriction of the Hilbert space and then traces it out. More formally: $\forall A \in \mathcal{N}$ and $\forall X \in \mathcal{H}$:

  $$\mathcal{E}_\tau(X) = \tau(A^* X)$$

  The τ denotes a faithful tracial state. This is a state whose density matrix ($|\tau\rangle \langle\tau|$) is the maximally mixed state ($\frac{1}{n}(I)$).

- **General adjoint operator** - We define two such operations on any $X, Y \in \mathcal{H}$:

$$U(X) = \mathcal{E}_\tau(X)(\mathcal{E}_\tau(\rho))^{\frac{-1}{2}}\rho^{\frac{1}{2}}$$

$$U(Y) = \mathcal{E}_\tau(Y\rho^{\frac{1}{2}})(\mathcal{E}_\tau(\rho))^{\frac{-1}{2}}$$

We use the definition above to compute $U^*U(X) = \mathcal{E}_\tau(X)$, and compute the relative-modular operator acting on the adjoint-operator:

$$(\Delta_{\sigma,\rho}(U(X)))^{\frac{1}{2}} = \sigma^{\frac{1}{2}}X(\mathcal{E}_\tau(\rho))^{\frac{-1}{2}}$$

We can then (after some matrix algebra) assert that:

$$\langle \Delta_{\sigma,\rho}(U(X)))^{\frac{1}{2}}|\Delta_{\sigma,\rho}(U(X)))^{\frac{1}{2}}\rangle = \langle (\Delta_{(\mathcal{E}_\tau(\sigma))^{\frac{-1}{2}},(\mathcal{E}_\tau(\rho))^{\frac{-1}{2}}})^{\frac{1}{2}}|(\Delta_{(\mathcal{E}_\tau(\sigma))^{\frac{-1}{2}},(\mathcal{E}_\tau(\rho))^{\frac{-1}{2}}})^{\frac{1}{2}}\rangle$$

This allows us to conclude that:

$$U^*\Delta_{\sigma,\rho}U = \Delta_{(\mathcal{E}_\tau(\sigma))^{\frac{-1}{2}},(\mathcal{E}_\tau(\rho))^{\frac{-1}{2}}}$$

By Jensen's inequality:

$$U^*f(\Delta_{\sigma,\rho})U \geq f(U^*\Delta_{\sigma,\rho}U)$$

Combining the equality and inequality above allows us to finally conclude that:

$$D_f^{id}(\rho\|\sigma) \geq D_f^{id}((\mathcal{N}\rho)\|(\mathcal{N}\sigma))$$

Note that the proof follows similarly to the case when $f = -\log(x)$ and we have the Umegaki entropy by introducing a representation $(U(X), U(Y))$, reducing it to a form that involves the action of the relative-modular operator, and then applying Jensen's inequality to it. The only difference is the form induced, whereby, in the case of the Umegaki entropy, it suffices to use Schur's lemma, while over here we rely on the tracial projection operator and how it "pulls-through" into the relative-modular operator.

Significance vs. standard DPI: Firstly, it does not matter what function f you use to define relative entropy as long as it is operator monotone decreasing: DPI will still hold. Furthermore, standard DPI stated that in the operational perspective, looking at components of the system will not give more information. This generalization states that there is no physically "allowed" operation that you can do to extract more information. That is, let's say that n measurements are needed to determine up to error $\varepsilon > 0$, that the state you have is ρ or σ. There is *no* operation that can be done before measuring that reduces the number of measurements required.

5 The Petz recovery map

5.1 Motivation

Consider the unitary map ϕ applied to a quantum state ρ, where U is a unitary matrix.

$$\phi(\rho) = U\rho U^\top$$

There exists another unitary map ϕ', which reverses or "undoes" the action of ϕ:

$$\phi'(\rho) = U^\top\rho U$$

as $\phi'\phi(\rho) = \phi'(U\rho U^\top) = U^\top U\rho U^\top U = \rho$ as $U^\top U = UU^\top = I$.

Notice that both ϕ and ϕ' are completely positive trace preserving maps. For two quantum states ρ and σ, let us apply DPI with ϕ as the CPTP map being applied. We get:

$$D(\rho||\sigma) \geq D(\phi(\rho)||\phi(\sigma))$$

For, for the two quantum states $\phi(\rho)$ and $\phi(\sigma)$, let us apply DPI with ϕ' as the CPTP map being applied. We get:

$$D(\phi(\rho)||\phi(\sigma)) \geq D(\phi'(\phi(\rho))||\phi'(\phi(\sigma))) = D(\rho||\sigma)$$

Combining the two above statements, we get:

$$D(\rho||\sigma) \geq D(\phi(\rho)||\phi(\sigma)) \geq D(\rho||\sigma)$$

which is try if and only if:

$$D(\rho||\sigma) = D(\phi(\rho)||\phi(\sigma))$$

Hence, we can see that for unitary map, DPI is saturated and both the relative entropies coincide. However, notice that it was not crucial for the map to be unitary, all that was required was there existed a map that could "undo" the operation. Furthermore, it is important the map that undoes the initial operation is same for both ρ and σ. Enter, the Petz recovery map.

5.2 DPI equality and Petz recovery map

Theorem 18 (Petz recovery map [5])**.** *For quantum states ρ and σ, and a map ϕ,*

$$D(\rho||\sigma) = D(\phi(\rho)||\phi(\sigma))$$

if and only if there exists a quantum map ϕ' (the Petz recovery map), such that:

$$\phi'(\phi(\rho)) = \rho$$

$$\phi'(\phi(\sigma)) = \sigma$$

Furthermore, without loss of generality, ϕ' can be written as, for quantum state X:

$$\phi'(X) = \rho^{1/2}(\phi(\rho)^{-1/2}X\phi(\rho)^{-1/2})\rho^{1/2}$$

Proof (if direction): Section 5.1
Proof (only if direction): We shall prove this as a corollary of Theorem 19.

5.3 Strengthening DPI via the Petz recovery map performance

Now, we know that if the Petz recovery map exists and undoes the operation for both ρ and σ, DPI is saturated. However, notice that we can always construct a recovery map from the perspective of one of the quantum states. For instance, for state ρ, for all density matrices, X: consider

$$\phi'(X) = \mathcal{R}_\rho(X) = \rho^{1/2}(\phi(\rho)^{-1/2}X\phi(\rho)^{-1/2})\rho^{1/2}$$

where ϕ is the CPCT map applied for DPI.

We can see that: $\mathcal{R}_\rho(\phi(\rho)) = \rho$. Hence, using Theorem 18, DPI will be satured if and only if:

$$\mathcal{R}_\rho(\phi(\sigma)) = \sigma$$

That is, if we can use the recovery map associated with one quantum state to undo the ϕ action on the other quantum state. Therefore, we can quantify the performance of the recovery map via the difference: $(\mathcal{R}_\rho(\phi(\sigma)) - \sigma)$. DPI already provides a bound that the new relative entropy (after application of ϕ) cannot be greater than the initial relative entropy. Intuitively, we may be believe the performance of the recovery map can provide a bound on how different the entropy can be: if performance is good, that is $(\mathcal{R}_\rho(\phi(\sigma)) - \sigma) \sim 0$, then perhaps, the entropies are close and vice-versa. The achievement of Carlen and Verhsynina [6] is exactly that.

Here, we shall be proving a limited version of their claim: we shall be dealing only with Umegaki entropies. In Section 5.4, we shall state their claim fully.

Theorem 19 (Carlen-Verhsynina [6])**.** *For quantum states ρ and σ, and CP/CT map ϕ:*

$$D_{(t)}(\rho||\sigma) - D_{(t)}(\phi(\rho)||\phi(\sigma)) \geq \left(\frac{\pi}{8}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2} \|\mathcal{R}_\rho(\phi(\sigma)) - \sigma\|_1^4$$

.

where $\Delta_{\sigma,\rho}$ is the relative modular operator defined in Section 4 and $\|\cdot\|_1$ is the trace norm.

Before proving this, first note that the "only if" condition of Theorem 18 is a corollary of Theorem 19:

$$D(\rho||\sigma) - D(\phi(\rho)||\phi(\sigma)) = 0 \implies \|\mathcal{R}_\rho(\phi(\sigma)) - \sigma\|_1^4 = 0$$
$$\implies \mathcal{R}_\rho(\phi(\sigma)) = \sigma$$

First let us define a new type of entropies: for which the operator monotone decreasing function f is given by: $f(x) = (t+x)^{-1}$ for some. Let's label this entropy $D_{(t)}$. To prove this we will rely on two auxiliary lemmas. We will first state the lemmas, the proof to which involves some matrix algebra and can be found in [6]. We will then state the final proof, which involves a clever application of Loewner's Theorem, an upper bound on the integral representation of the square-root function via a spectral norm bound on the Relative-Modular Operator, and a choice of limit that minimizes a constant term. A direct application of the first lemma into the proof will yield the main theorem. We begin by stating the two matrix algebra lemmas:

Lemma 20 (Bounded 1-norm)**.** $\forall$ operators X, Y, such that, $Tr(X^*X) = Tr(Y^*Y) = 1$:

$$||X^*X - Y^*Y||_1 \leq 2||X - Y||_2$$

The lemma above simply establishes an upper bound on the 1-norm of the difference of the Inner-Product of 2 matrices by the difference of their 2-norms, provided the Frobenius Norm of both operators is 1.

Lemma 21 (Skew-Symmetric composition of positive operators under Isometry)**.** Given an isometry embedding a Hilbert Space into another $U : \mathcal{K} \to \mathcal{H}$, and positive invertible operators $B \in \mathcal{K}, A \in \mathcal{H}$ related by isometric conjugation $U^*AU = B$, the stretch to any $v \in \mathcal{K}$ is:

$$\langle v|U^*A^{-1}Uv\rangle = \langle v|B^{-1}v\rangle + \langle w|Aw\rangle$$

where $w = UB^{-1}v - A^{-1}Uv$.

Some intuition for this lemma may be given by the fact that it decomposes the inverse map (A^{-1}) under conjugation by the isometry (U) into the invertible map of B along with the linear cross-conjugation terms of UB^{-1} and $A^{-1}U$.

Proof: Relative entropy difference as an integral: First let us use Lemma 21, with the following definitions of A, B and v:

$$A = (t + \Delta_{\sigma,\rho})$$
$$B = (t + \Delta_{\phi(\sigma),\phi(\rho)})$$
$$v = (\phi(\rho))^{1/2}$$

As per definition of relative entropy:

$$D_{(t)}(\rho||\sigma) - D_{(t)}(\phi(\rho)||\phi(\sigma)) = \langle \rho^{1/2}, (t + \Delta_{\sigma,\rho})^{-1}\rho^{1/2}\rangle - \langle \phi(\rho)^{1/2}, (t + \Delta_{\phi(\sigma),\phi(\rho)})^{-1}\phi(\rho)^{1/2}\rangle$$

Now using $U(\phi(\rho))^{1/2} = \rho^{1/2}$ and Lemma 21, we get:

$$\langle \rho^{1/2}, (t + \Delta_{\sigma,\rho})^{-1}\rho^{1/2}\rangle - \langle \phi(\rho)^{1/2}, (t + \Delta_{\phi(\sigma),\phi(\rho)})^{-1}\phi(\rho)^{1/2}\rangle = \langle w_t, (t + \Delta_{\sigma,\rho})w_t\rangle$$

with

$$w_t = U(t + \Delta_{\phi(\sigma),\phi(\rho)})^{-1}(\phi(\rho))^{1/2} - (t + \Delta_{\sigma,\rho})^{-1}\rho^{1/2}$$

Notice the interesting inverse characteristics of w_t and recall the integral representation of the square root function that Lowner's theorem gave us:

$$X^{1/2} = \frac{1}{\pi}\int_0^\infty t^{1/2}\left(\frac{1}{t} - \frac{1}{t+X}\right)\mathrm{d}t,$$

After, some pattern matching, we can observe that in fact, the following expression can be interpreted as an integral of w_t using Lowner's theorem:

$$U(\Delta_{\phi(\sigma),\phi(\rho)})^{1/2}(\phi(\rho))^{1/2} - (\Delta_{\sigma,\rho})^{1/2}\rho^{1/2} = \frac{1}{\pi}\int_0^\infty t^{1/2}w_t\mathrm{d}t\ .$$

From the proof of Theorem 17, the same expression can be reduced to:

$$U(\Delta_{\phi(\sigma),\phi(\rho)})^{1/2}(\phi(\rho))^{1/2} - (\Delta_{\sigma,\rho})^{1/2}\rho^{1/2} = (\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}$$

Comparing the two expressions:

$$(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2} \ = \frac{1}{\pi}\int_0^\infty t^{1/2}w_t\mathrm{d}t$$

Now we can take the norm on both sides:

$$\|(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 = \frac{1}{\pi}\left\|\int_0^\infty t^{1/2}w_t\mathrm{d}t\right\|_2$$

Bounding the integral:

We can break the integral up at some arbitary unit T and using the triangle inequality: we get that:

$$\frac{1}{\pi}\int_0^T t^{1/2}\|w_t\|_2 \mathrm{d}t + \frac{1}{\pi}\left\|\int_T^\infty t^{1/2} w_t \mathrm{d}t\right\|_2 \geq \frac{1}{\pi}\int_0^\infty t^{1/2} w_t \mathrm{d}t$$

Now applying the the Cauchy-Schwarz inequality to the first part:

$$(\int_0^T t^{1/2}\|w_t\|_2 \mathrm{d}t)^2 \leq T\int_0^T t\|w_t\|_2^2 \mathrm{d}t$$

Recall through Lemma 21, we found that

$$D_{(t)}(\rho||\sigma) - D_{(t)}(\phi(\rho)||\phi(\sigma)) = \langle w_t, (t+\Delta_{\sigma,\rho})w_t\rangle$$
$$\implies \geq t\|w_t\|^2$$

where the last inequality follows as $\Delta_{\sigma,\rho}$ is positive.

Using this (replacing $t\|w_t\|^2$ with the difference of relative entropy) we get that:

$$(\int_0^T t^{1/2}\|w_t\|_2 \mathrm{d}t)^2 \leq T\int_0^T t\|w_t\|_2^2 \mathrm{d}t \leq T\int_0^\infty \left(D_{(t)}(\rho||\sigma) - D_{(t)}(\phi(\rho)||\phi(\sigma))\right) \mathrm{d}t$$

Now notice that the integral of $\frac{1}{x+t}$ is $log(x+t)$. Also, note that there is a $log(\alpha)$ term with $\alpha \to \infty$ that appears, however, given that we are integrating a difference, the two large log terms get cancelled and were are left with the Umegaki relative entropy. Therefore, when we integrate:

$$T\int_0^\infty \left(D_{(t)}(\rho||\sigma) - D_{(t)}(\phi(\rho)||\phi(\sigma))\right) \mathrm{d}t = T(S(\rho||\sigma) - S(\phi(\rho)||\phi(\sigma)))$$

Given that the the norm of X is the square root of the highest singular value, we can state the following inequalities:

$$t^{1/2}\left(\frac{1}{tI} - \frac{1}{tI+X}\right) \leq t^{1/2}\left(\frac{1}{tI} - \frac{1}{tI+\|X\|I}\right) = \frac{\|X\|}{t^{1/2}(\|X\|I+tI)},$$

and now integrating,

$$\int_T^\infty t^{1/2}\left(\frac{1}{tI} - \frac{1}{tI+X}\right)\mathrm{d}t \leq \|X\|^{1/2}\left(\int_{T/\|X\|}^\infty \frac{1}{t^{1/2}(1+t)}\mathrm{d}t\right) \leq \frac{2\|X\|}{T^{1/2}}.$$

The spectra of $\phi(\sigma)$ and $\phi(\rho)$ lie in the convex hulls of the spectra of σ and ρ. This can be seen easily as a consequence of the fact that ϕ is trace-preserving and the sum of the trace gives the sum of the eigenvalues. Therefore, we can now bound the spectral norm of the relative-modular operator of σ, ρ from above by that of the relative-modular operator of $\phi(\sigma), \phi(\rho)$:

$$\|\Delta_{\phi(\sigma),\phi(\rho)}\| \leq \|\Delta_{\sigma,\rho}\|$$

Recalling the definition of w_t, we obtain

$$\left\|\int_T^\infty t^{1/2} w_t \mathrm{d}t\right\|_2 \leq \frac{4\|\Delta_{\sigma,\rho}\|}{T^{1/2}}\ .$$

Adding up the two components of the integeral, we get:

$$\|(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \leq \frac{1}{\pi}T^{1/2}(S(\rho||\sigma) - S(\phi(\rho)||\phi(\sigma)))^{1/2} \ + \ \frac{4\|\Delta_{\sigma,\rho}\|}{\pi T^{1/2}}\ .$$

Realize that the term bounded by the Spectral norm vanishes when we minimize T. To minimize T, we take the derivative of $\frac{4\|\Delta_{\sigma,\rho}\|}{T^{1/2}}$ and set it to 0, noticing that this happens in the limit that $T \to \infty$. This also tightens the splitting term in the triangle inequality and we get:

$$\|(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \leq \frac{4}{\pi}\|\Delta_{\sigma,\rho}\|^{1/2}(S(\rho||\sigma) - S(\phi(\rho)||\phi(\sigma)))^{1/4}$$

We get (*):

$$D_{log}(\rho||\sigma) - D_{log}(\phi(\rho)||\phi(\sigma)) \geq \left(\frac{\pi}{4}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\|(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2^4$$

Relating $(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}$ to the performance of the recovery map:

Now, if we simply apply Lemma 20, with $X = (\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2}$ and $Y = \sigma^{1/2}$, we get

$$2\|(\phi(\sigma))^{1/2}(\phi(\rho))^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \geq \|\rho^{1/2}(\phi(\rho)^{-1/2}\phi(\sigma)\phi(\rho)^{-1/2})\rho^{1/2} - \sigma\|_1 = \|\mathcal{R}_\rho(\phi(\sigma)) - \sigma\|_1 \ .$$

Applying this to (*), we get:

$$D_{log}(\rho||\sigma) - D_{log}(\phi(\rho)||\phi(\sigma)) \geq \left(\frac{\pi}{8}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\|\mathcal{R}_\rho(\phi(\sigma)) - \sigma\|_1^4$$

5.4 Further reading

For further generalizations of the DPI inequality to all f-relative quasi entropies, one may refer to the following sources: **Corollary 4.5** in [7] which states that in the setting of Theorem 19 with some additional constraints, with K and c as constants that can be determined, we can show that for a general f-relative quasi-entropy:

$$\|\mathcal{R}_\rho(\phi(\sigma)) - \sigma\|_1 \leq K(D_f(\rho||\sigma) - D_f(\phi(\rho)\|\phi(\sigma)))^{\frac{1}{4}\frac{1}{1+c}}$$

Furthermore, [8] provides further generalizations. See **Corollary 4.7** and **Theorem 4.4**.

References

[1] Hisaharu Umegaki. "Conditional expectation in an operator algebra, IV (entropy and information)". In: *Kodai Mathematical Seminar Reports*. Vol. 14. 2. Department of Mathematics, Tokyo Institute of Technology. 1962, pp. 59–85.

[2] Matthew J Donald. "On the relative entropy". In: *Communications in Mathematical Physics* 105.1 (1986), pp. 13–34.

[3] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[4] Vlatko Vedral. "The role of relative entropy in quantum information theory". In: *Reviews of Modern Physics* 74.1 (2002), p. 197.

[5] Dénes Petz. "Sufficient subalgebras and the relative entropy of states of a von Neumann algebra". In: *Communications in mathematical physics* 105.1 (1986), pp. 123–131.

[6] Eric A Carlen and Anna Vershynina. "Recovery map stability for the Data Processing Inequality". In: *arXiv preprint arXiv:1710.02409* (2017).

[7] Eric A Carlen and Anna Vershynina. "Recovery and the Data Processing Inequality for quasi-entropies". In: *IEEE Transactions on Information Theory* 64.10 (2018), pp. 6929–6938.

[8] Anna Vershynina. "On quantum quasi-relative entropy". In: *arXiv preprint arXiv:1810.00391* (2018).