PHYS 7895: Quantum Information Theory Fall 2015

Lecture 20 — November 2, 2015

Prof. Mark M. Wilde *Scribe: Mark M. Wilde*

1 Overview

In the previous lecture, we discussed a quantum entropy inequality called the monotonicity of quantum relative entropy. This has several important consequences, such as the strong subadditivity of quantum entropy, concavity of conditional quantum entropy, joint convexity of quantum relative entropy, etc. We also discussed how the conditional quantum entropy is continuous and gave an explicit bound for continuity with respect to the topology induced by the trace norm.

In this lecture, we prove the monotonicity of quantum relative entropy using a very recent proof technique. This method gives a physically meaningful improvement of the entropy inequality which has an interpretation in terms of "recoverability."

2 Introduction

The quantum entropy inequalities discussed previously lie at the core of quantum Shannon theory and in fact underly some important principles of physics such as the uncertainty principle. In fact, we will use these entropy inequalities to prove the converse parts of every coding theorem in this course. Their prominence in both quantum Shannon theory and other areas of physics motivates us to study them in more detail. The book delves into more depth regarding many of the classical entropy inequalities, and in the process, establishes necessary and sufficient conditions for the saturation of the inequalities, while also understanding the near saturation of the entropy inequalities. The aim of this lecture is to carry out a similar program for all of the quantum entropy inequalities presented in the previous chapter. The outcome will be a proof for the monotonicity of quantum relative entropy, with the added benefit of an understanding of the saturation and near saturation of this quantum entropy inequality.

The main result in this chapter can be summarized informally as follows: if the decrease in quantum relative entropy between two quantum states after a quantum channel is relatively small, then it is possible to perform a recovery channel, such that we can perfectly recover one state while approximately recovering the other. This can be interpreted as quantifying how well one can reverse the action of a quantum channel. Throughout, we take ρ, σ, and $\mathcal{N}$ as given in the following definition:

Definition 1. *Let $\rho \in \mathcal{D}(\mathcal{H})$ and let $\sigma \in \mathcal{L}(\mathcal{H})$ be positive semi-definite, such that* $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. *Let $\mathcal{N} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ be a quantum channel.*

The formal statement of the theorem is as follows:

Theorem 2. *Given ρ, σ, and $\mathcal{N}$ as in Definition 1, there exists a recovery channel $\mathcal{R}_{\sigma,\mathcal{N}} : \mathcal{L}(\mathcal{H}') \to \mathcal{L}(\mathcal{H})$, depending only on σ and $\mathcal{N}$, such that*

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -\log F(\rho, (\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)). \tag{1}$$

Given that the quantum fidelity F takes values between zero and one, we can immediately conclude that

$$-\log F(\rho, (\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)) \geq 0, \tag{2}$$

so that the above theorem implies the monotonicity of quantum relative entropy as a consequence. Furthermore, the recovery channel satisfying (1) has the property that it perfectly recovers σ from $\mathcal{N}(\sigma)$:

$$(\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\sigma) = \sigma, \tag{3}$$

which is something that we show later.

The proof given here for Theorem 2 relies on the method of complex interpolation and the notion of a Rényi generalization of a relative entropy difference. We review this background first before going through the proof. One of the consequences of Theorem 2 is to provide physically meaningful improvements to many quantum entropy inequalities discussed in the previous lecture, such as strong subadditivity, joint convexity of quantum relative entropy, and concavity of conditional quantum entropy. We explore one of these consequences in Section 7 and point to the book for the rest.

3 Schatten Norms and Complex Interpolation

The proof of Theorem 2 given here requires a bit of mathematical background before we can delve into it. So we first begin by defining the Schatten norms and several of their properties. We then review some essential results from complex analysis, that lead to a complex interpolation theorem known as the Stein–Hirschman interpolation theorem.

3.1 Schatten Norms and Duality

An important technical tool in the proof given here is the Schatten p-norm of an operator A, defined as

$$\|A\|_p \equiv \left[\text{Tr}\left\{|A|^p\right\}\right]^{1/p}, \tag{4}$$

where $A \in \mathcal{L}(\mathcal{H})$, $|A| \equiv \sqrt{A^\dagger A}$, and $p \geq 1$. We have already studied two special cases of this norm, which are the trace norm when $p = 1$ and the Hilbert–Schmidt norm when $p = 2$. One can show, along the same lines as we did for the trace norm, that $\|A\|_p$ is equal to the p-norm of the singular values of A. That is, if $\sigma_i(A)$ is the vector of singular values of A, then

$$\|A\|_p = \left[\sum_i \sigma_i(A)^p\right]^{1/p}. \tag{5}$$

The convention is for $\|A\|_\infty$ to be defined as the largest singular value of A because $\|A\|_p$ converges to this in the limit as $p \to \infty$. In the proof of Theorem 2, we repeatedly use the fact that $\|A\|_p$

is unitarily invariant. That is, $\|A\|_p$ is invariant with respect to linear isometries, in the sense that $\|A\|_p = \left\|UAV^\dagger\right\|_p$, where $U, V \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ are linear isometries satisfying $U^\dagger U = I_{\mathcal{H}}$ and $V^\dagger V = I_{\mathcal{H}}$. Isometric invariance follows from (5) and because these isometries do not change the singular values of A. From these norms, one can define information measures relating quantum states and channels, with the main one used here known as a Rényi generalization of a relative entropy difference.

Extending the Cauchy–Schwarz inequality is an important inequality known as the Hölder inequality:

$$|\langle A, B\rangle| = \left|\mathrm{Tr}\{A^\dagger B\}\right| \leq \|A\|_p \|B\|_q , \tag{6}$$

holding for $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and $A, B \in \mathcal{L}(\mathcal{H})$. When $p, q \in [1, \infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$, p and q are said to be Hölder conjugates of each other. One can see that Cauchy–Schwarz is a special case by picking $p = q = 2$. Observe that equality is achieved in (6) if A and B are such that $A^\dagger = a\,|B|^{q/p}\,U^\dagger$ for some constant $a \geq 0$ and where U is a unitary such that $B = U\,|B|$ is a left polar decomposition of B. The Hölder inequality along with the sufficient equality condition is enough for us to conclude the following variational expression for the p-norm in terms of its Hölder dual q-norm:

$$\|A\|_p = \max_{\|B\|_q \leq 1} \mathrm{Tr}\{A^\dagger B\}. \tag{7}$$

This expression can be very useful in calculations.

Exercise 3. *Prove that* $\|AB\|_1 \leq \|A\|_p \|B\|_q$ *for* $p, q \in [1, \infty]$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and* $A, B \in \mathcal{L}(\mathcal{H})$.

Throughout we adopt the usual convention and define $f(A)$ for a function f and a positive semi-definite operator A as follows: $f(A) \equiv \sum_{i:\lambda_i \neq 0} f(\lambda_i)|i\rangle\langle i|$, where $A = \sum_i \lambda_i |i\rangle\langle i|$ is a spectral decomposition of A. We denote the support of A by $\mathrm{supp}(A)$, and we let Π_A denote the projection onto the support of A.

3.2 Complex Analysis

We now review a few concepts from complex analysis. We will not prove these results in detail, but the purpose instead is to recall them, and the interested reader can follow references to books on complex analysis for details of proofs. The culmination of the development is the Stein–Hirschman complex interpolation theorem (Theorem 7).

The derivative of a complex-valued function $f : \mathbb{C} \to \mathbb{C}$ at a point $z_0 \in \mathbb{C}$ is defined in the usual way as

$$\left.\frac{df(z)}{dz}\right|_{z=z_0} = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}. \tag{8}$$

In order for this limit to exist, it must be the same for all possible directions that one could take in the complex plane to approach z_0, and this requirement demarcates a substantial difference between differentiability of real functions and complex ones. Complex differentiability shares several properties with real differentiability: it is linear and obeys the product rule, the quotient rule, and the chain rule. If f is complex differentiable at every point z_0 in an open set U, then we say that f is *holomorphic* on U.

There is a connection between real differentiability and complex differentiability, given by the Cauchy–Riemann equations. Let $f(x+iy) = u(x,y) + iv(x,y)$ where $x, y \in \mathbb{R}$ and $u, v : \mathbb{R} \to \mathbb{R}$. If f is holomorphic, then u and v have first partial derivatives with respect to x and y and satisfy the Cauchy–Riemann equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{9}$$

The converse is not always true. However, if the first partial derivatives of u and v are continuous and satisfy the Cauchy–Riemann equations, then f is holomorphic. An important holomorphic function for our purposes is given in the following exercise:

Exercise 4. *Verify that $f(z) = a^z = e^{[\ln a]z}$, where $a > 0$ and $z \in \mathbb{C}$, is a holomorphic function everywhere in the complex plane. (Hint: Use that $e^z = e^x\left[\cos(y) + i\sin(y)\right]$ for $z = x + iy$ and $x, y \in \mathbb{R}$.)*

Holomorphic functions have good closure properties. That is, the sums, products, and compositions of holomorphic functions are holomorphic as well, given that complex differentiation is linear and satisfies the product, quotient, and chain rules. Note that the quotient of two holomorphic functions is holomorphic wherever the denominator is not equal to zero.

The *maximum modulus principle* is an important principle that holomorphic functions obey. Formally, it is the following statement: let $f : \mathbb{C} \to \mathbb{C}$ be a function holomorphic on some connected, bounded open subset U of $\mathbb{C}$. If $z_0 \in U$ is such that $|f(z_0)| \geq |f(z)|$ for all z in a neighborhood of z_0, then the function f is constant on U. A consequence of this is that if f is not constant on a bounded, connected, open subset U of $\mathbb{C}$, then it achieves its maximum on the boundary of U.

The maximum modulus principle has an extension to an unbounded strip in $\mathbb{C}$, which we call the *maximum modulus principle on a strip*. Let S denote the standard strip in $\mathbb{C}$, $\overline{S}$ its closure, and $\partial\overline{S}$ its boundary:

$$S \equiv \{z \in \mathbb{C} : 0 < \mathrm{Re}\{z\} < 1\}, \tag{10}$$

$$\overline{S} \equiv \{z \in \mathbb{C} : 0 \leq \mathrm{Re}\{z\} \leq 1\}, \tag{11}$$

$$\partial\overline{S} \equiv \{z \in \mathbb{C} : \mathrm{Re}\{z\} = 0 \vee \mathrm{Re}\{z\} = 1\}. \tag{12}$$

Let $f : \overline{S} \to \mathbb{C}$ be bounded on $\overline{S}$, holomorphic on S, and continuous on $\partial\overline{S}$. Then the supremum of $|f|$ is attained on $\partial\overline{S}$. That is, $\sup_{z\in\overline{S}} |f(z)| = \sup_{z\in\partial\overline{S}} |f(z)|$.

The maximum modulus principle on a strip implies a result known as the Hadamard three-lines theorem:

Theorem 5 (Hadamard Three-Lines)**.** *Let $f : \overline{S} \to \mathbb{C}$ be a function that is bounded on $\overline{S}$, holomorphic on S, and continuous on the boundary $\partial\overline{S}$. Let $\theta \in (0,1)$ and $M(\theta) \equiv \sup_{t\in\mathbb{R}} |f(\theta + it)|$. Then $\ln M(\theta)$ is a convex function on $[0,1]$, implying that*

$$\ln M(\theta) \leq (1-\theta)\ln M(0) + \theta \ln M(1). \tag{13}$$

There is a strengthening of the Hadamard three-lines theorem due to Hirschman, which in fact implies the Hadamard three-lines theorem:

Theorem 6 (Hirschman)**.** *Let $f(z) : \overline{S} \to \mathbb{C}$ be a function that is bounded on $\overline{S}$, holomorphic on S, and continuous on the boundary $\partial\overline{S}$. Then for $\theta \in (0,1)$, the following bound holds*

$$\ln|f(\theta)| \leq \int_{-\infty}^{\infty} dt \ \left(\alpha_\theta(t) \ln\left[|f(it)|^{1-\theta}\right] + \beta_\theta(t) \ln\left[|f(1+it)|^{\theta}\right]\right), \tag{14}$$

where

$$\alpha_\theta(t) \equiv \frac{\sin(\pi\theta)}{2(1-\theta)\left[\cosh(\pi t) - \cos(\pi\theta)\right]}, \tag{15}$$

$$\beta_\theta(t) \equiv \frac{\sin(\pi\theta)}{2\theta\left[\cosh(\pi t) + \cos(\pi\theta)\right]}. \tag{16}$$

For a fixed $\theta \in (0,1)$, we have that $\alpha_\theta(t), \beta_\theta(t) \geq 0$ for all $t \in \mathbb{R}$ and

$$\int_{-\infty}^{\infty} dt \ \alpha_\theta(t) = \int_{-\infty}^{\infty} dt \ \beta_\theta(t) = 1 \ , \tag{17}$$

so that $\alpha_\theta(t)$ and $\beta_\theta(t)$ can be interpreted as probability density functions. Furthermore, we have that

$$\lim_{\theta \searrow 0} \beta_\theta(t) = \frac{\pi}{2}\left[\cosh(\pi t) + 1\right]^{-1} \equiv \beta_0(t) \ , \tag{18}$$

where β_0 is also a probability density function on $\mathbb{R}$. With these observations, we can see that Hirschman's theorem implies the Hadamard three-lines theorem, given that an expectation can never exceed a supremum.

3.3 Complex Interpolation of Schatten Norms

We can extend much of the development above to operator-valued functions, which is needed to prove Theorem 2. Let $G : \mathbb{C} \to \mathcal{L}(\mathcal{H})$ be an operator-valued function. We say that $G(z)$ is holomorphic if every function mapping z to a matrix entry is holomorphic. For our purposes in what follows, we are interested in operator-valued functions of the form A^z, where A is a positive semi-definite operator. In this case, we apply the usual convention and take $A^z = \sum_{i:\lambda_i \neq 0} \lambda_i^z |i\rangle\langle i|$, where $A = \sum_i \lambda_i |i\rangle\langle i|$ is an eigendecomposition of A with $\lambda_i \geq 0$ for all i. Given the result of Exercise 4 combined with the closure properties of holomorphic functions mentioned above, we can conclude that A^z is holomorphic if A is positive semi-definite.

We can now establish a version of the Hirschman theorem which applies to operator-valued functions and allows for bounding their Schatten norms. This is one of the main technical tools that we need to establish Theorem 2.

Theorem 7 (Stein–Hirschman)**.** *Let $G : \overline{S} \to L(\mathcal{H})$ be an operator-valued function that is bounded on $\overline{S}$, holomorphic on S, and continuous on the boundary $\partial\overline{S}$. Let $\theta \in (0,1)$ and define p_θ by*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1} \ , \tag{19}$$

where $p_0, p_1 \in [1,\infty]$. Then the following bound holds

$$\ln\|G(\theta)\|_{p_\theta} \leq \int_{-\infty}^{\infty} dt \ \left(\alpha_\theta(t) \ln\left[\|G(it)\|_{p_0}^{1-\theta}\right] + \beta_\theta(t) \ln\left[\|G(1+it)\|_{p_1}^{\theta}\right]\right) \ , \tag{20}$$

where $\alpha_\theta(t)$ and $\beta_\theta(t)$ are defined in (15)–(16)*.*

Proof. For fixed $\theta \in (0,1)$, let q_θ be the Hölder conjugate of p_θ, defined by

$$\frac{1}{p_\theta} + \frac{1}{q_\theta} = 1 \ . \tag{21}$$

Similarly, let q_0 and q_1 be Hölder conjugates of p_0 and p_1, respectively. From the sufficient equality condition for the Hölder inequality, we can find an operator X such that $\|X\|_{q_\theta} = 1$ and $\mathrm{Tr}\{XG(\theta)\} = \|G(\theta)\|_{p_\theta}$. We can write the singular value decomposition for X in the form $X = UD^{1/q_\theta}V$ (implying $\mathrm{Tr}\{D\} = 1$). For $z \in S$, define

$$X(z) \equiv UD^{\frac{1-z}{q_0} + \frac{z}{q_1}}V \ . \tag{22}$$

As a consequence, $X(z)$ is bounded on $\overline{S}$, holomorphic on S, and continuous on the boundary $\partial\overline{S}$. Also, observe that $X(\theta) = X$. Then the following function satisfies the requirements needed to apply Theorem 6:

$$g(z) \equiv \mathrm{Tr}\{X(z)G(z)\} \ . \tag{23}$$

Indeed, we have that

$$\ln \|G(\theta)\|_{p_\theta} = \ln |g(\theta)| \tag{24}$$

$$\leq \int_{-\infty}^{\infty} dt \ \left(\alpha_\theta(t) \ln \left[|g(it)|^{1-\theta} \right] + \beta_\theta(t) \ln \left[|g(1+it)|^{\theta} \right] \right) . \tag{25}$$

Now, from applying Hölder's inequality and the facts that $\|X(it)\|_{q_0} = 1 = \|X(1+it)\|_{q_1}$, we find that

$$|g(it)| = |\mathrm{Tr}\{X(it)G(it)\}| \leq \|X(it)\|_{q_0} \|G(it)\|_{p_0} = \|G(it)\|_{p_0} \, , \tag{26}$$

and

$$|g(1+it)| = |\mathrm{Tr}\{X(1+it)G(1+it)\}| \tag{27}$$

$$\leq \|X(1+it)\|_{q_1} \|G(1+it)\|_{p_1} \tag{28}$$

$$= \|G(1+it)\|_{p_1} \, . \tag{29}$$

Bounding (25) from above using these inequalities then gives (20). □

The theorem above is known as a complex interpolation theorem because it allows us to obtain estimates on the "intermediate" norm in terms of other norms which might be available. Furthermore, we are interpolating through the holomorphic family of operators given by $G(z)$.

4 Petz Recovery Map

The channel appearing in the lower bound of Theorem 2 has an explicit form and is constructed from a map known as the Petz recovery map, which we define as follows:

Definition 8. *Let $\sigma \in \mathcal{L}(\mathcal{H})$ be positive semi-definite, and let $\mathcal{N} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ be a quantum channel. The Petz recovery map $\mathcal{P}_{\sigma,\mathcal{N}} : \mathcal{L}(\mathcal{H}') \to \mathcal{L}(\mathcal{H})$ is a completely positive, trace-non-increasing linear map defined as follows for $Q \in \mathcal{L}(\mathcal{H}')$:*

$$\mathcal{P}_{\sigma,\mathcal{N}}(Q) \equiv \sigma^{1/2} \mathcal{N}^\dagger \left([\mathcal{N}(\sigma)]^{-1/2} \, Q \, [\mathcal{N}(\sigma)]^{-1/2} \right) \sigma^{1/2}. \tag{30}$$

The Petz recovery map $\mathcal{P}_{\sigma,\mathcal{N}}$ is linear, and it is completely positive because it is equal to a serial concatenation of three completely positive maps: $Q \to [\mathcal{N}(\sigma)]^{-1/2} Q [\mathcal{N}(\sigma)]^{-1/2}$, $Q \to \mathcal{N}^\dagger(Q)$, and $M \to \sigma^{1/2} M \sigma^{1/2}$ for $M \in \mathcal{L}(\mathcal{H})$. It is trace-non-increasing because the following holds for positive semi-definite Q:

$$\begin{aligned}
\operatorname{Tr}\{\mathcal{P}_{\sigma,\mathcal{N}}(Q)\} &= \operatorname{Tr}\left\{\sigma \mathcal{N}^\dagger\left([\mathcal{N}(\sigma)]^{-1/2} Q [\mathcal{N}(\sigma)]^{-1/2}\right)\right\} && (31)\\
&= \operatorname{Tr}\left\{\mathcal{N}(\sigma) [\mathcal{N}(\sigma)]^{-1/2} Q [\mathcal{N}(\sigma)]^{-1/2}\right\} && (32)\\
&= \operatorname{Tr}\{\Pi_{\mathcal{N}(\sigma)} Q\} && (33)\\
&\leq \operatorname{Tr}\{Q\}. && (34)
\end{aligned}$$

An important special case of the Petz recovery map occurs when σ and $\mathcal{N}$ are effectively classical. That is, suppose that $\mathcal{N}$ is a classical-to-classical channel with Kraus operators $\{\sqrt{N(y|x)}|y\rangle\langle x|\}$, where $N(y|x)$ is a conditional probability distribution. Suppose further that $\sigma = \sum_x q(x)|x\rangle\langle x|$, with $q(x) \geq 0$ for all x. In this case, one can check that the Petz recovery map is a classical-to-classical channel with Kraus operators $\{\sqrt{R(x|y)}|x\rangle\langle y|\}$, where $R(x|y)$ is a conditional probability distribution given by the Bayes theorem, satisfying

$$R(x|y)(Nq)(y) = N(y|x)q(x), \tag{35}$$

for all x and y, where $(Nq)(y) = \sum_x N(y|x)q(x)$. We leave the details of this calculation as an exercise for the reader and point out that this recovery channel appears in the refinement of the monotonicity of classical relative entropy in the book.

We can also define a partial isometric map $\mathcal{U}_{\sigma,t}$ in the following way:

$$\mathcal{U}_{\sigma,t}(M) \equiv \sigma^{it} M \sigma^{-it}, \tag{36}$$

and where $\sigma^{it}\sigma^{-it} = \Pi_\sigma$. In the case that σ is positive definite, $\mathcal{U}_{\sigma,t}$ is a unitary channel. We can then define a rotated or "swiveled" Petz map, which plays an important role in the construction of a recovery channel satisfying the lower bound in Theorem 2.

Definition 9 (Rotated Petz Map)**.** *Let $\sigma \in \mathcal{L}(\mathcal{H})$ be positive semi-definite, and let $\mathcal{N} : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ be a quantum channel. A rotated Petz map is defined as follows for $Q \in \mathcal{L}(\mathcal{H}')$:*

$$\mathcal{R}^t_{\sigma,\mathcal{N}}(Q) \equiv (\mathcal{U}_{\sigma,-t} \circ \mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{U}_{\mathcal{N}(\sigma),t})(Q). \tag{37}$$

Exercise 10 (Perfect Recovery)**.** *Verify that a rotated Petz map perfectly recovers σ from $\mathcal{N}(\sigma)$:*

$$\mathcal{R}^t_{\sigma,\mathcal{N}}(\mathcal{N}(\sigma)) = \sigma. \tag{38}$$

(To show this equality when σ is not invertible, first show that $\mathcal{N}^\dagger(\Pi_{\mathcal{N}(\sigma)}) \geq \Pi_\sigma$.)

5 Rényi Information Measure

Given ρ, σ, and $\mathcal{N}$ as in Definition 1, we define a Rényi information measure known as a Rényi generalization of a relative entropy difference:

$$\widetilde{\Delta}_\alpha(\rho,\sigma,\mathcal{N}) \equiv \frac{2\alpha}{\alpha-1} \ln \left\| \left([\mathcal{N}(\rho)]^{\frac{1-\alpha}{2\alpha}} [\mathcal{N}(\sigma)]^{\frac{\alpha-1}{2\alpha}} \otimes I_E \right) U \sigma^{\frac{1-\alpha}{2\alpha}} \rho^{1/2} \right\|_{2\alpha}, \tag{39}$$

where $\alpha \in (0,1) \cup (1,\infty)$ and $U : \mathcal{H} \to \mathcal{H}' \otimes \mathcal{H}_E$ is an isometric extension of the channel $\mathcal{N}$. That is, U is a linear isometry satisfying $\mathrm{Tr}_E\{U(\cdot)U^\dagger\} = \mathcal{N}(\cdot)$ and $U^\dagger U = I_{\mathcal{H}}$. Recall that all isometric extensions of a channel are related by an isometry acting on the environment system E, so that the definition in (39) is invariant under any such choice. Recall that the adjoint $\mathcal{N}^\dagger$ of a channel is given in terms of an isometric extension U as $\mathcal{N}^\dagger(\cdot) = U^\dagger\left((\cdot) \otimes I_E\right) U$.

The following lemma is one of the main reasons that we say that $\widetilde{\Delta}_\alpha(\rho,\sigma,\mathcal{N})$ is a Rényi generalization of a relative entropy difference. A proof is available in the book.

Lemma 11. *The following limit holds for ρ, σ, and $\mathcal{N}$ as given in Definition 1:*

$$\frac{1}{\ln 2}\lim_{\alpha\to 1}\widetilde{\Delta}_\alpha(\rho,\sigma,\mathcal{N}) = D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)). \tag{40}$$

For $\alpha = 1/2$, observe that

$$\begin{aligned}\widetilde{\Delta}_{1/2}(\rho,\sigma,\mathcal{N}) &= -\ln\left\|\left([\mathcal{N}(\rho)]^{1/2}\,[\mathcal{N}(\sigma)]^{-1/2}\otimes I_E\right)U_{S\to BE}\sigma^{1/2}\rho^{1/2}\right\|_1^2 &(41)\\ &= -\ln F\left(\rho,\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\rho))\right). &(42)\end{aligned}$$

where $F(\rho,\sigma) \equiv \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2$ is the quantum fidelity. Thus, if $\widetilde{\Delta}_\alpha(\rho,\sigma,\mathcal{N})$ were monotone non-decreasing with respect to α, we could combine these observations to conclude that

$$\begin{aligned}D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) &= \frac{1}{\ln 2}\widetilde{\Delta}_1(\rho,\sigma,\mathcal{N}) &(43)\\ &\overset{?}{\geq} \frac{1}{\ln 2}\widetilde{\Delta}_{1/2}(\rho,\sigma,\mathcal{N}) &(44)\\ &= -\log F\left(\rho,\mathcal{P}_{\sigma,\mathcal{N}}(\mathcal{N}(\rho))\right). &(45)\end{aligned}$$

If this were true, then we could conclude that Theorem 2 would be true with the recovery channel taken to be the Petz recovery map. However, it is not known whether this is true, and we will instead invoke the Stein–Hirschman theorem to conclude that a convex combination of rotated Petz maps satisfies the bound stated in Theorem 2.

6 Proof of the Recoverability Theorem

This section presents the proof of Theorem 2. In fact, we prove the following stronger statement, which implies Theorem 2 for the following recovery channel

$$\mathcal{R}_{\sigma,\mathcal{N}} = \int_{-\infty}^{\infty} dt\ \beta_0(t)\ \mathcal{R}_{\sigma,\mathcal{N}}^{t/2}, \tag{46}$$

due to the concavity of both the logarithm and the fidelity.

Theorem 12. *Let ρ, σ, and $\mathcal{N}$ be as given in Definition 1. Then the following inequality holds*

$$D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq -\int_{-\infty}^{\infty} dt\ \beta_0(t)\ \log\left[F\left(\rho,(\mathcal{R}_{\sigma,\mathcal{N}}^{t/2}\circ\mathcal{N})(\rho)\right)\right], \tag{47}$$

where $\mathcal{R}_{\sigma,\mathcal{N}}^{t/2}$ is a rotated Petz recovery map from Definition 9.

Proof. We can prove this result by employing Theorem 7. We first establish the inequality in (47). Let $U : \mathcal{H} \to \mathcal{H}' \otimes \mathcal{H}_E$ be an isometric extension of the channel $\mathcal{N}$. Pick

$$G(z) \equiv \left([\mathcal{N}(\rho)]^{z/2} [\mathcal{N}(\sigma)]^{-z/2} \otimes I_E \right) U \sigma^{z/2} \rho^{1/2}, \tag{48}$$

for $z \in \overline{S}$, $p_0 = 2$, $p_1 = 1$, and $\theta \in (0,1)$, which fixes $p_\theta = \frac{2}{1+\theta}$. The operator valued-function $G(z)$ satisfies the conditions needed to apply Theorem 7. For the choices above, we find that

$$\|G(\theta)\|_{2/(1+\theta)} = \left\| \left([\mathcal{N}(\rho)]^{\theta/2} [\mathcal{N}(\sigma)]^{-\theta/2} \otimes I_E \right) U \sigma^{\theta/2} \rho^{1/2} \right\|_{2/(1+\theta)}, \tag{49}$$

$$\begin{aligned} \|G(it)\|_2 &= \left\| \left([\mathcal{N}(\rho)]^{it/2} [\mathcal{N}(\sigma)]^{-it/2} \otimes I_E \right) U \sigma^{it} \rho^{1/2} \right\|_2 \\ &\leq \left\| \rho^{1/2} \right\|_2 \\ &= 1, \end{aligned} \tag{50}$$

$$\begin{aligned} \|G(1+it)\|_1 &= \left\| \left([\mathcal{N}(\rho)]^{1+it/2} [\mathcal{N}(\sigma)]^{-(1+it)/2} \otimes I_E \right) U \sigma^{(1+it)/2} \rho^{1/2} \right\|_1 \\ &= \left\| \left([\mathcal{N}(\rho)]^{\frac{it}{2}} [\mathcal{N}(\rho)]^{\frac{1}{2}} [\mathcal{N}(\sigma)]^{-\frac{it}{2}} [\mathcal{N}(\sigma)]^{-\frac{1}{2}} \otimes I_E \right) U \sigma^{\frac{1}{2}} \sigma^{\frac{it}{2}} \rho^{\frac{1}{2}} \right\|_1 \\ &= \left\| \left([\mathcal{N}(\rho)]^{1/2} [\mathcal{N}(\sigma)]^{-it/2} [\mathcal{N}(\sigma)]^{-1/2} \otimes I_E \right) U \sigma^{1/2} \sigma^{it/2} \rho^{1/2} \right\|_1 \\ &= \sqrt{F} \left(\rho, \left(\mathcal{U}_{\sigma,-t/2} \circ \mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{U}_{\mathcal{N}(\sigma),t/2} \right) (\mathcal{N}(\rho)) \right) \\ &= \sqrt{F}(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)). \end{aligned} \tag{51}$$

Then we can apply Theorem 7 to conclude that

$$\begin{aligned} &\ln \left\| \left([\mathcal{N}(\rho)]^{\theta/2} [\mathcal{N}(\sigma)]^{-\theta/2} \otimes I_E \right) U \sigma^{\theta/2} \rho^{1/2} \right\|_{2/(1+\theta)} \\ &\qquad \leq \int_{-\infty}^{\infty} dt \ \beta_\theta(t) \ln \left[F \left(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho) \right)^{\theta/2} \right]. \end{aligned} \tag{52}$$

This implies that

$$\begin{aligned} &-\frac{2}{\theta} \ln \left\| \left([\mathcal{N}(\rho)]^{\theta/2} [\mathcal{N}(\sigma)]^{-\theta/2} \otimes I_E \right) U \sigma^{\theta/2} \rho^{1/2} \right\|_{2/(1+\theta)} \\ &\qquad \geq -\int_{-\infty}^{\infty} dt \ \beta_\theta(t) \ \ln \left[F \left(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho) \right) \right]. \end{aligned} \tag{53}$$

Letting $\theta = (1-\alpha)/\alpha$, we see that this is the same as

$$\widetilde{\Delta}_\alpha(\rho, \sigma, \mathcal{N}) \geq -\int_{-\infty}^{\infty} dt \ \beta_{(1-\alpha)/\alpha}(t) \ \ln \left[F \left(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho) \right) \right]. \tag{54}$$

Since the inequality in (53) holds for all $\theta \in (0,1)$ and thus (54) holds for all $\alpha \in (1/2, 1)$, we can take the limit as $\alpha \nearrow 1$ and apply (40) and the dominated convergence theorem to conclude that (47) holds. □

As a corollary of the above theorem, we obtain equality conditions for the monotonicity of quantum relative entropy:

Corollary 13 (Equality Conditions)**.** *Let ρ, σ, and $\mathcal{N}$ be as given in Definition 1. Then*

$$D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \tag{55}$$

if and only if all rotated Petz recovery maps perfectly recover ρ from $\mathcal{N}(\rho)$:

$$\forall t \in \mathbb{R} : (\mathcal{R}^t_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho) = \rho. \tag{56}$$

Proof. Recall from Exercise 10 that, independent of the conditions in the statement of the corollary, we always have that $(\mathcal{R}^t_{\sigma,\mathcal{N}} \circ \mathcal{N})(\sigma) = \sigma$ for all $t \in \mathbb{R}$.

We start by proving the "only if" part. Suppose that $\forall t \in \mathbb{R} : (\mathcal{R}^t_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho) = \rho$. Then for a particular $t \in \mathbb{R}$, the monotonicity of quantum relative entropy implies that

$$\begin{aligned} D(\rho\|\sigma) &\geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) && (57)\\ D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) &\geq D((\mathcal{R}^t_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)\|(\mathcal{R}^t_{\sigma,\mathcal{N}} \circ \mathcal{N})(\sigma)) && (58)\\ &= D(\rho\|\sigma), && (59) \end{aligned}$$

which in turn imply that $D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$.

We now prove the "if part" of the theorem. Suppose that $D(\rho\|\sigma) = D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$. By Theorem 12, we can conclude that

$$\int_{-\infty}^{\infty} dt\ \beta_0(t)\ \left[-\log\left[F\left(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)\right)\right]\right] = 0. \tag{60}$$

Since $\beta_0(t)$ is a positive definite function for all $t \in \mathbb{R}$ and $-\log F \geq 0$, we can conclude that

$$-\log\left[F\left(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)\right)\right] = 0 \tag{61}$$

for all $t \in \mathbb{R}$, which is the same as $F(\rho, (\mathcal{R}^{t/2}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)) = 1$ for all $t \in \mathbb{R}$. We can then conclude that (56) holds because the fidelity between two states is equal to one if and only if the states are the same. □

7 Refinements of Quantum Entropy Inequalities

Theorem 2 leads to a strengthening of many quantum entropy inequalities, including strong subadditivity of quantum entropy, concavity of conditional entropy, joint convexity of relative entropy, non-negativity of quantum discord, and the Holevo bound. We go through the refinement of strong subadditivity and point to the book for the others.

7.1 Strong Subadditivity

Recall the conditional quantum mutual information of a tripartite state ρ_{ABC}:

$$I(A;B|C)_\rho \equiv H(AC)_\rho + H(BC)_\rho - H(C)_\rho - H(ABC)_\rho. \tag{62}$$

Strong subadditivity is the statement that $I(A;B|C)_\rho \geq 0$ for all tripartite states ρ_{ABC}.

Corollary 14 below gives an improvement of strong subadditivity. It is a direct consequence of Theorem 2 after choosing

$$\rho = \rho_{ABC}, \quad \sigma = \rho_{AC} \otimes I_B, \quad \mathcal{N} = \mathrm{Tr}_A, \tag{63}$$

so that

$$\mathcal{N}(\rho) = \rho_{BC}, \quad \mathcal{N}(\sigma) = \rho_C \otimes I_B, \quad \mathcal{N}^\dagger(\cdot) = (\cdot) \otimes I_A, \tag{64}$$

and

$$\begin{aligned} D(\rho\|\sigma) - D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) &= D(\rho_{ABC}\|\rho_{AC} \otimes I_B) - D(\rho_{BC}\|\rho_C \otimes I_B) \\ &= I(A;B|C)_\rho, \end{aligned} \tag{65}$$

$$\begin{aligned} \mathcal{P}_{\sigma,\mathcal{N}}(\cdot) &= \sigma^{1/2}\mathcal{N}^\dagger\left([\mathcal{N}(\sigma)]^{-1/2}(\cdot)[\mathcal{N}(\sigma)]^{-1/2}\right)\sigma^{1/2} \\ &= \rho_{AC}^{1/2}\left[\rho_C^{-1/2}(\cdot)\rho_C^{-1/2} \otimes I_A\right]\rho_{AC}^{1/2}. \end{aligned} \tag{66}$$

Corollary 14. *Let $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$. Then the following inequality holds*

$$I(A;B|C)_\rho \geq -\log\left[F(\rho_{ABC}, \mathcal{R}_{C\to AC}(\rho_{BC}))\right], \tag{67}$$

where the recovery channel $\mathcal{R}_{C\to AC} = \int_{-\infty}^{\infty} dt\ \beta_0(t)\ \mathcal{R}_{\rho_{AC},\mathrm{Tr}_A}^{t/2}$ perfectly recovers ρ_{AC} from ρ_C.