Recovery map stability for the Data Processing Inequality

Eric A. Carlen[1] and Anna Vershynina[2]

[1] *Department of Mathematics, Hill Center, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA*

[2] *Department of Mathematics, Philip Guthrie Hoffman Hall, University of Houston, 3551 Cullen Blvd., Houston, TX 77204-3008, USA*

November 6, 2019

Abstract

Let $\mathcal{M}$ be a finite dimensional von Neumann algebra and $\mathcal{N}$ a von Neumann subalgebra of it. For states ρ and σ on $\mathcal{M}$, let $\rho_{\mathcal{N}}$ and $\sigma_{\mathcal{N}}$ be the corresponding states induced on $\mathcal{N}$. The Data Processing Inequality (DPI) implies that $S(\rho||\sigma) \geq S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}})$ where $S(\rho||\sigma) := \mathrm{Tr}[\rho(\log\rho - \log\sigma)]$ is the relative entropy. Petz proved that there is equality if and only if $\sigma = \mathscr{R}_\rho(\sigma_{\mathcal{N}})$, where $\mathscr{R}_\rho$ is the *Petz recovery map*. We prove a quantitative version of Petz's theorem. In it simplest form, our bound is

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{8}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2} \|\mathscr{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1^4 \ .$$

where $\Delta_{\sigma,\rho}$ is the relative modular operator. Since $\|\Delta_{\sigma,\rho}\| \leq \|\rho^{-1}\|$, this yields a bound that is independent of σ. We also prove an analogous result with a more complicated constant in which the roles of ρ and σ are interchanged on the right.

Quantum information theoretic inequalities are usually much harder to prove, or differ from, their classical counterparts because classical proofs often rely on conditioning argument that do not carry over to the quantum setting. In particular, quantum conditional expectations rarely preserve expectations – something that always happens in the classical setting. We also prove a simple theorem characterizing states ρ and subalgebras $\mathcal{N}$ for which conditional expectations do preserve expectation with respect to ρ, illuminating the quantum obstacle to the existence of nicely behaved conditional expectations and the origins the Petz recovery map.

1 Introduction

1.1 The Data Processing Inequality

Let $\mathcal{M}$ be a finite dimensional von Neumann algebra, which we may regard as a subalgebra of $M_n(\mathbb{C})$, the $n \times n$ complex matrices. The Hilbert-Schmidt inner product $\langle \cdot, \cdot \rangle_{HS}$ on $M_n(\mathbb{C})$ is given in terms of the trace by $\langle X, Y\rangle_{HS} = \mathrm{Tr}[X^*Y]$. Let $\mathbf{1}$ denote the identity.

A state on $\mathcal{M}$ is a linear functional φ on $\mathcal{M}$ such that $\varphi(A^*A) \geq 0$ for $A \in \mathcal{M}$ and such that $\varphi(\mathbf{1}) = 1$. A state φ is *faithful* in case $\varphi(A^*A) > 0$ whenever $A \neq 0$, and is *tracial* in case $\varphi(AB) = \varphi(BA)$ for all $A, B \in \mathcal{M}$. Every state on $\mathcal{M}$ is of the form $X \mapsto \mathrm{Tr}[\rho X]$, where ρ is a *density matrix* in $\mathcal{M}$; i.e., a non-negative element ρ of $\mathcal{M}$ such that $\mathrm{Tr}[\rho] = 1$. This state is faithful if and only if ρ is invertible. It will be convenient to write $\rho(X) = \mathrm{Tr}[\rho X]$ to denote the state corresponding to a density matrix ρ. Given a faithful state ρ, the corresponding Gelfand-Naimark-Segal (GNS) inner product is given by $\langle X, Y\rangle_{GNS,\rho} := \rho(X^*Y)$.

In this finite dimensional setting, there is always a faithful tracial state τ on $\mathcal{M}$, namely the one whose density matrix is $n^{-1}\mathbf{1}$. The symbol τ is reserved throughout for this tracial state.

Let $\mathcal{N}$ be a von Neumann subalgebra of $\mathcal{M}$. Let $\mathscr{E}$ be any norm-contractive projection from $\mathcal{M}$ onto $\mathcal{N}$. (Norm contractive means that $\|\mathscr{E}(X)\| \leq \|X\|$ for all $X \in \mathcal{M}$. Throughout the paper, $\|\cdot\|$ without any subscript denotes the operator norm.) By a theorem of Tomiyama [38], $\mathscr{E}$ preserves positivity, $\mathscr{E}(\mathbf{1}) = \mathbf{1}$, and

$$\mathscr{E}(AXB) = A\mathscr{E}(X)B \qquad \text{for all} \quad A, B \in \mathcal{N},\ X \in \mathcal{M}\ . \tag{1.1}$$

Moreover, as Tomiyama noted, it follows from (1.1) and the positivity preserving property of $\mathscr{E}$ that

$$\mathscr{E}(X)^*\mathscr{E}(X) \leq \mathscr{E}(X^*X) \qquad \text{for all } X \in \mathcal{M}\ , \tag{1.2}$$

In fact, more is true. As is well known, every norm contractive projection is *completely positive*.

A *conditional expectation* from $\mathcal{M}$ onto $\mathcal{N}$, in the sense of Umegaki [42, 43, 44], is a unital projection from $\mathcal{M}$ onto $\mathcal{N}$ that is order preserving and such that (1.1) and (1.2) are satisfied. Since every conditional expectation $\mathscr{E}$ is a unital completely positive map, its adjoint with respect to the Hilbert-Schmidt inner product, $\mathscr{E}^\dagger$, is a completely positive trace preserving (CPTP) map, also known as a *quantum operation*. (Throughout this paper, a dagger † always denotes the adjoint with respect to the Hilbert-Schmidt inner product.)

Let $\mathscr{E}_\tau$ denote the orthogonal projection from $\mathcal{M}$ onto $\mathcal{N}$ with respect to the GNS inner product determined by τ. It is easy to see, using the tracial nature of τ, that $\mathscr{E}_\tau$ is in fact a conditional expectation, and since $\mathscr{E}_\tau = \mathscr{E}_\tau^\dagger$, $\mathscr{E}_\tau$ is a quantum operation.

1.1 Definition. For any state ρ on $\mathcal{M}$, $\rho_{\mathcal{N}}$ denotes the state on $\mathcal{N}$ given by $\rho_{\mathcal{N}} := \mathscr{E}_\tau(\rho)$ where, as always $\mathscr{E}_\tau$ denotes the tracial conditional expectation onto $\mathcal{N}$.

The restriction of a state ρ on $\mathcal{M}$ to $\mathcal{N}$ is of course a state on $\mathcal{N}$, and as such, it is represented by a unique density matrix belonging to $\mathcal{N}$, which is precisely $\rho_{\mathcal{N}}$.

1.2 Example. Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ be the tensor product of two finite dimensional Hilbert spaces. Let $\mathcal{M} = \mathcal{B}(\mathcal{H})$ be the algebra of all linear transformations on $\mathcal{H}$, and let $\mathcal{N}$ be the subalgebra $\mathbf{1}_{\mathcal{H}_1} \otimes \mathcal{B}(\mathcal{H}_2)$ consisting of all operators in $\mathcal{M}$ of the form $\mathbf{1}_{\mathcal{H}_1} \otimes A$, $A \in \mathcal{B}(\mathcal{H}_2)$. Then for the normalized trace τ, $\mathcal{E}_\tau(X) = d_1^{-1}\mathbf{1}_{\mathcal{H}_1} \otimes \mathrm{Tr}_1 X$ for all $X \in \mathcal{M}$ where d_1 is the dimension of $\mathcal{H}_1$ and where Tr_1 denotes the partial trace over $\mathcal{H}_1$.

Given two states ρ and σ on $\mathcal{M}$, the *Umegaki relative entropy of ρ with respect to σ* is defined [45] by

$$S(\rho||\sigma) := \mathrm{Tr}[\rho(\log\rho - \log\sigma)] \ . \tag{1.3}$$

Lindblad's inequality [20] states that with $\mathscr{E}_\tau$ being the tracial conditional expectation onto $\mathcal{N}$,

$$S(\rho||\sigma) \geq S(\mathscr{E}_\tau(\rho)||\mathscr{E}_\tau(\sigma)) \ . \tag{1.4}$$

Lindblad showed that the monotonicity (1.4) is equivalent to the joint convexity of the relative entropy $(\rho,\sigma) \mapsto S(\rho||\sigma)$, and this in turn is an immediate consequence of Lieb's Concavity Theorem [17]. In the case that $\mathcal{M} = \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ and $\mathcal{N} = \{\mathbf{1}_{\mathcal{H}_1} \otimes A,\ A \in \mathcal{B}(\mathcal{H}_2)\}$, (1.4) was proved by Lieb and Ruskai [19], who showed it to be equivalent to the Strong Subadditivity (SSA) of the von Neumann entropy.

Using the fact that Stinesping's Dilation Theorem [34] relates general CPTP maps to tracial expectation, Lindblad [21] was able to prove, using (1.4) that for any CPTP map $\mathscr{P}$,

$$S(\rho||\sigma) \geq S(\mathscr{P}(\rho)||\mathscr{P}(\sigma)) \ . \tag{1.5}$$

This is known as the *Data Processing Inequality* (DPI). Because of the simple relation between (1.4) and (1.5) the problem of determining the cases of equality in the Data Processing Inequality largely comes down the problem of determining the cases of equality in (1.4), which was solved by Petz [28, 29]. His necessary and sufficient condition for equality in (1.4) is closely connected with the problem of *quantum coarse graining*, and in particular a quantum coarse graining operation introduced by Acardi and Cecchini [2], whose dual is now now known as the *Petz recovery channel*, the CPTP map $\mathscr{R}_\rho$ given by

$$\mathscr{R}_\rho(\gamma) = \rho^{1/2}(\rho_{\mathcal{N}}^{-1/2}\gamma\rho_{\mathcal{N}}^{-1/2})\rho^{1/2} \ . \tag{1.6}$$

It is obvious that $\mathscr{R}_\rho(\rho_{\mathcal{N}}) = \rho$, so that $\mathscr{R}_\rho$ "recovers" ρ from $\rho_{\mathcal{N}}$. Petz proved [28, 29] that there is equality in (1.4) if and only if

$$\mathscr{R}_\rho(\sigma_{\mathcal{N}}) = \sigma \tag{1.7}$$

and that this is true if and only if

$$\mathscr{R}_\sigma(\rho_{\mathcal{N}}) = \rho \ . \tag{1.8}$$

There has been much recent work on *stability* for for the DPI: Suppose that ρ and σ are such that there is *approximate* equality in (1.4). To what extent do ρ and σ provide approximate solutions to Petz's equation (1.7) and (1.8)? The papers : [14, 18, 33, 35, 36, 48] all address this question.

1.2 Main results

In this paper we further develop an approach that we introduced in [9] for proving stability for analogs of the DPI for Rényi relative entropies. The Rényi relative entropies include the Umegaki relative entropy (1.3) as a limiting case, but taking advantage of the special structure on the Umegaki relative entropy, we are able to sharpen the stability bounds obtained in [9] for this case.

Our results in this direction are given in Theorem 1.5 and Corollary 1.7. These results are proved in Section 2.

Many classical information theoretic inequalities have simple proofs that rely on the use of conditional probabilities and conditional expectations, as we recall in the next subsection. An obstacle to proving quantum analogs of such inequalities is that in the presence of non-commutativity, there simply is no fully satisfactory analog of conditioning. For example, in the classical case, taking any conditional expectation of any random variable on any probability space preserves the expectation – the original random variable and its conditional expectation have the same expected value. A theorem of Takesaki [37] says that in the quantum setting, this is unfortunately only rarely the case. Takesaki's Theorem motivated Accardi and Cechini [2] to develop their expectation preserving quantum coarse graining operation, whose dual turns out to be the Petz recovery map. For these reasons, it is of interest to understand the nature of the obstacle blocking the general existence of expectation preserving conditional expectation. However, Takesaki's paper will only be accessible to readers who are well-versed in the rather subtle Tomita-Takesaki Theory. In section 3 we give a simple proof, valid at least in our finite dimensional setting, of a result that characterizes when expectation preserving conditional expectations will exist, and sheds new light, at least in this setting, on why they often fail to exist. The full statement is given in Theorem 3.1. Before giving precise statements we briefly review that corresponding classical problem.

1.3 The classical DPI

The classical analog of the DPI is relatively simple: Let Ω be a finite set. Let $\mathcal{F}$ be a non-trivial partition of Ω. Let $\mathcal{M}$ denote the functions on Ω, and let $\mathcal{N}$ denote the functions on Ω that are constant on each set of the partition $\mathcal{F}$. Then $\mathcal{M}$ and $\mathcal{N}$ are commutative von Neumann algebras, and $\mathcal{N}$ is a subalgebra of $\mathcal{M}$. Let X be a function on Ω such that $X(\omega) = X(\omega')$ if and only if ω and ω' belong to the same set in $\mathcal{F}$. Then X *generates* $\mathcal{F}$ in the sense that the sets constituting $\mathcal{F}$ are precisely the non-empty sets of the form $\{\omega \ : \ X(\omega) = x\}$.

Let ρ and σ be two strictly positive probability densities on the set Ω. Let τ denote the uniform probability density on Ω; i.e., $\tau(\omega) = |\Omega|^{-1}$ for all ω, where $|\Omega|$ is the cardinality of Ω. As above, let $\mathscr{E}_\tau$ denote the orthogonal projection of $\mathcal{M}$ onto $\mathcal{N}$, which is nothing other than the conditional expectation with respect the random variable X and the probability measure τ. As above, let $\rho_{\mathcal{N}} = \mathscr{E}_\tau\rho$ and $\sigma_{\mathcal{N}} = \mathscr{E}_\tau\sigma$. Then $\rho_{\mathcal{N}}$ is a "coarse grained" version of ρ, obtained by averaging ρ on the sets of the partition $\mathcal{F}$, making it constant on these.

Let $f(\omega|x)$ be the conditional density under ρ for ω given that $X(\omega) = x$, and likewise let $g(\omega|x)$ be the conditional density under σ for ω given that $X(\omega) = x$. That is

$$f(\omega|x) = \frac{\rho(\omega)}{\rho_{\mathcal{N}}(x)} \quad \text{and} \quad g(\omega|x) = \frac{\sigma(\omega)}{\rho_{\mathcal{N}}(x)} \ , \tag{1.9}$$

which, for each x in the range of X, are both probability densities on the set $\{\omega \ : \ X(\omega) = x\}$. Then

$$\rho(\omega) = \rho_{\mathcal{N}}(X(\omega))f(\omega|X(\omega)) \quad \text{and} \quad \sigma(\omega) = \sigma_{\mathcal{N}}(X(\omega))g(\omega|X(\omega)) \ , \tag{1.10}$$

and hence

$$\begin{aligned} S(\rho||\sigma) &= \sum_{\omega\in\Omega} \rho(\omega)(\log\rho(\omega) - \log\sigma(\omega)) \\ &= \sum_{\omega\in\Omega} \rho(\omega)([\log\rho_{\mathcal{N}}(\omega) - \log\sigma_{\mathcal{N}}(\omega)] + [\log f(\omega|X(\omega)) - \log g(\omega|X(\omega))]) \\ &= S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) + \sum_{\omega\in\Omega} \rho_{\mathcal{N}}(X(\omega)) f(\omega|X(\omega)) \left[\log f(\omega|X(\omega)) - \log g(\omega|X(\omega))\right] \quad (1.11) \end{aligned}$$

For each x in the range of X, it follows from Jensen's inequality that

$$\sum_{\{\omega\ :\ X(\omega)=x\}} f(\omega|x)\left[\log f(\omega|x) - \log g(\omega|x)\right] \geq 0\ , \tag{1.12}$$

and there is equality if and only if $f(\omega|x) = g(\omega|x)$ everywhere on $\{\omega\ :\ X(\omega) = x\}$. It follows that $S(\rho||\sigma) \geq S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}})$ with equality if and only if for each x in the range of X, $f(\omega|x) = g(\omega|x)$ everywhere on $\{\omega\ :\ X(\omega) = x\}$.

In this case, X is called a *sufficient statistic* for the pair $\{\rho,\sigma\}$: Suppose we are given an independent identically distributed sequence of points $\{\omega_j\}$, drawn according to one of the two probability densities ρ or σ, and we want to determine which it is. Because of (1.10) and $f = g$, the entire difference between $\rho(\omega)$ and $\sigma(\omega)$, lies in the difference between $\rho_{\mathcal{N}}(X(\omega))$ and $\sigma_{\mathcal{N}}(X(\omega))$. Therefore, it suffices to observe the sequence $\{X(\omega_j)\}$ in order to determine which of ρ or σ is governing the sequence of samples.

We can define a classical recovery map as follows: For any probability density $\gamma\in\mathcal{N}$, regarded as a probability density on the range of X, define $\mathscr{R}_\rho\gamma$ to be the probability density in $\mathcal{M}$ given by

$$\mathscr{R}_\rho\gamma(\omega) = \gamma(X(\omega))f(\omega|X(\omega))\ .$$

Therefore, we can express the condition for equality in the classical DPI as $\mathscr{R}_\rho\sigma_{\mathcal{N}} = \sigma$, and evidently this is true if and only if $\mathscr{R}_\sigma\rho_{\mathcal{N}} = \rho$. This is the classical analog of Petz's result. Moreover, in this notation, we have that

$$\sum_{\omega\in\Omega} \rho_{\mathcal{N}}(X(\omega)) f(\omega|X(\omega)) \left[\log f(\omega|X(\omega)) - \log g(\omega|X(\omega))\right] = S(\rho||\mathscr{R}_\sigma\rho_{\mathcal{N}})\ ,$$

so that (1.11) becomes

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq S(\rho||\mathscr{R}_\sigma\rho)\ . \tag{1.13}$$

Then by the classical Pinsker inequality,

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \frac{1}{2}\left(\sum_{\omega\in\Omega} |\rho(\omega) - \mathscr{R}_\sigma\rho_{\mathcal{N}}(\omega)|\right)^2\ . \tag{1.14}$$

It remains an open problem to prove quantum analogs of (1.13) or (1.14), even with worse constants on the right. In the case of (1.13), some modification such as worse constants is certainly required; the exact quantum analog is violated in some numerical examples; see [13]. Here we prove a quantum analog of (1.14) with a worse constant, and with the power raised from 2 to 4 on the right. The line of argument has to be entirely different from the one we have just employed in the classical case because there is no effective quantum replacement for "conditioning on the observable X".

It is therefore useful to find a way of describing the classical recovery map that does not refer explicitly to conditioning on the random variable X. Define $\mathscr{E}_\sigma$ to be the orthogonal projection of $\mathcal{M}$ onto $\mathcal{N}$ in $L^2(\sigma)$. Then for any random variable Y (i.e., any function on Ω), $\mathscr{E}_\sigma Y$ is the conditional expectation of Y given $\mathcal{N}$. The operation $Y \mapsto \mathscr{E}_\sigma Y$ yields a "coarse grained version" of Y that is constant on the sets in $\mathcal{F}$: With X and g as in (1.9),

$$\mathscr{E}_\sigma Y(\omega) = \sum_{\omega':\ X(\omega')=X(\omega)\}} g(\omega'|X(\omega)Y(\omega')\ .$$

It is clear from this formula that $\mathscr{E}_\sigma$ preserves positivity, and preserves expectations with respect to σ, That is,

$$\sigma(Y) = \sigma(\mathscr{E}_\sigma Y)\ . \tag{1.15}$$

Now let $\mathscr{E}_\sigma^\dagger$ be the dual operation taking states on $\mathcal{N}$ (probability densities on the range of X) to states on $\mathcal{M}$ (probability densities on Ω). It is easily seen that this is nothing other than $\mathscr{R}_\sigma$. That is, the classical recovery map $\mathscr{R}_\sigma$ is nothing other than the dual of the conditional expectation $\mathscr{E}_\sigma$, which is nothing other than the orthogonal projection of $\mathcal{M}$ onto $\mathcal{N}$ in $L^2(\sigma)$. This analytic specification of $\mathscr{R}_\sigma$, making no explicit mention of conditioning on X, provides a starting point for the construction of a quantum recovery map.

1.4 Quantum conditional expectations and quantum coarse graining

The discussion of the classical DPI brings us to the question as to whether for any faithful state ρ on $\mathcal{M}$ there exists a conditional expectation $\mathscr{E}$ from $\mathcal{M}$ onto $\mathcal{N}$ that preserves expectations with respect to ρ, i.e. such that

$$\rho(X) = \rho(\mathscr{E}(X)) \qquad \text{for all } X \in \mathcal{M}\ . \tag{1.16}$$

The property (1.16) says that "the expectation of a conditional expectation of an observable equals the expectation of the observable". If such a conditional expectation exists, then it is unique: *Any such conditional expectation must be the orthogonal projection of $\mathcal{M}$ on $\mathcal{N}$ with respect to the GNS inner product for the state ρ.* To see this, note that for all $X \in \mathcal{M}$ and all $A \in \mathcal{N}$, using (1.1),

$$\langle A, X\rangle_{GNS,\rho} := \rho(A^*X) = \rho(\mathscr{E}(A^*X)) = \rho(A^*\mathscr{E}(X)) = \langle A, \mathscr{E}(X)\rangle_{GNS,\rho}\ . \tag{1.17}$$

Suppose that $\mathscr{E}$ is a conditional expectation satisfying (1.16). Then since $\mathscr{E}$ is a unital completely positive map, $\mathscr{E}^\dagger$ is a CPTP map; i.e., a quantum channel. For any state γ on $\mathcal{N}$, and any $A \in \mathcal{M}$, we then have

$$\mathscr{E}^\dagger(A) = \gamma(\mathscr{E}(A))\ .$$

Then when (1.16) is satisfied, taking $\gamma = \rho_{\mathcal{N}}$, we have

$$\mathscr{E}^\dagger \rho_{\mathcal{N}}(A) = \rho_{\mathcal{N}}(\mathscr{E}(A)) = \rho(A) \ ,$$

and this means that $\mathscr{E}^\dagger$ is a quantum channel that "recovers" ρ from $\rho_{\mathcal{N}}$.

As we have already noted, $\mathscr{E}_\tau$ is a conditional expectation with the property (1.16). However, for non-tracial states ρ, a conditional expectation satisfying (1.16) need not exist.

A theorem of Takesaki [37] says, in our finite dimensional context, that for a faithful state ρ, there exists a conditional expectation $\mathscr{E}$ from $\mathcal{M}$ onto $\mathcal{N}$ if and only if $\rho A \rho^{-1} \in \mathcal{N}$ for all $A \in \mathcal{N}$, and in general this is not the case. We give a short proof of this and somewhat more in Section 3: In Theorem 3.1, we prove that $\mathscr{E}_\rho$, the orthogonal projection from $\mathcal{M}$ onto $\mathcal{N}$ in the GNS inner product with respect to ρ, is *real* (that is, it preserves self-adjointness) if and only if $\rho A \rho^{-1} \in \mathcal{N}$ for all $A \in \mathcal{N}$. Since every order preserving linear transformation is real, this precludes the general existence of conditional expectations satisfying (1.16) whenever $\mathcal{N}$ is not invariant under $X \mapsto \rho X \rho^{-1}$, thus implying Takesaki's Theorem (in this finite dimensional setting).

There is another inner product on $\mathcal{M}$ that is naturally induced by a faithful state ρ, namely the *Kubo-Martin-Schwinger (KMS) inner product*. It is defined by

$$\langle X, Y \rangle_{KMS,\rho} = \mathrm{Tr}[\rho^{1/2} X^* \rho^{1/2} Y] = \mathrm{Tr}[(\rho^{1/4} X \rho^{1/4})^* (\rho^{1/4} Y \rho^{1/4})] \ . \tag{1.18}$$

Evidently, for any $X \in \mathcal{M}$ and $Y \in \mathcal{N}$, by the Cauchy-Schwarz inequality,

$$|\langle X, Y \rangle_{KMS,\rho}|^2 = |\mathrm{Tr}[\rho^{1/2} X^* \rho^{1/2} Y]|^2 \leq \mathrm{Tr}[X^* \rho^{1/2} X \rho^{1/2}]\mathrm{Tr}[Y^* \rho^{1/2} Y \rho^{1/2}] \ . \tag{1.19}$$

If U is any unitary in $\mathcal{N}'$, the commutant of $\mathcal{N}$,

$$\mathrm{Tr}[Y^* \rho^{1/2} Y \rho^{1/2}] = \mathrm{Tr}[U^*(Y^* \rho^{1/2} Y \rho^{1/2})U] = \mathrm{Tr}[Y^*(U^* \rho U)^{1/2} Y (U^* \rho U)^{1/2}] \ .$$

Then since $\rho_{\mathcal{N}}$ can be written as an average of $U^* \rho U$ as U ranges over the unitaries in $\mathcal{N}'$, [39], it follows from the Lieb Concavity Theorem that $\mathrm{Tr}[Y^* \rho^{1/2} Y \rho^{1/2}] \leq \mathrm{Tr}[Y^* \rho_{\mathcal{N}}^{1/2} Y \rho_{\mathcal{N}}^{1/2}]$. Combining this with (1.19),

$$|\langle X, Y \rangle_{KMS,\rho}| \leq \|X\|_{KMS,\rho} \|Y\|_{KMS,\rho_{\mathcal{N}}} \tag{1.20}$$

Hence $Y \mapsto \langle X, Y \rangle_{KMS,\rho}$ is a bounded linear functional on $(\mathcal{N}, \langle \cdot\, , \cdot \rangle_{KMS,\rho_{\mathcal{N}}})$, and then there is a uniquely determined $\mathscr{A}_\rho(X) \in \mathcal{N}$ such that for all $Y \in \mathcal{N}$,

$$\langle X, Y \rangle_{KMS,\rho} = \langle \mathscr{A}_\rho(X), Y \rangle_{KMS,\rho_{\mathcal{N}}} \ . \tag{1.21}$$

Evidently, $X \mapsto \mathscr{A}_\rho(X)$ is linear.

1.3 Definition. Let ρ be a faithful state on $\mathcal{M}$. The *Accardi-Cecchini coarse graining operator* $\mathscr{A}_\rho$ from $\mathcal{M}$ to $\mathcal{N}$ is defined by (1.21).

The map $\mathscr{A}_\rho$ was introduced by Accardi and Cecchini [2], building on previous work by Accardi [1]. It is a "coarse graining" operation in that to each observable X in the larger algebra $\mathcal{M}$, it associates an observable $\mathscr{A}_\rho(X)$, in the smaller algebra $\mathcal{N}$, and measurement of $\mathscr{A}_\rho(X)$ will yield coarser information than a measurement of X itself. The same, of course, is true for conditional expectations,.

Since $\mathbf{1} \in \mathcal{N}$ by definition, for all $X \in \mathcal{M}$, $\langle \mathbf{1}, X\rangle_{KMS,\rho} = \langle \mathbf{1}, \mathscr{A}_\rho(X)\rangle_{KMS,\rho}$, and for all $X \in \mathcal{M}$, $\langle \mathbf{1}, X\rangle_{KMS,\rho} = \mathrm{Tr}[\sigma^{1/2}\mathbf{1}\sigma^{1/2}X] = \rho(X)$. Therefore

$$\rho(\mathscr{A}_\rho(X)) = \rho(X) \ . \tag{1.22}$$

Thus, unlike conditional expectations in general, the Accardi-Cecchini coarse-graining operator *always* preserves expectations with respect to ρ.

In the matricial setting, it is a particularly simple matter to derive an explicit expression for $\mathscr{A}_\rho$. By definition, for all $X \in \mathcal{M}$ and all $Y \in \mathcal{N}$,

$$\mathrm{Tr}[\rho_{\mathcal{N}}^{1/2} Y \rho_{\mathcal{N}}^{1/2} \mathscr{A}_\rho(X)] = \mathrm{Tr}[\rho^{1/2} Y \rho^{1/2} X] \ . \tag{1.23}$$

Make the change of variables $Z = \rho_{\mathcal{N}}^{1/2} Y \rho_{\mathcal{N}}^{1/2}$. Since $\rho_{\mathcal{N}}^{1/2}$ is invertible, as Y ranges over $\mathcal{N}$, Z ranges over $\mathcal{N}$. Hence

$$\mathrm{Tr}[Z\mathscr{A}_\rho(X)] = \mathrm{Tr}[(\rho^{1/2}\rho_{\mathcal{N}}^{-1/2} Z \rho_{\mathcal{N}}^{-1/2}\rho^{1/2})X] = \mathrm{Tr}[Z(\rho_{\mathcal{N}}^{-1/2}\rho^{1/2} X \rho^{1/2}\rho_{\mathcal{N}}^{-1/2})] \ . \tag{1.24}$$

Since the above holds for all $Z \in \mathcal{N}$, it follows that

$$\mathscr{A}_\rho(X) = \rho_{\mathcal{N}}^{-1/2}\mathscr{E}_\tau(\rho^{1/2} X \rho^{1/2})\rho_{\mathcal{N}}^{-1/2} \ . \tag{1.25}$$

It is evident from this formula that $\mathscr{A}_\rho$ is a completely positive unital map from $\mathcal{M}$ to $\mathcal{N}$, and therefore it is actually a contraction from $\mathcal{M}$ to $\mathcal{N}$. By Tomiyama's Theorem, it cannot in general be a projection of $\mathcal{M}$ onto $\mathcal{N}$. That is, if $X \in \mathcal{N}$, it is not necessarily the case that $\mathscr{A}_\rho(X) = X$.

1.4 Definition. The *Petz recovery map* $\mathscr{R}_\rho$ is the Hilbert-Schmidt adjoint of $\mathscr{A}_\rho$ [29]. That is, $\mathscr{R}_\rho = \mathscr{A}_\rho^\dagger$, or equivalently, for all density matrices $\gamma \in \mathcal{N}$

$$\mathrm{Tr}[\gamma\mathscr{A}_\rho(X)] = \mathrm{Tr}[\mathscr{R}_\rho(\gamma)X] \ .$$

As the dual of a unital completely positive map, $\mathscr{R}_\rho$ is a CPTP map. Moreover, it follows immediately from the definition and (1.25) that for all density matrices $\gamma \in \mathcal{N}$,

$$\mathscr{R}_\rho(\gamma) = \rho^{1/2}(\rho_{\mathcal{N}}^{-1/2}\gamma\rho_{\mathcal{N}}^{-1/2})\rho^{1/2} \ . \tag{1.26}$$

It is evident from this formula not only that $\mathscr{R}_\rho$ is a CPTP map, but that $\mathscr{R}_\rho(\rho_{\mathcal{N}}) = \rho$; i.e., $\mathscr{R}_\rho$ recovers ρ from $\rho_{\mathcal{N}}$. Now suppose that σ is another density matrix in $\mathcal{M}$ and that

$$\mathscr{R}_\rho(\sigma_{\mathcal{N}}) = \sigma \ . \tag{1.27}$$

Then by (1.27) and then the DPI, $S(\rho||\sigma) = S(\mathscr{R}_\rho(\rho_{\mathcal{N}})||\mathscr{R}_\rho(\sigma_{\mathcal{N}})) \leq S(\rho||\sigma)$. Hence when $\mathscr{R}_\rho(\sigma_{\mathcal{N}}) = \sigma$, there is equality in (1.4). The deeper result of Petz [28, 29] is that there is equality in (1.4) *only* in this case.

As noted above, $\mathscr{A}_\rho$ is not a projection onto $\mathcal{N}$. The fixed point set

$$\mathcal{C} := \{X \in \mathcal{N} : \ \mathscr{A}_\rho(X) = X \ \} \tag{1.28}$$

turns out to be a von Neumann subalgebra of $\mathcal{N}$, and is structure, already investigated in [2], plays an important role in classifying the set of states σ such that $\mathscr{R}_\rho(\sigma_{\mathcal{N}}) = \sigma$; see [24]. This classification gives a complete understanding of the cases of equality, and the remaining open problems concern stability.

Here we prove a stability bound for Petz's theorem on the cases of equality in (1.4). Our result involves the *relative modular operator* $\Delta_{\rho,\sigma}$ on $\mathcal{M}$ defined by

$$\Delta_{\sigma,\rho}(X) = \sigma X \rho^{-1} \tag{1.29}$$

for all $X \in \mathcal{M}$. This is the matricial version of an operator introduced in a more general von Neumann algebra context by Araki [3]. Our main result is:

1.5 Theorem. *Let ρ and σ be two states on $\mathcal{M}$ Let $\mathscr{E}_\tau$ be the tracial conditional expectation onto a von Neumann subalgebra $\mathcal{N}$, and let $\rho_{\mathcal{N}} = \mathscr{E}_\tau\rho$ and $\sigma_{\mathcal{N}} = \mathscr{E}_\tau\sigma$. Then, with $\|\cdot\|_2$ denoting the Hilbert-Schmidt norm,*

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{4}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2} \|\sigma_{\mathcal{N}}^{1/2}\rho_{\mathcal{N}}^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2^4 \ . \tag{1.30}$$

The quantity on the right hand side may be estimated in terms of the Petz recovery map. In Section 2 we prove:

1.6 Lemma. *Let ρ, σ and $\sigma_{\mathcal{N}}$ be specified as in Theorem 1.5. Then, with $\|\cdot\|_1$ denoting the trace norm,*

$$\|(\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \geq \frac{1}{2}\|\mathscr{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1 \ .$$

As an immediate Corollary of Theorem 1.5 and Lemma 1.6, we obtain

1.7 Corollary. *Let ρ and σ be two states on $\mathcal{M}$. Let $\mathscr{E}_\tau$ be the tracial conditional expectation onto a von Neumann subalgebra $\mathcal{N}$, and let $\rho_{\mathcal{N}} = \mathscr{E}_\tau\rho$ and $\sigma_{\mathcal{N}} = \mathscr{E}_\tau\sigma$. Then, with $\|\cdot\|_1$ denoting the trace norm,*

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{8}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2} \|\mathscr{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1^4 \ . \tag{1.31}$$

Note how the right hand side of (1.31) differs from the right hand side of (1.14): Apart from the constant and the power 4 in place of 2, the most striking difference is that the roles of ρ and σ are reversed. The expected result, with a worse constant, is obtained in Corollary 1.8.

Recall that the modular operator is the right multiplication by ρ^{-1} and left multiplication by σ, so $\|\Delta_{\sigma,\rho}\| \leq \|\rho^{-1}\|$, since $\|\sigma\| \leq 1$. While $\|\rho^{-1}\|$ might be considerably larger than $\|\Delta_{\sigma,\rho}\|$, a bound in terms of $\|\rho^{-1}\|$ has the merit that it is independent of σ:

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{8}\right)^4 \|\rho^{-1}\|^{-2} \|\mathscr{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1^4 \ . \tag{1.32}$$

Corollary 1.7 yields a result of Petz: With $\mathcal{M}$, $\mathcal{N}$, ρ and σ as above, $S(\rho||\sigma) = S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}})$ if and only if σ satisfies the Petz equation

$$\sigma = \mathscr{R}_\rho(\sigma_{\mathcal{N}}) \ . \tag{1.33}$$

Theorem 1.5 gives what appears to be a stronger condition on relating ρ, σ, $\rho_{\mathcal{N}}$ and $\sigma_{\mathcal{N}}$, namely that

$$\rho_{\mathcal{N}}^{-1/2}\rho^{1/2} = \sigma_{\mathcal{N}}^{-1/2}\sigma^{1/2} \ . \tag{1.34}$$

While validity of (1.34) immediately implies that σ satisfies the Petz equation (1.33), the converse is also true: By what we have noted above, when (1.33) is satisfied, $S(\rho||\sigma) = S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}})$, and then by Theorem 1.5, (1.34) is satisfied.

This may be made quantitative as follows: Letting L_A denote the operator of left multiplication by A,

$$L_{\rho_{\mathcal{N}}^{1/2}}L_{\sigma_{\mathcal{N}}^{-1/2}}(\sigma_{\mathcal{N}}^{1/2}\rho_{\mathcal{N}}^{-1/2}\rho^{1/2} - \sigma^{1/2}) = (\rho^{1/2} - \rho_{\mathcal{N}}^{1/2}\sigma_{\mathcal{N}}^{-1/2}\sigma^{1/2}) \ ,$$

and hence

$$\|\rho^{1/2} - \rho_{\mathcal{N}}^{1/2}\sigma_{\mathcal{N}}^{-1/2}\sigma^{1/2}\|_2 \leq \|L_{\rho_{\mathcal{N}}^{1/2}}\|\|L_{\sigma_{\mathcal{N}}^{-1/2}}\|\|\sigma_{\mathcal{N}}^{1/2}\rho_{\mathcal{N}}^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \ . \tag{1.35}$$

Since $\|L_{\rho_{\mathcal{N}}^{1/2}}\| = \|\rho_{\mathcal{N}}\|^{1/2}$ and $\|L_{\sigma_{\mathcal{N}}^{-1/2}}\| = \|\sigma_{\mathcal{N}}^{-1}\|^{1/2}$, we may combine (1.35) with (1.30) to obtain

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{4}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\|\rho_{\mathcal{N}}\|^{-2}\|\sigma_{\mathcal{N}}^{-1}\|^{-2}\|\rho_{\mathcal{N}}^{1/2}\sigma_{\mathcal{N}}^{-1/2}\sigma^{1/2} - \rho^{1/2}\|_2^4 \ , \tag{1.36}$$

which is the analog of (1.30) with a somewhat worse constant on the right, but the roles of ρ and σ interchanged there. Applying Lemma 1.6 once more, we obtain

1.8 Corollary. *Let ρ and σ be two states on $\mathcal{M}$. Let $\mathscr{E}_\tau$ be the tracial conditional expectation onto a von Neumann subalgebra $\mathcal{N}$, and let $\rho_{\mathcal{N}} = \mathscr{E}_\tau\rho$ and $\sigma_{\mathcal{N}} = \mathscr{E}_\tau\sigma$. Then*

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{8}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\|\rho_{\mathcal{N}}\|^{-2}\|\sigma_{\mathcal{N}}^{-1}\|^{-2}\|\mathscr{R}_\sigma(\rho_{\mathcal{N}}) - \rho\|_1^4 \ . \tag{1.37}$$

As above, bounding the norms of states by 1, we get a constant that depends only on the smallest eigenvalues of ρ and $\sigma_{\mathcal{N}}$

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{8}\right)^4 \|\rho^{-1}\|^{-2}\|\sigma_{\mathcal{N}}^{-1}\|^{-2}\|\mathscr{R}_\sigma(\rho_{\mathcal{N}}) - \rho\|_1^4 \ . \tag{1.38}$$

We noted above that σ solves the Petz equation if and only if (1.34) is satisfied, and then since (1.34) is symmetric in ρ and σ, $\sigma = \mathscr{R}_\rho\sigma_{\mathcal{N}}$ if and only if $\rho = \mathscr{R}_\sigma\rho_{\mathcal{N}}$, and hence

$$S(\rho||\sigma) = S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \iff S(\sigma||\rho) = S(\sigma_{\mathcal{N}}||\rho_{\mathcal{N}}) \ . \tag{1.39}$$

In physical applications, instead of the trace distance, one often considers an alternative measure of the closeness between two quantum states, the *fidelity* [40] , which for two states ρ and σ on $\mathcal{B}(\mathcal{H})$ is defined as

$$F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2. \tag{1.40}$$

By the Cauchy-Schwarz inequality, $0 \leq F(\rho,\sigma) \leq 1$, and $F(\rho,\sigma) = 1$ if and only if $\rho = \sigma$, and $F(\rho,\sigma) = 0$ if and only is the support of ρ is orthogonal to the support of σ. So in other words, the fidelity is zero when states are perfectly distinguishable, and one when they cannot be distinguished. Moreover, there is a relation between the trace distance $\|\rho - \sigma\|_1$ and fidelity

$$1 - \sqrt{F(\rho,\sigma)} \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho,\sigma)}\ . \tag{1.41}$$

From here and the Corollary 1.7 we obtain the quantitative version of the Petz's Theorem involving the fidelity between states

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{4}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\left(1 - \sqrt{F(\sigma, \mathscr{R}_\rho(\sigma_{\mathcal{N}}))}\right)^4\ . \tag{1.42}$$

Recent results reported in [14, 18, 46, 48] provide stability results for the DPI, but the lower bounds provided there involve quantities that are not so directly related to $\sigma - \mathscr{R}_\rho(\sigma_{\mathcal{N}})$ or to $\rho - \mathscr{R}_\sigma(\rho_{\mathcal{N}})$); e.g., "rotated" and "twirled" Petz recovery maps. For another fidelity type bound not explicitly involving the recover map, see [8, Theorem 2.2]. The appeal of the bounds proved in this paper is that they are expressed in terms of physically relevant distances between σ and $\mathscr{R}_\rho(\sigma_{\mathcal{N}})$, or between ρ and $\mathscr{R}_\sigma(\rho_{\mathcal{N}})$).

2 Stability for the Data Processing Inequality

We begin this section by recalling Petz's proof of the monotonicity of the quasi relative entropies S_f for operator convex f. Throughout this section, $\mathcal{N}$ is a von Neumann subalgebra of the finite dimensional von Neumann algebra $\mathcal{M}$, and ρ and σ are two density matrices in $\mathcal{M}$. $\mathscr{E}_\tau$ is the tracial conditional expectation onto $\mathcal{N}$, and $\rho_{\mathcal{N}} = \mathscr{E}_\tau\rho$ and $\sigma_{\mathcal{N}} = \mathscr{E}_\tau\sigma$. Finally $\mathcal{H}$ denotes $(\mathcal{M}, \langle\cdot,\cdot\rangle_{HS})$, where $\langle\cdot,\cdot\rangle_{HS}$ is the Hilbert-Schmidt inner product.

Define the operator U mapping $\mathcal{H}$ to $\mathcal{H}$ by

$$U(X) = \mathscr{E}_\tau(X)\rho_{\mathcal{N}}^{-1/2}\rho^{1/2}\ . \tag{2.1}$$

Note that for all $X \in \mathcal{N}$, $U(X) = X\rho_{\mathcal{N}}^{-1/2}\rho^{1/2}$. The adjoint operator on $\mathcal{H}$ is given by

$$U^*(Y) = \mathscr{E}_\tau(Y\rho^{1/2})\rho_{\mathcal{N}}^{-1/2} \tag{2.2}$$

for all $Y \in \mathcal{H} = \mathcal{M}$.

For $X \in \mathcal{M}$, $U^*U(X) = \mathscr{E}_\tau(\rho_{\mathcal{N}}^{-1/2}\mathscr{E}_\tau(X)\rho_{\mathcal{N}}^{-1/2}\rho) = \mathscr{E}_\tau(X)$. Hence $U^*U = \mathscr{E}_\tau$, the orthogonal projection in $\mathcal{H}$ onto $\mathcal{N}$. That is, U, restricted to $\mathcal{N}$, is an isometric embedding of $\mathcal{N}$ into $\mathcal{H} = \mathcal{M}$, but it is not the trivial isometric embedding by inclusion. Also, we see that on $\mathcal{N}$ the map U is isometric.

Now observe that for all $X \in \mathcal{N}$, $\Delta_{\sigma,\rho}^{1/2}(U(X)) = \sigma^{1/2}X\rho_{\mathcal{N}}^{-1/2}$, and hence for all $X \in \mathcal{N}$,

$$\begin{aligned}
\langle\Delta_{\sigma,\rho}^{1/2}(U(X)), \Delta_{\sigma,\rho}^{1/2}(U(X))\rangle &= \mathrm{Tr}((\rho_{\mathcal{N}})^{-1/2}X^*\sigma X(\rho_{\mathcal{N}})^{-1/2}) \\
&= \mathrm{Tr}((\rho_{\mathcal{N}})^{-1/2}X^*\sigma_{\mathcal{N}}X(\rho_{\mathcal{N}})^{-1/2}) \\
&= \langle\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}}^{1/2}(X), \Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}}^{1/2}(X)\rangle\ .
\end{aligned}$$

That is, on $\mathcal{N}$,

$$U^*\Delta_{\sigma,\rho}U = \Delta_{\sigma_\mathcal{N},\rho_\mathcal{N}} \ . \tag{2.3}$$

By the operator Jensen inequality, as operators on $(\mathcal{N}, \langle\cdot,\cdot\rangle_{HS})$,

$$U^* f(\Delta_{\sigma,\rho})U \geq f\left(U^*\Delta_{\sigma,\rho}U\right) \ . \tag{2.4}$$

Combining (2.3) and (2.4), and using the fact that $U(\rho_\mathcal{N})^{1/2} = \rho^{1/2}$,

$$\begin{aligned} S_f(\rho_\mathcal{N}||\sigma_\mathcal{N}) &= \langle(\rho_\mathcal{N})^{1/2}, f(\Delta_{\sigma_\mathcal{N},\rho_\mathcal{N}})(\rho_\mathcal{N})^{1/2}\rangle \\ &\leq \left\langle U(\rho_\mathcal{N})^{1/2}, f(\Delta_{\sigma,\rho})U(\rho_\mathcal{N})^{1/2}\right\rangle \\ &= \left\langle \rho^{1/2}, f(\Delta_{\sigma,\rho})\rho^{1/2}\right\rangle = S_f(\rho||\sigma) \ . \end{aligned}$$

This proves, following Petz, his monotonicity theorem for the quasi relative entropy S_f for the operator convex function.

Now consider the family of quasi relative entropies defined by functions $f_t(x) = (t+x)^{-1}$. Our immediate goal is to prove the inequality

$$S_{(t)}(\rho||\sigma) = \langle \rho^{1/2}, (t+\Delta_{\sigma,\rho})^{-1}\rho^{1/2}\rangle \geq \langle \rho_\mathcal{N}^{1/2}, (t+\Delta_{\sigma_\mathcal{N},\rho_\mathcal{N}})^{-1}\rho_\mathcal{N}^{1/2}\rangle = S_{(t)}(\rho_\mathcal{N}||\sigma_\mathcal{N}) \ . \tag{2.5}$$

2.1 Lemma. *Let U be a partial isometry embedding a Hilbert space $\mathcal{K}$ into a Hilbert space $\mathcal{H}$. Let B be an invertible positive operator on $\mathcal{K}$, A be an invertible positive operator on $\mathcal{H}$, and suppose that $U^*AU = B$. Then for all $v \in \mathcal{K}$,*

$$\langle v, U^*A^{-1}Uv\rangle = \langle v, B^{-1}v\rangle + \langle w, Aw\rangle \ , \tag{2.6}$$

where

$$w := UB^{-1}v - A^{-1}Uv \ . \tag{2.7}$$

Proof. We compute, using $U^*U = \mathbf{1}_\mathcal{K}$,

$$\begin{aligned} \langle w, Aw\rangle &= \langle UB^{-1}v - A^{-1}Uv, AUB^{-1}v - Uv)\rangle \\ &= \langle v, B^{-1}U^*AUB^{-1}v\rangle - 2\langle v, B^{-1}v\rangle + \langle v, U^*A^{-1}Uv\rangle \\ &= -\langle v, B^{-1}v\rangle + \langle v, U^*A^{-1}Uv\rangle \end{aligned}$$

□

Proof of Theorem 1.5. We apply Lemma 2.1 with $A := (t+\Delta_{\sigma,\rho})$, $B = (t+\Delta_{\sigma_\mathcal{N},\rho_\mathcal{N}})$ and $v := (\rho_\mathcal{N})^{1/2}$, and with U defined as above. The lemma's condition, $U^*AU = B$, follows from (2.3) and the fact that $U^*U = \mathbf{1}_\mathcal{K}$. Therefore, applying Lemma 2.1 with $U(\rho_\mathcal{N})^{1/2} = \rho^{1/2}$,

$$\begin{aligned} S_{(t)}(\rho||\sigma) - S_{(t)}(\rho_\mathcal{N}||\sigma_\mathcal{N}) &= \langle \rho^{1/2}, (t+\Delta_{\sigma,\rho})^{-1}\rho^{1/2}\rangle - \langle \rho_\mathcal{N}^{1/2}, (t+\Delta_{\sigma_\mathcal{N},\rho_\mathcal{N}})^{-1}\rho_\mathcal{N}^{1/2}\rangle \\ &= \langle w_t, (t+\Delta_{\sigma,\rho})w_t\rangle \geq t\|w_t\|^2, \end{aligned} \tag{2.8}$$

where, recalling that $U(\rho_{\mathcal{N}})^{1/2} = \rho^{1/2}$,

$$w_t := U(t + \Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}})^{-1}(\rho_{\mathcal{N}})^{1/2} - (t + \Delta_{\sigma,\rho})^{-1}\rho^{1/2} \ . \tag{2.9}$$

Using the integral representation of the square root function,

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

and $U(\rho_{\mathcal{N}})^{1/2} = \rho^{1/2}$ once more, we conclude that

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

On the other hand,

$$\begin{aligned} U(\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}})^{1/2}(\rho_{\mathcal{N}})^{1/2} - (\Delta_{\sigma,\rho})^{1/2}\rho^{1/2} &= U(\sigma_{\mathcal{N}})^{1/2} - \sigma^{1/2} \\ &= (\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2} \ . \end{aligned}$$

Therefore, combining the last two equalities and taking the Hilbert space norm associated with $\mathcal{H}$, for any $T > 0$,

$$\begin{aligned} \|(\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 &= \frac{1}{\pi}\left\|\int_0^\infty t^{1/2} w_t \mathrm{d}t\right\|_2 \\ &\leq \frac{1}{\pi}\int_0^T t^{1/2}\|w_t\|_2 \mathrm{d}t + \frac{1}{\pi}\left\|\int_T^\infty t^{1/2} w_t \mathrm{d}t\right\|_2 \ . \end{aligned} \tag{2.10}$$

We estimate these two terms separately. For the first term, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \left(\int_0^T t^{1/2}\|w_t\|_2 \mathrm{d}t\right)^2 &\leq T\int_0^T t\|w_t\|_2^2 \mathrm{d}t \\ &\leq T\int_0^\infty \big(S_{(t)}(\rho||\sigma) - S_{(t)}(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}})\big)\,\mathrm{d}t \\ &= T(S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}})) \ . \end{aligned} \tag{2.11}$$

For the second term in (2.10), note that for any positive operator X

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

and hence

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

The spectra of $\sigma_{\mathcal{N}}$ and $\rho_{\mathcal{N}}$ lie in the convex hulls of the spectra of σ and ρ respectively. It follows that $\|\Delta_{\sigma_{\mathcal{N}},\rho_{\mathcal{N}}}\| \leq \|\Delta_{\sigma,\rho}\|$. Therefore, recalling the definition of w_t in (2.9), we obtain

$$\left\|\int_T^\infty t^{1/2} w_t \mathrm{d}t\right\|_2 \leq \frac{4\|\Delta_{\sigma,\rho}\|}{T^{1/2}} \ . \tag{2.12}$$

Combining (2.10), (2.11) and (2.12) we obtain

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

Optimizing in T,

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

Rearranging terms

$$S(\rho||\sigma) - S(\rho_{\mathcal{N}}||\sigma_{\mathcal{N}}) \geq \left(\frac{\pi}{4}\right)^4 \|\Delta_{\sigma,\rho}\|^{-2}\|(\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2^4 \ . \tag{2.13}$$

□

We now prove the lemma leading from (1.30) to (1.31).

2.2 Lemma. *For any operators X and Y with* $\mathrm{Tr}[X^*X] = \mathrm{Tr}[Y^*Y] = 1$. *Then*

$$\|X^*X - Y^*Y\|_1 \leq 2\|X - Y\|_2 \ . \tag{2.14}$$

Proof. Recall that for any operator A, $\|A\|_1 = \sup\{|\mathrm{Tr}[ZA] \ : \ \|Z\| \leq 1\}$ where $\|\cdot\|$ denotes the operator norm. For any contraction Z, using cyclicity of the trace we have

$$\begin{aligned} |\mathrm{Tr}[Z(X^*X - Y^*Y)]| &\leq |\mathrm{Tr}[Z(X^* - Y^*)X + ZY^*(X - Y)]| \\ &\leq |\mathrm{Tr}[(X^* - Y^*)XZ| + |\mathrm{Tr}[ZY^*(X - Y)] \\ &\leq (\mathrm{Tr}(X^* - Y^*)(X - Y)])^{1/2}(\mathrm{Tr}[X^*Z^*ZX])^{1/2} \\ &+ (\mathrm{Tr}(X^* - Y^*)(X - Y)])^{1/2}(\mathrm{Tr}[Y^*Z^*ZY])^{1/2} \\ &\leq 2\|X - Y\|_2 \ . \end{aligned}$$

□

Applying this with $X = (\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2}$ and $Y = \sigma^{1/2}$, we get

$$\|(\sigma_{\mathcal{N}})^{1/2}(\rho_{\mathcal{N}})^{-1/2}\rho^{1/2} - \sigma^{1/2}\|_2 \geq \frac{1}{2}\|\mathscr{R}_\rho(\sigma_{\mathcal{N}}) - \sigma\|_1 \ .$$

3 Conditional expectations and the Petz Recovery Map

Recall from the introduction that classically, for *any* integrable random variable X on *any* probability space, if $\mathscr{E}(X)$ is its conditional expectation with respect to any other random variable, then both X and $\mathscr{E}(X)$ have the *same* expected value. Conditional expectations in the quantum setting need not have this property; in fact, they usually do not, as is shown by a theorem of Takesaki [37]. Also as explained in the introduction, taking a classical conditional expectation is essentially a "coarse graining" operation. Coarse graining is important in quantum statistical mechanics, and one wants observables and their coarse grained versions to have the same expected values; see Accardi and Cechinni [2], who took Takesaki's Theorem as their starting point in the work leading to their coarse graining operator, which, as we have explained, has the Petz recovery map as its dual.

In short, the quantum obstacle to the existence of expectation preserving conditional expectations is the *raison d'ete* for the Petz recovery map. Takesaki's deep paper makes use of the subtle Tomita-Takesaki Theory which limits its accessibility. It is therefore desirable to give an elementary elucidation of this quantum obstacle in our present finite dimensional setting. Wolf [47] gives a clear account of a number of aspects of quantum conditional expectations in the finite dimensional setting. He takes advantage of the fact that any finite dimensional von Neumann algebra is isometrically isomorphic to the direct sum of a finite number of complete matrix algebras. He then proves what may be viewed as a finite dimensional version of Tomiyama's Theorem [38] characterizing conditional expectations as norm one projections. He does not discuss Takesaki's Theorem, and we are unaware of any elementary exposition of it. Perhaps it would be possible to give an elementary proof of it by using the structure theory for finite dimensional von Neumann algebras, as in [47]. Here we take an approach that is elementary and yields a result that is not implied by Takesaki's Theorem, and uses methods that are not strictly limited to the finite dimensional case.

Our starting point is a fact explained in the Introduction: If ρ is a faithful state on $\mathcal{M}$, and $\mathcal{N}$ is a von Neumann subalgebra of $\mathcal{M}$, then there exists a conditional expectation $\mathscr{E}$ from $\mathcal{M}$ to $\mathcal{N}$ such that for all $X \in \mathcal{M}$, $\rho(X) = \rho(\mathscr{E}(X))$ if and only if the orthogonal projection onto $\mathcal{N}$ in the GNS inner product induced by ρ is a conditional expectation.

Thus we can rephrase our question about the obstacle to the existence of expectation preserving conditional expectations as: For which faithful states ρ is the orthogonal projection onto $\mathcal{N}$ in the GNS inner product induced by ρ, $\mathscr{P}_\rho$, is actually a conditional expectation? Conditional expectations, as defined in the Introduction, are always *real*; i.e., they send Hermitan operators to Hermitian operators. Therefore, the first part of the next theorem says that $\mathscr{P}_\rho$ can be a conditional expectation only if $\mathcal{N}$ is invariant under Δ_ρ.

3.1 Theorem. *Let $\mathcal{M}$ be a finite dimensional von Neumann algebra, and let $\mathcal{N}$ be a von Neumann subalgebra of $\mathcal{M}$. Let ρ be a faithful state on $\mathcal{M}$, and let Δ_ρ be the* modular operator *on $\mathcal{M}$ defined by $\Delta_\rho(X) = \rho X \rho^{-1}$. Let $\mathscr{P}_\rho$ be the orthogonal projection from $\mathcal{M}$ onto $\mathcal{N}$ in the GNS inner product induced by ρ. Then:*

(1) $\mathscr{P}_\rho$ is real; i.e., it preserves self-adjointness, if and only if $\mathcal{N}$ is invariant under Δ_ρ.

(2) $\mathcal{N}$ is invariant under Δ_ρ if and only if for all $A \in \mathcal{N}$,

$$\Delta_\rho(A) = \Delta_{\rho_\mathcal{N}}(A) \ , \tag{3.1}$$

in which case $\Delta^t_\rho(A) = \Delta^t_{\rho_\mathcal{N}}(A)$ for all $t \in \mathbb{R}$. Furthermore, (3.1) *is valid for all $A \in \mathcal{N}$ if and only if $\mathscr{A}_\rho(A) = A$ for all $A \in \mathcal{N}$.*

3.2 Remark. Part *(2)* of Theorem 3.1 is due to Accardi and Cecchini [2, Theorem 5.1]. In our finite dimensional context, we give a very simple proof; most of the proof below is devoted to *(1).*

Proof of Theorem 3.1. Suppose that $\mathscr{P}_\rho$ is real. Then for all $X \in \text{Null}(\mathscr{P}_\rho)$, $0 = (\mathscr{P}_\rho(X))^* = \mathscr{P}_\rho(X^*)$, so that $\text{Null}(\mathscr{P}_\rho)$ is a self adjoint subspace of $\mathcal{M}$. Let m denote the dimension of $\text{Null}(\mathscr{P}_\rho)$. Then, applying the Gram-Schmidt Algorithm, one can produce an orthonormal basis $\{H_1,\dots,H_m\}$ of $\text{Null}(\mathscr{P}_\rho)$ consisting of self-adjoint elements of $\mathcal{M}$.

The map $X \mapsto X\rho^{1/2}$ is unitary from $(\mathcal{M}, \langle\cdot,\cdot\rangle_{GNS,\rho})$ to $(\mathcal{M}, \langle\cdot,\cdot\rangle_{HS})$. Therefore for all $A \in \mathcal{N}$, and each $j = 1,\dots,m$, $\langle A\rho^{1/2}, H_j\rho^{1/2}\rangle_{HS} = 0$. Then since the map $X \mapsto X^*$ is an (antilinear) isometry on $(\mathcal{M}, \langle\cdot,\cdot\rangle_{HS,\rho})$,

$$0 = \langle (H_j\rho^{1/2})^*, (A\rho^{1/2})^*, \rangle_{HS} = \text{Tr}[H_j\rho A^*] = \text{Tr}[H_j\Delta_\rho(A^*)\rho] = \langle H_j, \Delta_\rho(A^*), \rangle_{GNS,\rho} \ .$$

Therefore, $\Delta_\rho(A^*)$ is orthogonal to $\text{Null}(\mathscr{P}_\rho)$ in $(\mathcal{M}, \langle\cdot,\cdot\rangle_{GNS,\rho})$, and hence $\Delta_\rho(A^*) \in \mathcal{N}$. Since A is arbitrary in $\mathcal{N}$, it follows that $\mathcal{N}$ is invariant under Δ_ρ.

For the converse, suppose that $\mathcal{N}$ is invariant under Δ_ρ. Then $\mathcal{N}$ is invariant under Δ_ρ^s for all $s \in \mathbb{R}$, and in particular, $\mathcal{N}$ is invariant under $\Delta_\rho^{1/2}$. Consequently, $\rho^{1/2}\mathcal{N} = \mathcal{N}\rho^{1/2}$ as subspaces of $\mathcal{M}$; let $\mathcal{K}$ denote this subspace, which is evidently self-adjoint. Let $\mathcal{K}^\perp$ demote its Hilbert-Schmidt orthogonal complement in $\mathcal{M}$. Let $H = H^* \in \mathcal{M}$. Then there is a unique $A \in \mathcal{N}$ such that $H\rho^{1/2} - A\rho^{1/2} \in \mathcal{K}^\perp$. Thus,

$$H\rho^{1/2} = (H\rho^{1/2} - A\rho^{1/2}) + A\rho^{1/2} \tag{3.2}$$

is the orthogonal decomposition of $H\rho^{1/2}$ with respect to $\mathcal{K}$. Again since $X \mapsto X\rho^{1/2}$ is unitary from $(\mathcal{M}, \langle\cdot,\cdot\rangle_{GNS,\rho})$ to $(\mathcal{M}, \langle\cdot,\cdot\rangle_{HS})$, $\mathscr{P}_\rho(H) = A$. We must show that $A = A^*$.

Since $X \mapsto X^*$ is an isometry for the Hilbert-Schmidt inner product, and since $\mathcal{K}$ and $\mathcal{K}^\perp$ are self adjoint,

$$\rho^{1/2}H = (\rho^{1/2}H - \rho^{1/2}A^*) + \rho^{1/2}A^* \tag{3.3}$$

is the orthogonal decomposition of $\rho^{1/2}H$ with respect to $\mathcal{K}$. In particular, $(\rho^{1/2}H - \rho^{1/2}A^*) \in \mathcal{K}^\perp$. Hence for any $Z \in \mathcal{N}$,

$$0 = \langle Z\rho^{1/2}, (\rho^{1/2}H - \rho^{1/2}A^*)\rangle_{HS} = \text{Tr}[Z^*(\rho^{1/2}H - \rho^{1/2}A^*)\rho^{1/2}] = \langle \rho^{1/2}Z, H\rho^{1/2} - A^*\rho^{1/2}\rangle_{HS} \ ,$$

and hence $H\rho^{1/2} - A^*\rho^{1/2} \in \mathcal{K}^\perp$. Now apply $\Delta_\rho^{-1/2}$ to both sides of (3.3) to obtain

$$H\rho^{1/2} = (H\rho^{1/2} - A^*\rho^{1/2}) + A^*\rho^{1/2} \ .$$

By what we have just shown, this is the orthogonal decomposition of $H\rho^{1/2}$, and must coincide with (3.2). Hence $A^* = A$. This proves *(1)*.

To prove *(2)*, note first of all that when (3.1) is valid for all $A \in \mathcal{N}$, then Δ_ρ preserves $\mathcal{N}$ since the right side evidently belongs to $\mathcal{N}$.

Now suppose the Δ_ρ preserves $\mathcal{N}$. Let $A, B \in \mathcal{N}$. Then $A^*\Delta_{\rho_\mathcal{N}}(\Delta_\rho(B)) \in \mathcal{N}$, and then by the definition of $\mathscr{E}_\tau$ and cyclicity of the trace,

$$\text{Tr}[\rho(A^*\rho_\mathcal{N}\rho^{-1}B\rho\rho_\mathcal{N}^{-1})] = \text{Tr}[\rho_\mathcal{N}(A^*\rho_\mathcal{N}\rho^{-1}B\rho\rho_\mathcal{N}^{-1})] = \text{Tr}[A^*\rho_\mathcal{N}\rho^{-1}B\rho] = \text{Tr}[\rho_\mathcal{N}(\rho^{-1}B\rho A^*)] \ .$$

In the same way, using the fact that $(\rho^{-1}B\rho A) \in \mathcal{N}$ and cyclicity of the trace,

$$\mathrm{Tr}[\rho_{\mathcal{N}}(\rho^{-1}B\rho A^*)] = \mathrm{Tr}[B\rho A^*] = \mathrm{Tr}[\rho A^* B] \ .$$

Altogether, $\langle A, \Delta_{\rho_N}(\Delta_\rho^{-1}(B))\rangle_{GNS,\rho} = \langle A, B\rangle_{GNS,\rho}$. Since $\Delta_{\rho_N}(\Delta_\rho^{-1}(B)) \in \mathcal{N}$, and A is arbitrary in $\mathcal{N}$, $\Delta_{\rho_N}(\Delta_\rho^{-1}(B)) = B$, and hence $\Delta_\rho^{-1}(B) = \Delta_{\rho_N}^{-1}(B)$. Then $\Delta_\rho^{-n}(B) = \Delta_{\rho_N}^{-n}(B)$ for all $n \in \mathbb{N}$, and then it follows that $\Delta_\rho^t(B) = \Delta_{\rho_N}^t(B)$ for all $t \in \mathbb{R}$.

Finally, we show that (3.1) is valid for all $A \in \mathcal{N}$, then $\mathscr{A}_\rho(A) = A$ for all $A \in \mathcal{N}$:

$$\mathscr{E}_\tau(\rho^{1/2}A\rho^{1/2}) = \mathscr{E}_\tau(\Delta_\rho^{1/2}(A)\rho) = \Delta_{\rho_{\mathcal{N}}}^{1/2}(A)\mathscr{E}_\tau(\rho) = \rho_{\mathcal{N}}^{1/2}A\rho_{\mathcal{N}}^{1/2} \ .$$

Therefore,

$$\mathscr{A}_\rho(A) = \rho_{\mathcal{N}}^{-1/2}\mathscr{E}_\tau(\rho^{1/2}A\rho^{1/2})\rho_{\mathcal{N}}^{-1/2} = A \ .$$

On the other hand, when $A = \mathscr{A}_\rho(A)$ for all $A \in \mathcal{N}$, $\mathscr{A}_\rho$ is a norm one projection onto $\mathcal{N}$, and by Tomiyama's Theorem [38], it is a conditional expectation, and it satisfies $\rho(\mathscr{A}_\rho(X)) = \rho(X)$ for all $X \in \mathcal{M}$. Therefore, it must coincide with $\mathscr{P}_\rho$, the orthogonal projection form $\mathcal{M}$ onto $\mathcal{N}$ in the GNS inner product induced by ρ. Hence $\mathscr{P}_\rho$ is a conditional expectation. By what we proved earlier, this means that $\mathcal{N}$ is invariant under Δ_ρ, and then that (3.1) is valid for all $A \in \mathcal{N}$. □

3.3 Theorem. *Let $\mathscr{P}_\rho$ denote the orthogonal projection of $\mathcal{M}$ onto $\mathcal{N}$ in the GNS inner product induced by ρ. Then*

(1) $\mathscr{P}_\rho$ is a conditional expectation if and only if $\mathcal{N}$ is invariant under Δ_ρ.

(2) $\mathscr{P}_\rho$ is a conditional expectation if and only if $\mathscr{P}_\rho$ is real.

Proof. Theorem 3.1 says that when Δ_ρ does not leave $\mathcal{N}$ invariant, $\mathscr{P}_\rho$ is not even real, and hence is not a conditional expectation. On the other hand, when Δ_ρ leaves $\mathcal{N}$ invariant, by part *(2)* of Theorem 3.1, $\mathscr{A}_\rho$ is a norm-one projection onto $\mathcal{N}$. By Tomiyama's Theorem, $\mathscr{A}_\rho$ is a conditional expectation that preserves expectation with respect to ρ. By remarks made in the introduction, this means that $\mathscr{A}_\rho = \mathscr{P}_\rho$, and hence that $\mathscr{P}_\rho$ is a conditional expectation. This proves *(1)*.

It is evident that if $\mathscr{P}_\rho$ is a conditional expectation, this $\mathscr{P}_\rho$ is real. On the other hand, if $\mathscr{P}_\rho$ is real, then by Theorem 3.1, $\mathcal{N}$ is invariant under Δ_ρ, and now *(2)* follows from *(1)*. □

Acknowledgments. The authors are grateful to Mark Wilde and Lin Zhang for comments and questions that have led us to add some reformulations in this version. EAC was partially supported by NSF grants DMS 1501007 and DMS 1764254. AV is grateful to EAC for hosting her visits to Rutgers University, during which this work was partially completed. AV is partially supported by NSF grant DMS 1812734.

References

[1] L. Accardi, *Non commutative Markov chains*, Proc. School of Math. Phys. Camerino (1974).

[2] L. Accardi and C. Cecchini, *Conditional Expectations in von Neumann algebras and A Theorem of Takesaki* Jor. Func. Analysis **45**, 245 - 273 (1982)

[3] H. Araki, *Relative entropy of state of von Neumann algebras,* Publ. RIMS Kyoto Univ. **9**, 809 - 833 (1976)

[4] Bhatia, *Matrix analysis,* Springer-Verlag, New York, 1997

[5] M.-D. Choi, *A Schwarz inequality for positive linear maps on C^*-algebras,* Illinois J. Math **18**: 4, 565-574 (1974)

[6] E. A. Carlen and E. H. Lieb, *Optimal Hypercontractivity for Fermi Fields and Related Non-commutative Integration Inequalities*, Comm. Math. Phys., **155**, 1993 pp. 27-46.

[7] E. A. Carlen and E. H. Lieb, *Bounds for entanglement via an extension of strong subadditivity of entropy.* Lett. Math. Phys. **101** (2012), no. 1, 1–11

[8] E. A. Carlen and E. H. Lieb, *Remainder Terms for Some Quantum Entropy Inequalities*, Jour. Math. Phys., **55**, 042201 (2014)

[9] E. A. Carlen and A. Vershynina, *Recovery and the Data Processing Inequality for Quasi-Entropies*, IEEE Trans. Info. Thy., **64**, 6929 - 6938 (2018)

[10] M.-D. Choi, *Completely positive linear maps on complex matrices,* Linear Algebra and its Applications, **10**: 3, 285-290 (1975)

[11] E. B. Davies, *Markovian master equations,* Commun. Math. Phys., **39**, 91- 110 (1974)

[12] C. Davis, *Various averaging operations onto subalgebras*, Illinois J. Math. **3** 538-553 (1959)

[13] H. Fawzi and O. Fawzi, *Efficient optimization of the quantum relative entropy*, J. Phys. A: Math. Theor., **51**, 154003 (2018)

[14] O. Fawzi, R. Renner, *Quantum conditional mutual information and approximate Markov chains*, Commun. Math. Phys. 340(2), 2015

[15] P. Hayden, R. Jozsa, D. Petz, A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality,* Communications in mathematical physics, **246**:2, 359-374 (2004)

[16] R. V. Kadison, *A generalized Schwarz inequality and algebraic invariants for operator algebra*, Ann. of Mafh. **56** 494 - 503 (1952)

[17] E. H. Lieb, *Convex trace functions and the Wigner-Yanase-Dyson conjecture*, Advances in Math., **11**, 267-288 (1973)

[18] M. Junge, R. Renner, D. Sutter, M. Wilde, A. Winter, *Universal recovery from a decrease of quantum relative entropy* arXiv:1509.07127, 2015

[19] E. H. Lieb, M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy,* Journal of Mathematical Physics **14:12**, 1938-1941 (1973)

[20] G. Lindblad, *Expectations and Entropy Inequalities for Finite Quantum Systems*, Commun. math. Phys. **39**, 111-119 (1974)

[21] G. Lindblad, *Completely Positive Maps and Entropy Inequalities,* Commun. Math. Phys. **40**, 147-151 (1975)

[22] G. Lindblad *The general no-cloning theorem*, Lett. Math. Phys., **47**:2, 189-196 (1999)

[23] M. Koashi, N. Imoto, *Operations that do not disturb partially known quantum states,* Phys. Rev. A, **66**: 2, 022318 (2002)

[24] M. Mosonyi and D. Petz, *Structure of Sufficient Quantum Coarse-Grainings*, Lett. in Math. Phys., **66**, 19-30, (2004)

[25] M. Nielsen, D. Petz, *A simple proof of the strong subadditivity inequality,* Quantum Information & Computation, **6**, 507 - 513 (2005)

[26] D. Petz, *Quasi-entropies for states of a von Neumann algebra,* Publ. RIMS. Kyoto Univ. **21**, 781-800 (1985)

[27] D. Petz, *Quasi-entropies for finite quantum systems,* Rep. Math. Phys. **23**, 57 - 65 (1986)

[28] D. Petz, *Sufficient subalgebras and the relative entropy of states of a von Neumann algebra,* Comm. Math. Phys. **105**:1, 123-131 (1986).

[29] D. Petz, *Sufficiency of channels over von Neumann algebras,* Quart. J. Math. Oxford Ser. (2), **39**:153, 97-108 (1988)

[30] D. Petz, D. Virosztek. *Some inequalities for quantum Tsallis entropy related to the strong subadditivity,* arXiv:1403.7062 (2014)

[31] D. W. Robinson, D. Ruelle, *Mean Entropy of States in Classical Statistical Mechanis,* Communications in Mathematical Physics **5**, 288 (1967)

[32] M. B. Ruskai, *Inequalities for Quantum Entropy: A Review with Conditions for Equality,* J. Math. Phys. **43**, 4358-4375 (2002); erratum **46**, 019901 (2005)

[33] N. Sharma, *Equality Conditions for Quantum Quasi-Entropies Under Monotonicity and Joint-Convexity,* Nat. Conf. Commun., 2014

[34] W. F. Stinespring, *Positive Functions on C*-algebras*, Proc. American Math. Soc., **6**, 211-216 (1955)

[35] D. Sutter, M. Berta and M. Tomamichel, *Multivariate Trace Inequlaities*, Comm. Math. Phys., **352**, 37-58 (2017)

[36] D. Sutter, *Approximate quantum Markov chains*, Springer Briefs in Mathematical Physics, Springer, Nerlin, 2018.

[37] M. Takesaki, *Conditional Expectations in von Neumann Algebras*, J. Funct. Anal. **9**, 306 - 321. (1972)

[38] J. Tomiyama, *On the projection of norm. one in W*-algebras*, Proc. Jupnn Acad. **33** 608 - 612 (1957)

[39] A. Uhlmann, *Endlich Dimensionale Dichtematrizen, II,* Wiss. Z. Karl-MarxUniversity Leipzig, Math-Naturwiss. **22**:2, 139 (1973)

[40] A. Uhlmann, *The "transition probability" in the state space of a *-algebra,* Reports on Mathematical Physics, **9**(2):273?279, (1976)

[41] A. Uhlmann. *Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory,* Communications in Mathematical Physics **54**: 1, 21-32 (1977)

[42] H. Umegaki, *Conditional expectation in an operator algebra*, Tokohu Math. J. **6** 177-181. (1954).

[43] H. Umegaki, *Conditional expectation in an operator algebra, II*, Tokohu Math. J. **8**, 86-100 (1956),

[44] H. Umegaki, *Conditional expectation in an operator algebra, III*, Kodai Marh. Sem. Rep. **11** 51-64 (1959) 39. H. UMEGAKI, Conditional

[45] H. Umegaki, *Conditional Expectation in an Operator Algebra. IV. Entropy and Information,* Kodai Math. Sem. Rep. **14**, 59-85 (1962)

[46] M. Wilde, *Recoverability in quantum information theory*, Proc. R. Soc. A. **471**(2182) The Royal Society, (2015)

[47] M. Wolf, *Quantum Channels and Operations: Guided Tour*, T.U. Munich Prerprint, July 5, 2012,

[48] L. Zhang, *A Strengthened Monotonicity Inequality of Quantum Relative Entropy: A Unifying Approach Via Rényi Relative Entropy,* Letters in Mathematical Physics 106(4): 557-573, (2016)