



HAL
open science

On the Monotonicity of relative entropy: A Comparative Study of Petz's and Uhlmann's Approaches

Santiago Matheus, Francesco Bottacin, Edoardo Provenzi

► To cite this version:

Santiago Matheus, Francesco Bottacin, Edoardo Provenzi. On the Monotonicity of relative entropy: A Comparative Study of Petz's and Uhlmann's Approaches. *Entropy*, 2025, 27 (9), pp.954. <10.3390/e27090954>. <hal-05215210v2>

HAL Id: hal-05215210

<https://hal.science/hal-05215210v2>

Submitted on 14 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

On the Monotonicity of relative entropy: A Comparative Study of Petz’s and Uhlmann’s Approaches

Santiago Matheus^{*1}, Francesco Bottacin^{†1} and Edoardo Provenzi^{‡2}

¹Università degli studi di Padova, Dipartimento di matematica, Via Trieste, 63, 35121, Padova, Italy

²Université de Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, 351 Cours de la Libération, Talence, France

Abstract

We revisit the monotonicity of relative entropy under the action of quantum channels, a foundational result in quantum information theory. Among the several available proofs, we focus on those by Petz and Uhlmann, which we reformulate within a unified, finite-dimensional operator-theoretic framework. In the first part, we examine Petz’s strategy, identify a subtle flaw in his original use of Jensen’s contractive operator inequality, and point out how it was corrected to restore the validity of his line of reasoning. In the second part, we develop Uhlmann’s approach, which is based on interpolations of positive sesquilinear forms and applies automatically also to non-invertible density operators. By comparing these two approaches, we highlight their complementary strengths: Petz’s method is more direct and clear, Uhlmann’s is more abstract and general. Our treatment aims to clarify the mathematical structure underlying the monotonicity of relative entropy and to make these proofs more accessible to a broader audience interested in both the foundations and the applications of quantum information theory.

1 Introduction

The (quantum) relative entropy is a central concept in quantum information theory, quantifying the distinguishability between quantum states and serving as a key tool in the analysis of information-processing tasks. One of its most important properties is the *monotonicity under quantum channels*¹, which expresses the idea that state distinguishability cannot increase during the dynamical evolution of a quantum system that interacts with an environment. This property is also known as Data Processing Inequality (DPI).

The concept of relative entropy was first introduced by Umegaki in 1962 [26], in the setting of σ -finite von Neumann algebras, and later extended to arbitrary von Neumann algebras by Araki in 1976 [1] by means of the Tomita–Takesaki modular theory.

The proof of monotonicity of relative entropy is closely related to the proof of the strong subadditivity (SSA) of the von Neumann entropy. The latter serves as a measure of the mixedness of quantum states, and the validity of the SSA ensures that quantum uncertainty behaves consistently across composite systems, placing fundamental constraints on how information and correlations can be distributed among subsystems.

The first step toward proving SSA was made in 1968 by Lanford and Robinson [10], who established the subadditivity of the von Neumann entropy and conjectured its stronger form.

^{*}santiago.javier.matheusrocha@studenti.unipd.it

[†]bottacin@math.unipd.it

[‡]edoardo.provenzi@math.u-bordeaux.fr

¹A quantum channel is a completely positive trace-preserving linear map, see later for a rigorous definition.

In 1973, Lieb [11], building on earlier work by Wigner, Yanase, and Dyson, proved several key properties concerning the convexity and concavity of operator functions and trace functionals.

These results enabled Lieb and Ruskai to establish the full proof of strong subadditivity of the von Neumann entropy for both finite and infinite-dimensional Hilbert spaces later that same year, in their landmark paper [12]. In that work, they also derived, though without emphasizing it as such, the monotonicity of relative entropy under the partial trace operation, which constitutes a special case of quantum channel.

The first explicit and general proof of the monotonicity of relative entropy under the action of quantum channels was provided two years later by Lindblad in his seminal 1975 paper [13], thanks to the results established by Lieb and Ruskai in [12].

A further breakthrough was achieved by Uhlmann in 1977, who extended the property of monotonicity to a broader class of transformations: the adjoints of unital Schwarz maps, see [25].

The equivalence between SSA and the monotonicity of relative entropy was rigorously established by Petz in the 1980s [19, 20]. Later, in the work [18], Petz proposed a new proof of monotonicity and posed the question of whether this property also holds for positive (but not necessarily completely positive) trace-preserving linear maps. This question remained open for several years, until it was affirmatively resolved in 2023 by Müller-Hermes and Reeb in [15]. Other pertinent references related to the monotonicity of relative entropy are [22, 4, 29, 8].

In this paper, we focus specifically on the demonstration strategies proposed by Petz and Uhlmann in the papers [18] and [25], respectively, which we believe offer particularly interesting and useful complementary perspectives. Petz’s proof relies on operator-algebraic methods and an explicit representation of the relative entropy in terms of a suitable inner product, an idea inspired by the previously quoted work of Araki.

However, as we show, his original use of Jensen’s operator inequality contains a subtle flaw when applied in its contractive form. We point out how Petz himself and Nielsen corrected the problem in order to restore the validity of Petz’s original approach.

Uhlmann’s proof, on the other hand, is formulated in terms of interpolations of positive sesquilinear forms, a technique that naturally extends to non-invertible states and arbitrary quantum channels.

Despite their foundational importance, both Petz’s and Uhlmann’s proofs are often considered technically demanding and conceptually opaque. Petz’s approach, while elegant, involves intricate manipulations of operator inequalities that can obscure the overall structure of the argument. Uhlmann’s method introduces a formalism that is unfamiliar to many working in quantum information theory and is rarely presented in full detail in the literature.

We aim to clarify and systematize both strategies by reformulating them in a unified, finite-dimensional operator-theoretic framework, thus making both proofs more accessible to a wider audience.

The paper is structured as follows. In Section 2, we recall the necessary preliminaries on operator convexity, partial trace, and quantum channels. In Section 3, we analyze Petz’s proof, highlight its limitations, and provide a rigorous correction. In Section 4, we develop Uhlmann’s approach in detail and derive the monotonicity of relative entropy in the general setting.

2 Mathematical preliminaries

In this section, we start recalling the basic definitions and results needed for the rest of the paper.

Given a finite-dimensional Hilbert state space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ over the field $\mathbb{F} = \mathbb{R}$ or \mathbb{C} , $\mathcal{B}(\mathcal{H})$ indicates the \mathbb{F} -algebra of linear (bounded) operators $A : \mathcal{H} \rightarrow \mathcal{H}$.

We recall that $A \in \mathcal{B}(\mathcal{H})$ is:

- *positive semi-definite*, written $A \geq 0$, if $\langle x, Ax \rangle_{\mathcal{H}} \geq 0$ for all $x \in \mathcal{H}$;
- *positive-definite*, written $A > 0$, if $\langle x, Ax \rangle_{\mathcal{H}} > 0$ for all $x \neq 0_{\mathcal{H}} \in \mathcal{H}$;
- *Hermitian*, if $A = A^{\dagger}$, where A^{\dagger} is the *adjoint* operator of A , defined by the formula $\langle A^{\dagger}x, y \rangle_{\mathcal{H}} = \langle x, Ay \rangle_{\mathcal{H}}$ for all $x, y \in \mathcal{H}$.

If $\mathbb{F} = \mathbb{C}$, then a positive semi-definite, or positive-definite, operator is automatically Hermitian, but this is not the case if $\mathbb{F} = \mathbb{R}$.

Suppose now that \mathcal{H} is the state space of a quantum system. The density operator (also called density matrix) $\rho \in \mathcal{B}(\mathcal{H})$ associated to a given state s of the system is positive semi-definite, Hermitian and such that $\text{Tr}(\rho) = 1$. Hence, ρ has eigenvalues $0 \leq \lambda_j \leq 1$ which sum up to 1.

As is well known, if s is a pure state, then ρ is a rank-one orthogonal projector and thus not invertible.

$\mathcal{B}(\mathcal{H})$ becomes itself a \mathbb{F} -Hilbert space when it is endowed with the Hilbert-Schmidt operator inner product

$$\langle A, B \rangle_{\mathcal{B}(\mathcal{H})} := \text{Tr}(A^\dagger B), \quad A, B \in \mathcal{B}(\mathcal{H}). \quad (1)$$

The subset of $\mathcal{B}(\mathcal{H})$ given by Hermitian operators on \mathcal{H} , indicated with $\mathcal{B}_H(\mathcal{H})$, is a real Hilbert space with respect to the inner product inherited from $\mathcal{B}(\mathcal{H})$ and also a partially ordered set with respect to the *Löwner ordering*, defined as follows: for all $A, B \in \mathcal{B}_H(\mathcal{H})$, $A \leq B \iff B - A \geq 0$.

Given a function $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$ and $A \in \mathcal{B}_H(\mathcal{H})$ with spectral decomposition $A = UDU^\dagger$, with U unitary and D diagonal with entries given by the eigenvalues λ_j of A , all supposed to belong to I , we write as usual

$$f(A) = Uf(D)U^\dagger, \quad (2)$$

where $f(D)$ is diagonal with non-trivial entries given by $f(\lambda_j)$.

If, for every finite-dimensional \mathcal{H} and every couple of operators $A, B \in \mathcal{B}_H(\mathcal{H})$, we have

$$A \leq B \implies f(A) \leq f(B), \quad (3)$$

then f is said to be *operator monotone* on I . Instead, if we have

$$f\left(\frac{A+B}{2}\right) \leq \frac{f(A)+f(B)}{2}, \quad (4)$$

then f is said to be *operator convex* on I . If the last inequality is reversed, f is said *operator concave* on I .

By Löwner's theorem, see e.g. [2] chapter V.4, $f : I \rightarrow \mathbb{R}$ is operator monotone if and only if it has an analytic continuation that maps the upper half-plane \mathbb{H}_+ into itself. Noticeable examples of operator monotone functions on $(0, +\infty)$ are $x \mapsto \log(x)$ and $x \mapsto -1/x$.

If $f : I \rightarrow \mathbb{R}$ is a continuous and operator monotone function on I , then $\forall a \in I$ the function $F : I \rightarrow \mathbb{R}$ given by

$$F(x) = \int_a^x f(t)dt \quad (5)$$

is operator convex on I , see again [2]. So, since $t \mapsto -1/t$ is operator monotone on $(0, +\infty)$,

$$x \mapsto \int_1^x -\frac{1}{t}dt = -\log(x) \quad (6)$$

is operator convex on $(0, +\infty)$, which implies that $x \mapsto \log(x)$ is operator concave on $(0, +\infty)$.

As it is well-known, a convex function $f : I \rightarrow \mathbb{R}$ satisfies the Jensen inequality

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i), \quad (7)$$

for all $x_1, \dots, x_n \in I$ and non-negative weights $\lambda_1, \dots, \lambda_n$ such that $\sum_{i=1}^n \lambda_i = 1$.

The Jensen inequality can be generalized to operator convex functions, as stated in the following celebrated theorem, proven by Hansen and Pedersen [6].

Theorem 2.1 (Jensen’s operator inequality) For a continuous function $f : I \rightarrow \mathbb{R}$, the following conditions are equivalent.

1. f is operator convex on I .
2. For each natural number $n \geq 1$ the following inequality holds

$$f\left(\sum_{i=1}^n A_i^\dagger X_i A_i\right) \leq \sum_{i=1}^n A_i^\dagger f(X_i) A_i, \quad (8)$$

for every n -tuple of bounded Hermitian operators X_1, \dots, X_n on an arbitrary Hilbert space \mathcal{H} , with spectra contained in I , and every n -tuple of operators (A_1, \dots, A_n) on \mathcal{H} satisfying $\sum_{i=1}^n A_i^\dagger A_i = id_{\mathcal{H}}$.

3. $f(V^\dagger X V) \leq V^\dagger f(X) V$ for every isometry $V : \mathcal{K} \rightarrow \mathcal{H}$ from an arbitrary Hilbert space \mathcal{K} on an arbitrary Hilbert space \mathcal{H} and every Hermitian operator X on \mathcal{H} with spectrum contained in I .

An immediate consequence of this theorem is the following corollary, also known as the *contractive Jensen’s operator inequality*.

Corollary 2.1 (Contractive version) Let $f : I \rightarrow \mathbb{R}$ be a continuous function and suppose that $0 \in I$. Then f is operator convex on I and $f(0) \leq 0$ if and only if for some, hence for every, $n \geq 1$ inequality (8) is valid for every n -tuple of bounded Hermitian operators X_1, \dots, X_n on an arbitrary Hilbert space \mathcal{H} , with spectra contained in I , and every n -tuple of operators (A_1, \dots, A_n) on \mathcal{H} satisfying $\sum_{i=1}^n A_i^\dagger A_i \leq id_{\mathcal{H}}$.

The reason for the adjective ‘contractive’ can be easily understood by setting $n = 1$: in this case, it follows that f is operator convex on an interval I containing 0 with $f(0) \leq 0$ if and only if

$$f(A^\dagger X A) \leq A^\dagger f(X) A, \quad (9)$$

for every bounded Hermitian operator X with spectrum in I and every operator A such that $A^\dagger A \leq id$, which implies that A is a contraction, i.e. $\|Ax\| \leq \|x\|$ for all $x \in \mathcal{H}$ or, equivalently, $\|A\| \leq 1$.

Let us now consider two interacting quantum systems a and b with Hilbert state spaces \mathcal{H}_a and \mathcal{H}_b , respectively. The associated composite quantum system ab has Hilbert state space given by $\mathcal{H}_{ab} = \mathcal{H}_a \otimes \mathcal{H}_b$.

In the following, we indicate with X_a, X_b, X_{ab} generic operators of $\mathcal{B}_H(\mathcal{H}_a)$, $\mathcal{B}_H(\mathcal{H}_b)$ and $\mathcal{B}_H(\mathcal{H}_{ab})$, respectively.

The *partial trace* Tr_b over the subsystem b is a ‘superoperator’, i.e. a linear map $\text{Tr}_b \in \mathcal{B}(\mathcal{B}_H(\mathcal{H}_{ab}), \mathcal{B}_H(\mathcal{H}_a))$, defined as the linear extension to the whole $\mathcal{B}_H(\mathcal{H}_{ab})$ of the map

$$\text{Tr}_b(X_a \otimes X_b) = \text{Tr}(X_b) X_a. \quad (10)$$

Tr_b satisfies the following property

$$\text{Tr}(\text{Tr}_b(X_{ab}) X_a) = \text{Tr}(X_{ab}(X_a \otimes id_b)). \quad (11)$$

see e.g. [7], page 100.

The partial trace Tr_b is:

- *trace-preserving*: $\text{Tr}(X_{ab}) = \text{Tr}(\text{Tr}_b(X_{ab}))$;
- *positive*: if $X_{ab} \geq 0$, then $\text{Tr}_b(X_{ab}) \geq 0$.

As a consequence, Tr_b maps states of \mathcal{H}_{ab} into states of \mathcal{H}_a .

Moreover, Tr_b is *completely positive*, i.e. $\text{Tr}_b \otimes id_{\mathcal{H}}$ is a positive linear map for all Hilbert spaces \mathcal{H} . A trace-preserving and completely positive (or ‘CPTP’) linear map is called a *channel*.

The partial trace is actually one of the three main elements of every channel. In fact, thanks to the Stinespring theorem, see [23], given any channel $\mathcal{C} \in \mathcal{B}(\mathcal{B}_H(\mathcal{H}_a), \mathcal{B}_H(\mathcal{H}_a))$, there exist an operator $Y \in \mathcal{B}_H(\mathcal{H}_b)$ and a unitary operator U on $\mathcal{H}_a \otimes \mathcal{H}_b$ such that:

$$\mathcal{C}(X) := \text{Tr}_b(U(X \otimes Y)U^\dagger), \quad \forall X \in \mathcal{B}_H(\mathcal{H}_a). \quad (12)$$

By the Riesz representation theorem, we can define the adjoint of the partial trace Tr_b as the only operator $\text{Tr}_b^\dagger : \mathcal{B}_H(\mathcal{H}_a) \rightarrow \mathcal{B}_H(\mathcal{H}_{ab})$ satisfying

$$\langle X_{ab}, \text{Tr}_b^\dagger(X_a) \rangle_{\mathcal{B}_H(\mathcal{H}_{ab})} = \langle \text{Tr}_b(X_{ab}), X_a \rangle_{\mathcal{B}_H(\mathcal{H}_a)}. \quad (13)$$

Writing the inner products explicitly and using property (11), we find

$$\text{Tr}(X_{ab} \text{Tr}_b^\dagger(X_a)) = \text{Tr}(\text{Tr}_b(X_{ab})X_a) = \text{Tr}(X_{ab}(X_a \otimes id_b)), \quad (14)$$

which allows us to write the explicit action of Tr_b^\dagger as

$$\text{Tr}_b^\dagger(X_a) = X_a \otimes id_b. \quad (15)$$

This formula implies that $(\text{Tr}_b^\dagger(X_a))^\dagger = \text{Tr}_b^\dagger(X_a^\dagger)$ and that Tr_b^\dagger is *unital*, i.e. it maps the identity of its domain to the identity of its range

$$\text{Tr}_b^\dagger(id_a) = id_{ab}. \quad (16)$$

Being a completely positive unital transformation, Tr_b^\dagger is a so-called *Schwarz map*, see [3] Corollary 2.8, i.e. it satisfies the following inequality:

$$\text{Tr}_b^\dagger(X_a^\dagger)\text{Tr}_b^\dagger(X_a) \leq \text{Tr}_b^\dagger(X_a^\dagger X_a). \quad (17)$$

This property is shared by the adjoint of any channel \mathcal{C} .

2.1 Relative entropy in quantum information theory

We devote a separate subsection to relative entropy, as its proper treatment entails the detailed development of several technical aspects.

This is essential for addressing certain subtleties that are often overlooked in the literature, yet play a crucial role in the rigorous analysis of relative entropy.

Given any finite-dimensional Hilbert space \mathcal{H} and any density operator $\rho \in \mathcal{B}_H(\mathcal{H})$ with eigenvalues $\lambda_j \geq 0$, we indicate with $\text{Sp}(\rho)$ its spectrum and with $\text{Sp}_+(\rho)$ the subset of $\text{Sp}(\rho)$ composed only by *strictly positive* eigenvalues.

It will be convenient to have the following index sets at hand: $I(\rho) = \{j : \lambda_j \in \text{Sp}(\rho)\}$ and $I_+(\rho) = \{j : \lambda_j \in \text{Sp}_+(\rho)\}$.

The *support* of ρ , denoted with $\text{supp}(\rho)$, is the subset of \mathcal{H} defined as follows

$$\begin{aligned} \text{supp}(\rho) &:= \text{span}\{x_j \in \mathcal{H} : x_j \text{ is an eigenvector of } \rho \text{ with eigenvalue } \lambda_j \in \text{Sp}_+(\rho)\} \\ &= \bigoplus_{j \in I_+(\rho)} E_{\lambda_j}, \end{aligned} \quad (18)$$

where E_{λ_j} is the eigenspace relative to the eigenvalue λ_j .

From this definition and the spectral theorem for Hermitian operators, it immediately follows that

$$\text{supp}(\rho) = \ker(\rho)^\perp = \text{Im}(\rho). \quad (19)$$

This implies that $\mathcal{H} = \ker(\rho) \oplus \text{supp}(\rho)$, so the positive-definite operator $\rho|_{\text{supp}(\rho)}$ has a trivial kernel, hence it is invertible.

In particular, if ρ is positive-definite then ρ is invertible everywhere and its image and support coincide with the entire \mathcal{H} . On the other hand, non-invertible density operators, for instance those corresponding to pure states, have support strictly included in \mathcal{H} .

The deviation from purity, or mixedness, of ρ can be measured by its *von Neumann entropy*, defined as follows

$$S(\rho) := -\text{Tr}(\rho \log \rho). \quad (20)$$

The condition $S(\rho) = 0$ is satisfied if and only if ρ is pure. In the literature, the precise definition of the logarithm of a matrix may vary according to the specific aims and context in which the matrix logarithm is employed.

For the purposes of this paper, the key property that the logarithm of a density operator ρ , or more generally of a Hermitian operator, must satisfy is $\exp(\log \rho) = \rho$. For this reason, we adopt a definition of \log via functional calculus in the extended real numbers with the conventions:

$$\log(0) = -\infty, \quad \exp(-\infty) = 0, \quad 0 \log 0 = 0, \quad (21)$$

of course justified by the limits $\log \varepsilon \rightarrow -\infty$ as $\varepsilon \rightarrow 0^+$, $\exp(-M) \rightarrow 0^+$ as $M \rightarrow +\infty$, and $\varepsilon \log \varepsilon \rightarrow 0$, as $\varepsilon \rightarrow 0^+$.

Using the spectral theorem, if ρ is decomposed as

$$\rho = \sum_{j \in I_+(\rho)} \lambda_j P_j + 0P_0, \quad (22)$$

where P_j denotes the orthogonal projector onto the eigenspace E_{λ_j} , $j \in I_+(\rho)$, and P_0 is the orthogonal projector on $\ker(\rho)$, then, using the previous conventions, we have

$$\log \rho := \sum_{j \in I_+(\rho)} \log(\lambda_j) P_j + (-\infty) P_0, \quad (23)$$

and so

$$\exp(\log \rho) = \sum_{j \in I_+(\rho)} \exp(\log \lambda_j) P_j + \exp(-\infty) P_0 = \sum_{j \in I_+(\rho)} \lambda_j P_j + 0P_0 = \rho. \quad (24)$$

Since the projectors satisfy the orthogonality relation $P_i P_j = \delta_{ij} P_j$ and by the convention $0 \log 0 = 0$, which accounts for the zero eigenvalue, we obtain

$$S(\rho) = -\text{Tr}(\rho \log \rho) = - \sum_{j \in I_+(\rho)} \lambda_j \log(\lambda_j) \text{Tr}(P_j), \quad (25)$$

where $\text{Tr}(P_j)$ is the multiplicity of the eigenvalue λ_j . In the literature, it is actually more common to write just

$$S(\rho) = - \sum_{j \in I_+(\rho)} \lambda_j \log(\lambda_j), \quad (26)$$

with the understanding that each eigenvalue is repeated according to its multiplicity.

The von Neumann entropy does not, by itself, tell us how one state differs from another. To capture the distinguishability of two states, the concept of *relative entropy* (sometimes called the *quantum Kullback–Leibler divergence*) must be introduced.

The definition of the relative entropy between two density operators ρ and σ is subjected to a condition on the compatibility of their supports, precisely:

$$S(\rho \parallel \sigma) := \begin{cases} \text{Tr}(\rho \log \rho - \rho \log \sigma) & \text{if } \text{supp}(\rho) \subseteq \text{supp}(\sigma) \\ +\infty & \text{if } \text{supp}(\rho) \not\subseteq \text{supp}(\sigma) \end{cases}, \quad (27)$$

which is known as the ‘*support-based definition*’ of relative entropy.

In order to understand why the condition $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ is both necessary and sufficient for $S(\rho||\sigma)$ to be well-defined, we first notice that

$$\text{Tr}(\rho \log \rho - \rho \log \sigma) = \text{Tr}(\rho \log \rho) - \text{Tr}(\rho \log \sigma), \quad (28)$$

so, the first term of formula (28) is minus the von Neumann entropy, which is always finite. In order to analyze the second term, let us consider, alongside eq. (22), the analogous spectral decomposition of σ

$$\sigma = \sum_{k \in I_+(\sigma)} \mu_k \Pi_k + 0 \Pi_0, \quad (29)$$

so that

$$\log \sigma = \sum_{k \in I_+(\sigma)} \log(\mu_k) \Pi_k + \log(0) \Pi_0, \quad (30)$$

where Π_k is the orthogonal projector on the eigenspace relative to the positive eigenvalue μ_k of σ , and Π_0 is the projector on $\ker(\sigma)$. Then,

$$\text{Tr}(\rho \log \sigma) = \sum_{j \in I_+(\rho)} \sum_{k \in I_+(\sigma)} \lambda_j \log(\mu_k) \text{Tr}(P_j \Pi_k) + \sum_{j \in I_+(\rho)} \lambda_j \log(0) \text{Tr}(P_j \Pi_0), \quad (31)$$

where the contribution of P_0 does not appear due to the convention $0 \log(0) = 0$.

The first term is always finite, since only strictly positive eigenvalues appear. Instead, the behavior of the second term depends on the value of $\text{Tr}(P_j \Pi_0)$. To compute it, let us consider any eigenbasis $(x_j)_{j \in I_+(\rho)}$ of $\text{supp}(\rho)$, so that $P_j(x_j) = x_j$ for all $j \in I_+(\rho)$, and use the fact that orthogonal projectors are Hermitian and idempotent to write

$$\begin{aligned} \text{Tr}(P_j \Pi_0) &= \sum_{j \in I_+(\rho)} \langle x_j, P_j \Pi_0 x_j \rangle = \sum_{j \in I_+(\rho)} \langle P_j x_j, \Pi_0 x_j \rangle = \sum_{j \in I_+(\rho)} \langle x_j, \Pi_0 x_j \rangle \\ &= \sum_{j \in I_+(\rho)} \langle x_j, \Pi_0 \Pi_0 x_j \rangle = \sum_{j \in I_+(\rho)} \langle \Pi_0 x_j, \Pi_0 x_j \rangle = \sum_{j \in I_+(\rho)} \|\Pi_0 x_j\|^2 \geq 0. \end{aligned} \quad (32)$$

The second term in eq. (31) diverges to $-\infty$ if and only if $\text{Tr}(P_j \Pi_0) > 0$, i.e. when there exists at least one $j \in I_+(\rho)$ such that $x_j \in \text{supp}(\rho)$ has a non-trivial projection onto $\ker(\sigma)$, i.e. $\Pi_0 x_j$ is non null, which is equivalent to saying that $x_j \notin \ker(\sigma)^\perp = \text{supp}(\sigma)$, therefore

$$\text{Tr}(\rho \log \sigma) = -\infty \iff \text{Tr}(P_j \Pi_0) > 0 \iff \text{supp}(\rho) \not\subseteq \text{supp}(\sigma), \quad (33)$$

which justifies the definition given in (27).

Instead, if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, then $\text{Tr}(P_j \Pi_0) = 0$ and, using again the convention $0 \log(0) = 0$, the second term in eq. (31) vanishes and we remain just with the first term, which can be written in an alternative form: consider now the orthonormal eigenbases $(x_j)_{j \in J_+(\rho)}$, $(y_k)_{k \in I_+(\sigma)}$ of $\text{supp}(\rho)$ and $\text{supp}(\sigma)$, respectively, then we have the following well-known formula, see e.g. [14],

$$\text{Tr}(P_j \Pi_k) = |\langle x_j, y_k \rangle|^2, \quad (34)$$

which shows that the factor $\text{Tr}(P_j \Pi_k)$ codifies the transition probability between the pure states represented by the unit vectors x_j and y_k .

In summary, when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, the relative entropy between ρ and σ can be written explicitly as follows

$$\begin{aligned} S(\rho||\sigma) &= -S(\rho) - \sum_{j \in I_+(\rho)} \sum_{k \in I_+(\sigma)} \lambda_j \log(\mu_k) \text{Tr}(P_j \Pi_k) \\ &= \sum_{j \in I_+(\rho)} \lambda_j \left(\log(\lambda_j) - \sum_{k \in I_+(\sigma)} \log(\mu_k) |\langle x_j, y_k \rangle|^2 \right). \end{aligned} \quad (35)$$

Two cases are particularly relevant in practical contexts:

- if $\rho, \sigma > 0$, then their supports coincide with the entire Hilbert space \mathcal{H} and so their relative entropy is the finite real number given by eq. (35);
- instead, if $\rho > 0$ but σ is not, then their relative entropy is infinite. This happens, for instance, when ρ is a full-rank mixed state and σ is a pure state.

An equivalent and useful definition of the von Neumann and relative entropy appears in the literature (see, e.g., [28]). Rather than relying on the support of the density operators, this alternative definition is based on a regularization procedure.

Specifically, given a state ρ on \mathcal{H} , one defines the regularized operator

$$\rho_\varepsilon := \rho + \varepsilon id_{\mathcal{H}}, \quad \varepsilon > 0. \quad (36)$$

ρ_ε is a positive-definite Hermitian operator and the spectral decompositions of ρ_ε and $\log \rho_\varepsilon$ are

$$\rho_\varepsilon = \sum_{j \in I(\rho)} (\lambda_j + \varepsilon) P_j, \quad \log \rho_\varepsilon = \sum_{j \in I(\rho)} \log(\lambda_j + \varepsilon) P_j. \quad (37)$$

We have

$$(\lambda_j + \varepsilon) \log(\lambda_j + \varepsilon) \xrightarrow{\varepsilon \rightarrow 0^+} \begin{cases} \lambda_j \log(\lambda_j) & \text{if } j \in I_+(\rho) \\ 0 & \text{if } \lambda_j = 0 \end{cases}, \quad (38)$$

so the von Neumann entropy of ρ is well-defined via the limit

$$S(\rho) := - \lim_{\varepsilon \rightarrow 0^+} \text{Tr}(\rho_\varepsilon \log \rho_\varepsilon). \quad (39)$$

Similarly, setting $\sigma_\varepsilon := \sigma + \varepsilon id_{\mathcal{H}}$, $\varepsilon > 0$, the relative entropy between ρ and σ can be defined as

$$S(\rho \parallel \sigma) := \lim_{\varepsilon \rightarrow 0^+} \text{Tr}(\rho_\varepsilon \log \rho_\varepsilon - \rho_\varepsilon \log \sigma_\varepsilon), \quad (40)$$

known as the ‘regularized definition’ of relative entropy.

Let us verify that the regularized and support-based definitions of relative entropy coincide. Equation (40) splits in two terms, the first equals minus the regularized definition of the von Neumann entropy, which is finite by eq. (39), so the only issue to address concerns the second term of eq. (40). For that, we write the spectral decomposition of σ_ε as follows

$$\sigma_\varepsilon = \sum_{k \in I_+(\sigma)} (\mu_k + \varepsilon) \Pi_k + \varepsilon \Pi_0. \quad (41)$$

Repeating the same computations performed in the case of the support-based definition of $S(\rho \parallel \sigma)$, we obtain

$$\text{Tr}(\rho_\varepsilon \log \sigma_\varepsilon) = \sum_{k \in I_+(\sigma)} \log(\mu_k + \varepsilon) \text{Tr}(\rho_\varepsilon \Pi_k) + \log(\varepsilon) \text{Tr}(\rho_\varepsilon \Pi_0). \quad (42)$$

If $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, then $\text{Tr}(\rho_\varepsilon \Pi_0) \rightarrow 0$ when $\varepsilon \rightarrow 0^+$ and the last term in eq. (42) is absent, so the limit converges to the correct value.

Instead, if $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, then $\text{Tr}(\rho_\varepsilon \Pi_0) \rightarrow \alpha > 0$ when $\varepsilon \rightarrow 0^+$, consequently the second term in eq. (42) diverges to $-\infty$, thus matching the behavior of the support-based definition.

The relative entropy has several important properties, see e.g. [27].

- *Klein’s inequality*: $S(\rho \parallel \sigma) \geq 0$ for all ρ, σ , and $S(\rho \parallel \sigma) = 0$ if and only if $\rho = \sigma$. This property motivates why the relative entropy, despite lacking symmetry in its arguments, is taken to be a measure of distinguishability of states in quantum theories.
- *Invariance under unitary conjugation*: $S(U\rho U^\dagger \parallel U\sigma U^\dagger) = S(\rho \parallel \sigma)$, for all unitary operator U acting on the same Hilbert space as ρ and σ .
- *Additivity w.r.t. tensor product*: $S(\rho_1 \otimes \rho_2 \parallel \sigma_1 \otimes \sigma_2) = S(\rho_1 \parallel \sigma_1) + S(\rho_2 \parallel \sigma_2)$, for all density operators $\sigma_j, \rho_j, j = 1, 2$.

The *monotonicity of S under partial trace* is represented by the inequality

$$S(\mathrm{Tr}_b(\rho) \parallel \mathrm{Tr}_b(\sigma)) \leq S(\rho \parallel \sigma), \quad (43)$$

which, together with Stinespring's theorem and the three previously mentioned properties of S , permits to prove that quantum distinguishability does not increase under the action of a generic channel \mathcal{C}

$$S(\mathcal{C}(\rho) \parallel \mathcal{C}(\sigma)) \leq S(\rho \parallel \sigma), \quad (44)$$

a formula also known as *data processing inequality* (DPI). In fact:

$$\begin{aligned} S(\mathcal{C}(\rho) \parallel \mathcal{C}(\sigma)) &= S(\mathrm{Tr}_b(U(\rho \otimes X)U^\dagger) \parallel \mathrm{Tr}_b(U(\sigma \otimes X)U^\dagger)) \\ &\leq S(U(\rho \otimes X)U^\dagger \parallel U(\sigma \otimes X)U^\dagger) \\ &= S(\rho \otimes X \parallel \sigma \otimes X) = S(\rho \parallel \sigma) + S(X, X) \\ &= S(\rho \parallel \sigma). \end{aligned} \quad (45)$$

Inequality (44) explains why, in the quantum information literature, a channel is often referred to as a *coarse-graining* procedure, a term borrowed from statistical mechanics.

This terminology reflects the idea that information about different quantum states is lost through the action of the channel, as previously distinguishable states may become indistinguishable after the transformation.

While the first three properties of S mentioned above are relatively straightforward to prove, its monotonicity under partial trace is considerably more subtle. In the next two sections, we provide a detailed analysis of the proof originally proposed by Petz and Uhlmann.

3 Petz's proof of the monotonicity of the relative entropy under partial trace

In this section, we examine the strategy proposed by Petz in [18] for proving the monotonicity of relative entropy under the partial trace operation, which is based on a clever reformulation of the expression of the relative entropy as a suitable inner product inspired by an analogous construction by Araki [1].

We will show that Petz's proof is flawed due to an incorrect application of the contractive version of Jensen's operator inequality. We will explain how this issue can be circumvented, thus restoring the validity of the overall Petz's approach. Furthermore, we will show how to extend it to incorporate also non-invertible density operators.

The notation that will be used throughout this section is detailed below:

- given the finite-dimensional Hilbert spaces $(\mathcal{H}_a, \langle \cdot, \cdot \rangle_a)$ and $(\mathcal{H}_b, \langle \cdot, \cdot \rangle_b)$, we define $\mathcal{H}_{ab} := \mathcal{H}_a \otimes \mathcal{H}_b$, with inner product $\langle \cdot, \cdot \rangle_{ab}$ induced by those of \mathcal{H}_a and \mathcal{H}_b ;
- operators of $\mathcal{B}_H(\mathcal{H}_a)$ will be denoted as X, Y, Z and operators of $\mathcal{B}_H(\mathcal{H}_{ab})$ by A, B, C ;
- $\rho, \sigma \in \mathcal{B}_H(\mathcal{H}_{ab})$ are two positive-definite (invertible) density operators²: $\rho, \sigma > 0$.

In order to rewrite the relative entropy as an inner product, we must introduce some suitable superoperators.

Precisely, fixed $B, C \in \mathcal{B}_H(\mathcal{H}_{ab})$, $B > 0$, consider the Hermitian superoperators $R_B, L_C, \Delta_{B,C} \in \mathcal{B}_H(\mathcal{B}_H(\mathcal{H}_{ab}))$ given by the right and left operator multiplications and the relative modular operator, respectively, i.e.

$$R_B(A) := AB, \quad L_C(A) := CA, \quad \Delta_{B,C}(A) := L_C \circ R_B^{-1}(A) = CAB^{-1}. \quad (46)$$

²Actually, for the following analysis, only ρ has to be invertible, however, as we have noted with the support-based definition of the relative entropy, if $\rho > 0$, then its support is the entire \mathcal{H} and so, for our analysis to be meaningful, we also have to demand $\sigma > 0$.

Clearly, $R_B^{-1} = R_{B^{-1}}$, $L_C^{-1} = L_{C^{-1}}$ and

$$[L_C, R_B] = [L_C, R_B^{-1}] = 0. \quad (47)$$

All these superoperators are Hermitian, in fact

$$\langle R_B(A), C \rangle_{ab} = \text{Tr}((AB)^\dagger C) = \text{Tr}(BAC) = \text{Tr}(ACB) = \langle A, R_B(C) \rangle_{ab}, \quad (48)$$

and similarly for L_C . Regarding $\Delta_{B,C}$, using eq. (47) we have

$$\Delta_{B,C}^\dagger = (L_C R_B^{-1})^\dagger = (R_B^{-1})^\dagger L_C = (R_B^\dagger)^{-1} L_C = R_B^{-1} \circ L_C = L_C \circ R_B^{-1} = \Delta_{B,C}. \quad (49)$$

If we consider the specific case in which $A = \rho > 0$ and $B = \sigma > 0$, then we obtain

$$\begin{aligned} \log(\Delta_{\rho,\sigma}) &= \log(L_\sigma \circ R_\rho^{-1}) = \log(\exp(\log(L_\sigma)) \exp(\log(R_\rho^{-1}))) \\ &= \log(\exp(\log(L_\sigma) - \log(R_\rho))) = \log(L_\sigma) - \log(R_\rho). \end{aligned} \quad (50)$$

Since $\log(L_\sigma) = L_{\log(\sigma)}$ and $\log(R_\rho) = R_{\log(\rho)}$, applying formula (50) to $\rho^{1/2}$ gives

$$\log(\Delta_{\rho,\sigma})(\rho^{1/2}) = \log(\sigma)\rho^{1/2} - \rho^{1/2}\log(\rho). \quad (51)$$

This last identity, the cyclic property of the trace and the property $[\rho^{1/2}, \log(\rho)] = 0$ imply

$$\begin{aligned} S(\rho\|\sigma) &= \text{Tr}[\rho \log(\rho) - \rho \log(\sigma)] = \text{Tr}[\rho^{1/2}(\rho^{1/2} \log(\rho) - \rho^{1/2} \log(\sigma))] \\ &= \text{Tr}[(\rho^{1/2} \log(\rho) - \rho^{1/2} \log(\sigma))\rho^{1/2}] \\ &= \text{Tr}[\rho^{1/2} \log(\rho)\rho^{1/2} - \rho^{1/2} \log(\sigma)\rho^{1/2}] \\ &= \text{Tr}[\rho^{1/2} \rho^{1/2} \log(\rho) - \rho^{1/2} \log(\sigma)\rho^{1/2}] \\ &= -\text{Tr}[\rho^{1/2}(\log(\sigma)\rho^{1/2} - \rho^{1/2} \log(\rho))] \\ &= -\langle \rho^{1/2}, \log(\Delta_{\rho,\sigma})(\rho^{1/2}) \rangle_{ab}, \end{aligned} \quad (52)$$

which is the ‘inner product-reformulation’ of the relative entropy between the positive-definite states ρ and σ that we were searching for.

We can obtain an analogous formula for the relative entropy of the partial traces of ρ and σ . To do that, if we fix $Y, Z \in \mathcal{B}_H(\mathcal{H}_a)$, $Y > 0$, then we can define the Hermitian superoperators $R_Y^a, L_Z^a, \Delta_{\rho,\sigma}^a \in \mathcal{B}_H(\mathcal{B}_H(\mathcal{H}_a))$ as follows

$$R_Y^a(X) := XY, \quad L_Z^a(X) := ZX, \quad \Delta_{Y,Z}^a(X) := L_Z \circ R_Y^{-1}(X) = ZXY^{-1}. \quad (53)$$

By carrying out computations analogous to those in eq. (52), but this time using the superoperators $R_{\text{Tr}_b(\rho)}^a, L_{\text{Tr}_b(\sigma)}^a, \Delta_{\rho,\sigma}^a := L_{\text{Tr}_b(\sigma)}^a \circ R_{\text{Tr}_b(\rho)}^a$, we obtain

$$S(\text{Tr}_b(\rho)\|\text{Tr}_b(\sigma)) = -\langle \text{Tr}_b(\rho)^{1/2}, \log(\Delta_{\rho,\sigma}^a)(\text{Tr}_b(\rho)^{1/2}) \rangle_a. \quad (54)$$

Due to the minus sign in front of the inner products appearing in eqs. (52) and (54), we have that *the monotonicity of the relative entropy under partial trace is equivalent to:*

$$\langle \rho^{1/2}, \log(\Delta_{\rho,\sigma})(\rho^{1/2}) \rangle_{ab} \leq \langle \text{Tr}_b(\rho)^{1/2}, \log(\Delta_{\rho,\sigma}^a)(\text{Tr}_b(\rho)^{1/2}) \rangle_a. \quad (55)$$

These inner products are defined on two different Hilbert spaces, $\mathcal{B}_H(\mathcal{H}_{ab})$ and $\mathcal{B}_H(\mathcal{H}_a)$, in order to perform a meaningful comparison and prove the inequality, Petz introduced a superoperator $V_\rho : \mathcal{B}_H(\mathcal{H}_a) \rightarrow \mathcal{B}_H(\mathcal{H}_{ab})$ through the explicit formula

$$V_\rho(X \text{Tr}_b(\rho)^{1/2}) := \text{Tr}_b^\dagger(X) \rho^{1/2}, \quad (56)$$

which serves as a bridge between the reduced state $\text{Tr}_b(\rho)$ and the full state ρ . While, as we are going to show, this definition is computationally effective, it may at first appear somewhat ad hoc.

Actually, a seemingly more natural choice for V_ρ would have been Tr_b^\dagger for two reasons: first, as for V_ρ , Tr_b^\dagger is a map between $\mathcal{B}_H(\mathcal{H}_a)$ and $\mathcal{B}_H(\mathcal{H}_{ab})$ and, second, it satisfies the Schwarz inequality (17), which in the following will play a crucial role in the proof of inequality (55).

It turns out that V_ρ is tightly related to the adjoint of the partial trace Tr_b , not w.r.t. the original inner products of \mathcal{H}_a and \mathcal{H}_{ab} , but w.r.t. suitably *weighted inner products* that naturally emerge from the previous structural analysis of the relative entropy in terms of superoperators.

In fact, the reformulations of the relative entropy obtained in eqs. (52) and (54), respectively, involve inner products weighted by $\rho^{1/2}$ and $\text{Tr}_b(\rho)^{1/2}$. This observation suggests that the correct notion of adjoint to consider for Tr_b is the one defined relative to the inner products³

$$\langle A, B \rangle_{ab, \rho} := \langle R_{\rho^{-1/2}}(A), B \rangle_{ab}, \quad \langle X, Y \rangle_{a, \rho} := \langle R_{\text{Tr}_b(\rho)^{-1/2}}^a(X), Y \rangle_a. \quad (57)$$

We have

$$\begin{aligned} \langle X, \text{Tr}_b(A) \rangle_{a, \rho} &= \langle R_{\text{Tr}_b(\rho)^{-1/2}}^a(X), \text{Tr}_b(A) \rangle_a = \langle \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\rho)^{-1/2}}^a(X), A \rangle_{ab} \\ &= \langle R_{\rho^{-1/2}} \circ R_{\rho^{1/2}} \circ \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\rho)^{-1/2}}^a(X), A \rangle_{ab} \\ &= \langle R_{\rho^{1/2}} \circ \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\rho)^{-1/2}}^a(X), A \rangle_{ab, \rho}. \end{aligned} \quad (58)$$

So, the adjoint of Tr_b w.r.t. the weighted inner products defined above is the operator $\text{Tr}_b^{\dagger, \rho} := R_{\rho^{1/2}} \circ \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\rho)^{-1/2}}^a$, which, for all $X \in \mathcal{B}_H(\mathcal{H}_a)$, satisfies

$$(R_{\rho^{1/2}} \circ \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\rho)^{-1/2}}^a)(X \text{Tr}_b(\rho)^{1/2}) = (R_{\rho^{1/2}} \circ \text{Tr}_b^\dagger)(X) = \text{Tr}_b^\dagger(X) \rho^{1/2}, \quad (59)$$

therefore $V_\rho = \text{Tr}_b^{\dagger, \rho}$, i.e.

$$V_\rho = R_{\rho^{1/2}} \circ \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\rho)^{-1/2}}^a. \quad (60)$$

From this fact, we obtain that the explicit action of V_ρ on any $X \in \mathcal{B}_H(\mathcal{H}_a)$ is

$$V_\rho(X) = (X \text{Tr}_b(\rho)^{-1/2} \otimes id_b) \rho^{1/2}, \quad (61)$$

and this implies immediately that V_ρ transforms $\text{Tr}_b(\rho)^{1/2}$ into $\rho^{1/2}$

$$V_\rho(\text{Tr}_b(\rho)^{1/2}) = \rho^{1/2}. \quad (62)$$

Using repeatedly the cyclic property of the trace and Schwarz's inequality (17) satisfied by Tr_b^\dagger , we can prove that V_ρ is a contraction

$$\begin{aligned} \|V_\rho(X \text{Tr}_b(\rho)^{1/2})\|^2 &= \langle \text{Tr}_b^\dagger(X) \rho^{1/2}, \text{Tr}_b^\dagger(X) \rho^{1/2} \rangle_{ab} = \text{Tr}[(\text{Tr}_b^\dagger(X) \rho^{1/2})^\dagger \text{Tr}_b^\dagger(X) \rho^{1/2}] \\ &= \text{Tr}[\rho^{1/2} \text{Tr}_b^\dagger(X^\dagger) \text{Tr}_b^\dagger(X) \rho^{1/2}] = \text{Tr}[\text{Tr}_b^\dagger(X^\dagger) \text{Tr}_b^\dagger(X) \rho] \\ &\leq \text{Tr}[\text{Tr}_b^\dagger(X^\dagger X) \rho] = \langle \text{Tr}_b^\dagger(X^\dagger X), \rho \rangle_{ab} = \langle X^\dagger X, \text{Tr}_b(\rho) \rangle_a \\ &= \text{Tr}(X^\dagger X \text{Tr}_b(\rho)^{1/2} \text{Tr}_b(\rho)^{1/2}) = \text{Tr}[\text{Tr}_b(\rho)^{1/2} X^\dagger X \text{Tr}_b(\rho)^{1/2}] \\ &= \text{Tr}[(X \text{Tr}_b(\rho)^{1/2})^\dagger X \text{Tr}_b(\rho)^{1/2}] = \langle X \text{Tr}_b(\rho)^{1/2}, X \text{Tr}_b(\rho)^{1/2} \rangle_a \\ &= \|X \text{Tr}_b(\rho)^{1/2}\|^2. \end{aligned} \quad (63)$$

Note now that

$$\begin{aligned} \langle \Delta_{\rho, \sigma}^a(X) \text{Tr}_b(\rho)^{1/2}, X \text{Tr}_b(\rho)^{1/2} \rangle_a &= \langle \text{Tr}_b(\sigma) X \text{Tr}_b(\rho)^{-1/2}, X \text{Tr}_b(\rho)^{1/2} \rangle_a \\ &= \text{Tr}[(\text{Tr}_b(\sigma) X \text{Tr}_b(\rho)^{-1/2})^\dagger X \text{Tr}_b(\rho)^{1/2}] \\ &= \text{Tr}[\text{Tr}_b(\rho)^{-1/2} X^\dagger \text{Tr}_b(\sigma) X \text{Tr}_b(\rho)^{1/2}] \\ &= \text{Tr}[X X^\dagger \text{Tr}_b(\sigma)]. \end{aligned} \quad (64)$$

³The positive-definiteness of these inner products is guaranteed by the fact that $\rho, \text{Tr}_b(\rho) > 0$.

Using the equality just proven, we can show that $V_\rho^\dagger \Delta_{\rho,\sigma} V_\rho \leq \Delta_{\rho,\sigma}^a$, in fact

$$\begin{aligned}
\langle V_\rho^\dagger \Delta_{\rho,\sigma} V_\rho (X \text{Tr}_b(\rho)^{1/2}), X \text{Tr}_b(\rho)^{1/2} \rangle_a &= \langle \Delta_{\rho,\sigma} V_\rho (X \text{Tr}_b(\rho)^{1/2}), V_\rho (X \text{Tr}_b(\rho)^{1/2}) \rangle_{ab} \\
&= \langle \Delta_{\rho,\sigma} (\text{Tr}_b^\dagger(X) \rho^{1/2}), \text{Tr}_b^\dagger(X) \rho^{1/2} \rangle_{ab} \\
&= \langle \sigma \text{Tr}_b^\dagger(X) \rho^{-1/2}, \text{Tr}_b^\dagger(X) \rho^{1/2} \rangle \\
&= \text{Tr}[\rho^{-1/2} \text{Tr}_b^\dagger(X^\dagger) \sigma \text{Tr}_b^\dagger(X) \rho^{1/2}] \\
&= \text{Tr}[\text{Tr}_b^\dagger(X) \text{Tr}_b^\dagger(X^\dagger) \sigma] \\
&\leq \text{Tr}[\text{Tr}_b^\dagger(X X^\dagger) \sigma] = \langle \text{Tr}_b^\dagger(X X^\dagger), \sigma \rangle_{ab} \\
&= \langle X X^\dagger, \text{Tr}_b(\sigma) \rangle_a = \text{Tr}[X X^\dagger \text{Tr}_b(\sigma)] \\
&= \langle \Delta_{\rho,\sigma}^a(X) \text{Tr}_b(\rho)^{1/2}, X \text{Tr}_b(\rho)^{1/2} \rangle_a,
\end{aligned} \tag{65}$$

where we have used again Schwarz's inequality and, to write the last equality, we have applied eq. (64).

Since $\log(x)$ is an operator monotone function, the inequality $V_\rho^\dagger \Delta_{\rho,\sigma} V_\rho \leq \Delta_{\rho,\sigma}^a$ implies

$$\log(V_\rho^\dagger \Delta_{\rho,\sigma} V_\rho) \leq \log(\Delta_{\rho,\sigma}^a), \tag{66}$$

and so

$$\begin{aligned}
S(\text{Tr}_b(\rho) \| \text{Tr}_b(\sigma)) &= \langle \text{Tr}_b(\rho)^{1/2}, -\log(\Delta_{\rho,\sigma}^a)(\text{Tr}_b(\rho)^{1/2}) \rangle_a \\
&\leq \langle \text{Tr}_b(\rho)^{1/2}, -\log(V_\rho^\dagger \Delta_{\rho,\sigma} V_\rho)(\text{Tr}_b(\rho)^{1/2}) \rangle_a.
\end{aligned} \tag{67}$$

Petz's strategy to make the relative entropy $S(\rho \| \sigma)$ appear on the right-hand side of the previous inequality consists in using the fact that V_ρ is a contraction to apply the contractive Jensen operator inequality (9) with $f(x) = -\log(x)$. In this way, due to (62), we would get

$$\begin{aligned}
S(\text{Tr}_b(\rho) \| \text{Tr}_b(\sigma)) &\leq \langle \text{Tr}_b(\rho)^{1/2}, -\log(V_\rho^\dagger \Delta_{\rho,\sigma} V_\rho)(\text{Tr}_b(\rho)^{1/2}) \rangle_a \\
&\leq \langle \text{Tr}_b(\rho)^{1/2}, V_\rho^\dagger(-\log(\Delta_{\rho,\sigma}))V_\rho(\text{Tr}_b(\rho)^{1/2}) \rangle_a \\
&= \langle V_\rho \text{Tr}_b(\rho)^{1/2}, -\log(\Delta_{\rho,\sigma})V_\rho(\text{Tr}_b(\rho)^{1/2}) \rangle_{ab} \\
&= \langle \rho^{1/2}, -\log(\Delta_{\rho,\sigma})(\rho^{1/2}) \rangle_{ab} \\
&= S(\rho \| \sigma),
\end{aligned} \tag{68}$$

and so the proof of the monotonicity of S w.r.t. the partial trace Tr_b would be achieved.

For this argument to be valid, $-\log(x)$ is required to be operator convex, which is true, to be well defined at $x = 0$, and to satisfy $-\log(0) \leq 0$.

Petz circumvented the lack of definition of $-\log(x)$ in $x = 0$ by using the following integral identity

$$\begin{aligned}
\int_0^{+\infty} \left(\frac{1}{x+\xi} - \frac{1}{1+\xi} \right) d\xi &= \lim_{M \rightarrow +\infty} (\log(x+\xi) - \log(1+\xi)) \Big|_0^M = \\
&= \lim_{M \rightarrow +\infty} \log \left(\frac{x+M}{1+M} \right) - \log(x) \\
&= -\log(x).
\end{aligned} \tag{69}$$

Since the integral over $[0, +\infty)$ coincides with that over $(0, +\infty)$, we may restrict our attention to strictly positive values of ξ . If we denote by id_{ab} the identity operator on $\mathcal{B}_H(\mathcal{H}_{ab})$, we have

$$\Delta_{\rho,\sigma,\xi} \equiv (\Delta_{\rho,\sigma} + \xi id_{ab})^{-1} - (id_{ab} + \xi id_{ab})^{-1} = (\Delta_{\rho,\sigma} + \xi id_{ab})^{-1} - id_{ab}(1+\xi)^{-1}, \tag{70}$$

moreover

$$\langle \rho^{1/2}, -id_{ab}(1+\xi)^{-1} \rho^{1/2} \rangle_{ab} = -(1+\xi)^{-1} \text{Tr}(\rho) = -(1+\xi)^{-1}, \tag{71}$$

thus

$$\langle \rho^{1/2}, \Delta_{s,t,\xi} \rho^{1/2} \rangle_{ab} = \langle \rho^{1/2}, (\Delta_{\rho,\sigma} + \xi id_{ab})^{-1} \rho^{1/2} \rangle_{ab} - (1 + \xi)^{-1}. \quad (72)$$

It follows that

$$S(\rho \parallel \sigma) = \int_0^\infty \left(\langle \rho^{1/2}, (\Delta_{\rho,\sigma} + \xi id_{ab})^{-1} \rho^{1/2} \rangle_{ab} - (1 + \xi)^{-1} \right) d\xi, \quad (73)$$

and, analogously,

$$S(\text{Tr}_b(\rho) \parallel \text{Tr}_b(\sigma)) = \int_0^\infty \left(\langle \text{Tr}_b(\rho)^{1/2}, (\Delta_{\rho,\sigma}^a + \xi id_a)^{-1} \text{Tr}_b(\rho)^{1/2} \rangle_a - (1 + \xi)^{-1} \right) d\xi. \quad (74)$$

For all $\xi \in (0, +\infty)$, the function g_ξ given by $x \mapsto (x + \xi)^{-1} - (1 + \xi)^{-1}$ is well defined for $x = 0$, it is operator convex and operator monotone (decreasing), see [2] chapter V.1. Thanks to this last property we have

$$(\Delta_{\rho,\sigma}^a + \xi)^{-1} \leq (V_\rho^\dagger \Delta_{\rho,\sigma}^a V_\rho + \xi)^{-1}. \quad (75)$$

However, $g_\xi(0) = \xi^{-1} - (1 + \xi)^{-1} > 0$ for all $\xi \in (0, +\infty)$, so *the contractive Jensen operator inequality cannot be used to write*

$$(V_\rho^\dagger \Delta_{\rho,\sigma}^a V_\rho + \xi)^{-1} \leq V_\rho^\dagger (\Delta_{\rho,\sigma}^a + \xi)^{-1} V_\rho, \quad (76)$$

which would lead to the proof of the monotonicity of the relative entropy with computation analogous to those shown in formula (68).

A simple counterexample illustrating the failure of inequality (76) arises in the scalar case, where a contraction reduces to a multiplication by a real coefficient $\alpha \in (0, 1]$, and satisfies $\alpha^\dagger = \alpha$. The inequality

$$(\alpha x \alpha + \xi)^{-1} \leq \alpha (x + \xi)^{-1} \alpha, \quad (77)$$

is *false* for all $\xi \in (0, +\infty)$, as shown in Figure 1.

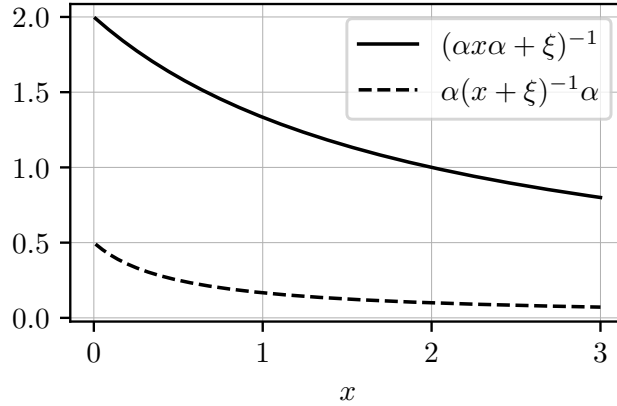


Figure 1: Comparison of $(\alpha x \alpha + \xi)^{-1}$ and $\alpha (x + \xi)^{-1} \alpha$, illustrating the failure of the Jensen-type inequality in the scalar case, with $\alpha = 0.5$ and $\xi = 0.5$.

Note that the inequality $-\log(V^\dagger X V) \leq -V^\dagger \log(X) V$ is also false for a generic contraction V , as we show in Figure 2 with a counterexample in the scalar case.

To summarize, showing that V_ρ is a contraction does not permit to use the contractive version of Jensen's operator inequality to prove the monotonicity of the relative entropy.

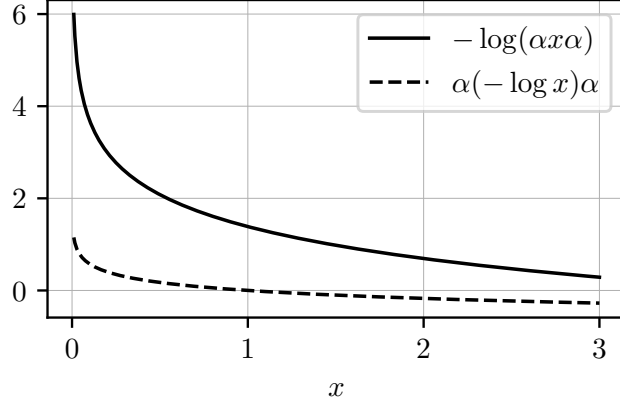


Figure 2: Counterexample showing that the inequality $-\log(\alpha x \alpha) \leq -\alpha \log(x) \alpha$ fails for $\alpha = 0.5$.

3.1 Correction of Petz's strategy to prove the relative entropy monotonicity

There is a simple correction of Petz's strategy that restores the validity of his approach to prove the monotonicity of the relative entropy in a rigorous way. The correction was provided by Petz himself, together with Nielsen, in [16]. The same line of reasoning can be found also in [24] and [9].

The key idea of the correction lies in recognizing that the operator V_ρ is not merely a contraction, but an isometry, i.e. $V_\rho^\dagger V_\rho$ is the identity operator on $\mathcal{B}_H(\mathcal{H}_a)$. In fact, using eqs. (11) and (56), we have

$$\begin{aligned}
\langle Y \text{Tr}_b(\rho)^{1/2}, V_\rho^\dagger V_\rho(X \text{Tr}_b(\rho)^{1/2}) \rangle_a &= \langle V_\rho(Y \text{Tr}_b(\rho)^{1/2}), V_\rho(X \text{Tr}_b(\rho)^{1/2}) \rangle_{ab} \\
&= \langle \text{Tr}_b^\dagger(Y) \rho^{1/2}, \text{Tr}_b^\dagger(X) \rho^{1/2} \rangle_{ab} \\
&= \text{Tr}[(Y^\dagger \otimes id_b)(X \otimes id_b) \rho] \\
&= \text{Tr}[(Y^\dagger X \otimes id_b) \rho] = \text{Tr}[Y^\dagger X \text{Tr}_b(\rho)] \quad (78) \\
&= \text{Tr}[\text{Tr}_b(\rho)^{1/2} Y^\dagger X \text{Tr}_b(\rho)^{1/2}] \\
&= \text{Tr}[(Y \text{Tr}_b(\rho)^{1/2})^\dagger X \text{Tr}_b(\rho)^{1/2}] \\
&= \langle Y \text{Tr}_b(\rho)^{1/2}, X \text{Tr}_b(\rho)^{1/2} \rangle_a.
\end{aligned}$$

This allows us to apply point (iii) of Theorem 2.1 with $\mathcal{H} = \mathcal{B}_H(\mathcal{H}_{ab})$, $\mathcal{K} = \mathcal{B}_H(\mathcal{H}_a)$ and $X = \Delta_{\rho, \sigma}$, which ensures that the inequality

$$-\log(V_s^\dagger \Delta_{\rho, \sigma} V_\rho) \leq -V_s^\dagger \log(\Delta_{\rho, \sigma}) V_\rho, \quad (79)$$

holds true, confirming the validity of the computations in (68), thereby establishing the monotonicity of the relative entropy.

3.2 Extension of Petz's proof to non-invertible density operators

The corrected version of Petz's strategy for proving the monotonicity of relative entropy can be extended to include non-invertible density operators. To this end, the interplay between the support-based and the regularized definition of relative entropy given in (27) and (40), respectively, will prove to be useful.

First of all, note that if $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$, then $S(\rho \parallel \sigma) = +\infty$, and the monotonicity statement is trivially true.

We therefore restrict our attention to the case $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, which ensures that the relative entropy is finite. Within this setting, we assume that ρ is non-invertible, while σ may or may not be invertible. This covers all cases not addressed by Petz's original analysis in [18].

To establish the monotonicity of relative entropy in this broader context, we must first ensure that $S(\text{Tr}_b(\rho) \parallel \text{Tr}_b(\sigma)) < +\infty$, i.e. that $\text{supp}(\text{Tr}_b(\rho)) \subseteq \text{supp}(\text{Tr}_b(\sigma))$. Observe that it is not necessary to check this condition when $\rho, \sigma > 0$, since in that case $\text{Tr}_b(\rho)$ and $\text{Tr}_b(\sigma)$ are positive-definite as well, and thus their supports coincide with \mathcal{H}_a .

We need a preliminary lemma regarding the kernel of positive semi-definite operators.

Lemma 3.1 *Let T_1 and T_2 be Hermitian positive semi-definite operators on a finite-dimensional Hilbert space \mathcal{H} . Then*

$$\ker(T_1 + T_2) = \ker(T_1) \cap \ker(T_2). \quad (80)$$

Proof. Proving the inclusion $\ker(T_1) \cap \ker(T_2) \subseteq \ker(T_1 + T_2)$ is trivial: if we have $x \in \ker(T_1) \cap \ker(T_2)$, then $0_{\mathcal{H}} = T_1x + T_2x = (T_1 + T_2)x$, so $x \in \ker(T_1 + T_2)$.

To show the inclusion $\ker(T_1 + T_2) \subseteq \ker(T_1) \cap \ker(T_2)$, we first notice that, thanks to the positive-semidefiniteness of T_1 and T_2 , for all $x \in \mathcal{H}$ we have

$$\langle (T_1 + T_2)x, x \rangle = \langle T_1x, x \rangle + \langle T_2x, x \rangle \geq 0, \quad (81)$$

with equality to 0 if and only if $\langle T_1x, x \rangle = \langle T_2x, x \rangle = 0$, i.e. if and only if $T_1x = T_2x = 0$ because the eigenvalues of both T_1 and T_2 are non-negative.

Hence, if $x \in \ker(T_1 + T_2)$, then $\langle (T_1 + T_2)x, x \rangle = 0$, so, due to the previous considerations, $T_1x = T_2x = 0$, and therefore $x \in \ker(T_1) \cap \ker(T_2)$. \square

Proposition 3.1 *For all density operators ρ, σ such that $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$, it holds that*

$$\text{supp}(\text{Tr}_b(\rho)) \subseteq \text{supp}(\text{Tr}_b(\sigma)). \quad (82)$$

Proof. Using the notation of section 2.1, consider again the spectral decompositions of ρ and σ

$$\rho = \sum_{j \in I_+(\rho)} \lambda_j P_j + 0P_0, \quad \sigma = \sum_{k \in I_+(\sigma)} \mu_k \Pi_k + 0\Pi_0, \quad (83)$$

where P_j and Π_k are the orthogonal projectors onto the eigenspaces corresponding to the positive eigenvalues λ_j of ρ and μ_k of σ , respectively. Then, the following operator sums yield the orthogonal projectors onto $\text{supp}(\rho)$ and $\text{supp}(\sigma)$, respectively

$$P = \sum_{j \in I_+(\rho)} P_j, \quad \Pi = \sum_{k \in I_+(\sigma)} \Pi_k. \quad (84)$$

Using the linearity and positivity of the partial trace and the fact that multiplying a positive semi-definite operator by a strictly positive scalar does not change its support, we have

$$\begin{aligned} \text{supp}(\text{Tr}_b(\rho)) &= \text{supp}\left(\sum_{j \in I_+(\rho)} \lambda_j \text{Tr}_b(P_j)\right) = \text{supp}\left(\sum_{j \in I_+(\rho)} \text{Tr}_b(P_j)\right) \\ &= \text{supp}\left(\text{Tr}_b\left(\sum_{j \in I_+(\rho)} P_j\right)\right) = \text{supp}(\text{Tr}_b(P)). \end{aligned} \quad (85)$$

The same argument applies to σ and Π , so that

$$\text{supp}(\text{Tr}_b(\sigma)) = \text{supp}(\text{Tr}_b(\Pi)). \quad (86)$$

Moreover, according to a standard result about orthogonal projectors, the inclusion $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ is equivalent to the fact that the operator $Q := \Pi - P$ is an orthogonal projector on $\text{supp}(\sigma) \cap \text{supp}(\rho)^\perp = \text{supp}(\sigma) \cap \ker(\rho)$. By the linearity of the partial trace, we can write

$$\text{Tr}_b(\Pi) = \text{Tr}_b(P) + \text{Tr}_b(Q), \quad (87)$$

moreover, using the fact that Tr_b is a positive map and Lemma 3.1, we have:

$$\begin{aligned} \text{supp}(\text{Tr}_b(\Pi)) &= \text{supp}(\text{Tr}_b(P) + \text{Tr}_b(Q)) \\ &= [\ker(\text{Tr}_b(P) + \text{Tr}_b(Q))]^\perp \\ &= [\ker(\text{Tr}_b(P)) \cap \ker(\text{Tr}_b(Q))]^\perp \\ &= \text{span}(\ker(\text{Tr}_b(P))^\perp \cup \ker(\text{Tr}_b(Q))^\perp) \\ &= \text{supp}(\text{Tr}_b(P)) + \text{supp}(\text{Tr}_b(Q)) \supseteq \text{supp}(\text{Tr}_b(P)), \end{aligned} \quad (88)$$

having used a standard property of the orthogonal complement of the intersection of two vector subspaces. Therefore,

$$\text{supp}(\text{Tr}_b(\rho)) = \text{supp}(\text{Tr}_b(P)) \subseteq \text{supp}(\text{Tr}_b(\Pi)) = \text{supp}(\text{Tr}_b(\sigma)), \quad (89)$$

as claimed. \square

While the support-based definition of relative entropy is useful to prove eq. (82), in order to prove the monotonicity of relative entropy for a non-invertible density operator ρ , the regularized definition (40) turns out to be more adequate.

Since both $\rho_\varepsilon > 0$ and $\sigma_\varepsilon > 0$ for any arbitrarily small $\varepsilon > 0$, the modular operator $\Delta_{\rho_\varepsilon, \sigma_\varepsilon}$ is well-defined. Similarly, the modular operator $\Delta_{\rho_\varepsilon, \sigma_\varepsilon}^a$ is also well-defined.

By applying the same steps used in the corrected version of Petz's proof and using equation (52), we obtain that, for any $\varepsilon > 0$

$$\begin{aligned} \text{Tr}[\text{Tr}_b(\rho_\varepsilon) \log \text{Tr}_b(\rho_\varepsilon) - \text{Tr}_b(\rho_\varepsilon) \log \text{Tr}_b(\sigma_\varepsilon)] &= \langle \text{Tr}_b(\rho_\varepsilon)^{1/2}, -\log(\Delta_{\rho_\varepsilon, \sigma_\varepsilon}^a) \text{Tr}_b(\rho_\varepsilon)^{1/2} \rangle_a \\ &\leq \langle \rho_\varepsilon^{1/2}, -\log(\Delta_{\rho_\varepsilon, \sigma_\varepsilon}) \rho_\varepsilon^{1/2} \rangle_{ab} \\ &= \text{Tr}[\rho_\varepsilon \log \rho_\varepsilon - \rho_\varepsilon \log \sigma_\varepsilon]. \end{aligned} \quad (90)$$

Notice that we can deal with $\rho_\varepsilon, \sigma_\varepsilon$ instead of ρ, σ even if they do not have unit trace because this property is used in Petz's original (and flawed) argument only in equation (71), which does not play any role in the corrected version outlined in Section 3.1.

By continuity, taking the limit $\varepsilon \rightarrow 0^+$ on both sides of the previous inequality leads to

$$\begin{aligned} S(\text{Tr}_b(\rho) || \text{Tr}_b(\sigma)) &= \lim_{\varepsilon \rightarrow 0^+} \text{Tr}[\text{Tr}_b(\rho_\varepsilon) \log \text{Tr}_b(\rho_\varepsilon) - \text{Tr}_b(\rho_\varepsilon) \log \text{Tr}_b(\sigma_\varepsilon)] \\ &\leq \lim_{\varepsilon \rightarrow 0^+} \text{Tr}[\rho_\varepsilon \log \rho_\varepsilon - \rho_\varepsilon \log \sigma_\varepsilon] = S(\rho || \sigma), \end{aligned} \quad (91)$$

which completes the proof for non-invertible ρ and general σ .

As a final remark, we note that, in [18], Petz claimed that his proof applies not only to quantum channels but also to adjoints of unital Schwarz maps. However, that claim relies on the flawed argument that we have pointed out. We have also shown that the validity of Petz's proof can be restored by proving that V_ρ is an isometry, a property that holds if we are dealing with the partial trace but not with adjoint of a general unital Schwarz map.

4 Uhlmann's proof of the monotonicity of the relative entropy under partial trace

The proof offered by Uhlmann in [25] (see also [17] for a review) was more general than that of Petz because it naturally encompassed also the case of non-invertible density operators. However,

thanks to the correction and generalization of Petz's proof outlined in the previous section, now we can state the two procedures have the same generality.

Uhlmann's proof is based on the concept of interpolations of positive sesquilinear forms, that we recall in the following subsection. Coherently with the analysis developed so-far, we will consider only the case of finite dimensional vector spaces.

4.1 Interpolations of positive sesquilinear forms

Let V be a vector space of finite dimension d over the field $\mathbb{F} = \mathbb{R}$ or \mathbb{C} , and let $\mathcal{F}(V)$ be the set of sesquilinear forms⁴ over V , assumed to be linear in the second variable and conjugate-linear in the first.

We say that $\alpha \in \mathcal{F}(V)$ is *positive* if $\alpha(v, v) \geq 0 \forall v \in V$ and we denote the space of positive sesquilinear forms over V by $\mathcal{F}_+(V)$.

We can endow $\mathcal{F}_+(V)$ with a Löwner-like partial ordering: given $\alpha, \beta \in \mathcal{F}_+(V)$, we say that $\beta \leq \alpha$ if $\alpha - \beta \geq 0$.

Fixing a basis of V , for any form $\alpha \in \mathcal{F}(V)$ there exists a unique Hermitian operator $T \in \text{End}_{\mathbb{F}}(V)$ such that for all $v, w \in V$, written in coordinates as $x, y \in \mathbb{F}^d$, one has

$$\alpha(v, w) = x^\dagger T y. \quad (92)$$

It can be easily proven that, for a positive form $\alpha \in \mathcal{F}_+(V)$, the kernel of α coincides with the *isotropic cone*

$$\ker(\alpha) = \{v \in V : \alpha(v, v) = 0\}, \quad (93)$$

and they are both equal to the kernel of the positive semi-definite operator T .

Let now \mathcal{H} be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, $h: V \rightarrow \mathcal{H}$ be a linear surjective map and $A \in \mathcal{B}_H(\mathcal{H})$. Then $\alpha \in \mathcal{F}_+(V)$ is said to be *represented by* (\mathcal{H}, h, A) , indicated with $\alpha \sim (\mathcal{H}, h, A)$, if

$$\alpha(v, w) = \langle h(v), A h(w) \rangle_{\mathcal{H}}, \quad \forall v, w \in V. \quad (94)$$

Two representations of positive sesquilinear forms $\alpha \sim (\mathcal{H}, h, A)$ and $\beta \sim (\mathcal{H}, h, B)$ are said to be *compatible* if $[A, B] = 0$. As we shall see shortly, compatibility is a key concept in constructing a functional calculus for sesquilinear forms.

The following theorem shows that compatible representations exist and its constructive proof provides a (non-unique) way to build them.

Theorem 4.1 *Let $\alpha, \beta \in \mathcal{F}_+(V)$. Then there exist representations $\alpha \sim (\mathcal{H}, h, A)$ and $\beta \sim (\mathcal{H}, h, B)$ (with the same mapping h) such that $[A, B] = 0$.*

Proof. Let $N \subset V$ be the kernel of the form $\alpha + \beta$

$$N = \{v \in V : \alpha(v, v) + \beta(v, v) = 0\}. \quad (95)$$

By fixing a basis of V , we can associate to α and β two Hermitian and positive semi-definite operators $T_1, T_2 \in \text{End}_{\mathbb{F}}(V)$ via eq. (92). It follows that

$$N = \ker(\alpha + \beta) = \ker(T_1 + T_2) = \ker(T_1) \cap \ker(T_2), \quad (96)$$

where the last equality is provided by Lemma 3.1.

Now, by setting $\mathcal{H} := V/N$ we can define the following inner product:

$$\begin{aligned} \langle \cdot, \cdot \rangle : \quad \mathcal{H} \times \mathcal{H} &\longrightarrow \mathbb{F} \\ (v + N, w + N) &\longmapsto \langle v + N, w + N \rangle = \alpha(v, w) + \beta(v, w). \end{aligned} \quad (97)$$

Note that this inner product is well-defined thanks to the equalities (96).

⁴The results of this subsection encompass also the case of *bilinear* forms.

By taking h as the (surjective) quotient map $h : V \rightarrow \mathcal{H}$, $v \mapsto h(v) = v + N$, we can define the following positive sesquilinear forms on \mathcal{H} :

$$\tilde{\alpha}(h(v), h(w)) = \tilde{\alpha}(v + N, w + N) := \alpha(v, w), \quad (98)$$

$$\tilde{\beta}(h(v), h(w)) = \tilde{\beta}(v + N, w + N) := \beta(v, w), \quad (99)$$

for all $v, w \in V$.

Thanks to the Riesz representation theorem, there exists a unique couple of positive Hermitian operators $A, B \in \mathcal{B}_H(\mathcal{H})$ such that

$$\tilde{\alpha}(h(v), h(w)) = \langle h(v), A h(w) \rangle = \alpha(v, w), \quad (100)$$

$$\tilde{\beta}(h(v), h(w)) = \langle h(v), B h(w) \rangle = \beta(v, w). \quad (101)$$

It follows that, for all $v, w \in V$:

$$\begin{aligned} \langle h(v), h(w) \rangle &= \alpha(v, w) + \beta(v, w) = \tilde{\alpha}(h(v), h(w)) + \tilde{\beta}(h(v), h(w)) \\ &= \langle h(v), (A + B) h(w) \rangle, \end{aligned} \quad (102)$$

hence $A + B = id_{\mathcal{H}}$, which implies that A and B commute. \square

Consider now $\mathbb{R}_+^2 = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$ and let J be the set of homogeneous (of degree 1), measurable and locally bounded functions $f : \mathbb{R}_+^2 \rightarrow \mathbb{R}$. It is possible to develop the concept of function of positive sesquilinear forms thanks to the following theorem, whose quite lengthy proof can be consulted in [21].

Theorem 4.2 *Let V be a vector space, $\alpha, \beta \in \mathcal{F}_+(V)$ and let $\alpha \sim (\mathcal{H}, h, A)$ and $\beta \sim (\mathcal{H}, h, B)$ be two compatible representations of α and β . Then, for any $f \in J$, the function $\gamma : V \times V \rightarrow \mathbb{F}$,*

$$\gamma(v, w) := \langle h(v), f(A, B)h(w) \rangle, \quad v, w \in V, \quad (103)$$

is a well-defined sesquilinear form on V , i.e. γ is independent of the choice of the representations and, for any given f , γ depends only on $\alpha, \beta \in \mathcal{F}_+(V)$.

Note that on the right-hand side of (103), $f(A, B)$ is intended as an operator function, which is well-defined since A and B commute and so they can be simultaneously diagonalized.

Combining Theorem 4.1 and Theorem 4.2, every $f \in J$ can be extended to a function of positive sesquilinear forms. Keeping the same symbol for simplicity, we can define

$$f : \mathcal{F}_+(V) \times \mathcal{F}_+(V) \longrightarrow \mathcal{F}(V), \quad (104)$$

as follows: given $\alpha, \beta \in \mathcal{F}_+(V)$ and any compatible representations $\alpha \sim (\mathcal{H}, h, A)$ and $\beta \sim (\mathcal{H}, h, B)$, the sesquilinear form $f(\alpha, \beta)$ is represented as

$$f(\alpha, \beta) \sim (\mathcal{H}, h, f(A, B)). \quad (105)$$

The elements of J used to define the concept of interpolation of positive sesquilinear forms are the positive functions $f^t(x, y) = x^{1-t}y^t$, where $t \in [0, 1]$. Given $\alpha, \beta \in \mathcal{F}_+(V)$, we call *interpolation from α to β* the positive sesquilinear form

$$\begin{aligned} \gamma_{\alpha \rightarrow \beta}^t &:= f^t(\alpha, \beta) : V \times V \longrightarrow \mathbb{F} \\ (v, w) &\longmapsto f^t(\alpha, \beta)(v, w), \end{aligned} \quad (106)$$

which means that for any compatible representations $\alpha \sim (\mathcal{H}, h, A)$ and $\beta \sim (\mathcal{H}, h, B)$,

$$\gamma_{\alpha \rightarrow \beta}^t(v, w) = \langle h(v), A^{1-t}B^t h(w) \rangle, \quad \forall v, w \in V. \quad (107)$$

The interpolation is said to go from α to β because, clearly, $\gamma_{\alpha \rightarrow \beta}^0 = \alpha$ and $\gamma_{\alpha \rightarrow \beta}^1 = \beta$. Moreover, the interpolation of two interpolations is another interpolation: given $t_1, t_2 \in [0, 1]$, we have

$$\begin{aligned} \gamma_{\gamma_{\alpha \rightarrow \beta}^{t_1} \rightarrow \gamma_{\alpha \rightarrow \beta}^{t_2}}^t(v, w) &= \langle h(v), (A^{1-t_1} B^{t_1})^{1-t} (A^{1-t_2} B^{t_2})^t h(w) \rangle \\ &= \langle h(v), A^{1-t_1-t+t_1t} A^{t-t_2t} B^{t_1-t_1t} B^{t_2t} h(w) \rangle \\ &= \langle h(v), A^{1-(t_1(1-t)+t_2t)} B^{t_1(1-t)+t_2t} h(w) \rangle \\ &= \gamma_{\alpha \rightarrow \beta}^{t'}(v, w), \end{aligned} \quad (108)$$

with $t' = t_1(1-t) + t_2t$.

The interpolation of α, β computed at the value $t = 1/2$ corresponds to a particularly important positive sesquilinear form, indicated with

$$\sqrt{\alpha\beta} := \gamma_{\alpha \rightarrow \beta}^{1/2}, \quad (109)$$

and called *geometric mean* of α, β . Clearly, we have

$$\sqrt{\alpha\beta}(v, w) = \gamma_{\alpha \rightarrow \beta}^{1/2}(v, w) = \langle h(v), A^{1/2} B^{1/2} h(w) \rangle, \quad \forall v, w \in V. \quad (110)$$

An important property of the geometric mean of α and β will emerge in connection with the following concept: a positive sesquilinear form $r \in \mathcal{F}_+(V)$ is said to be *dominated* by $\alpha, \beta \in \mathcal{F}_+(V)$ if

$$|r(v, w)|^2 \leq \alpha(v, v)\beta(w, w), \quad \forall v, w \in V. \quad (111)$$

The following theorem establishes that $\sqrt{\alpha\beta}$ is the ‘maximal’ positive sesquilinear form dominated by α and β .

Theorem 4.3 *Let V be a vector space and let $\alpha, \beta \in \mathcal{F}_+(V)$, then $\sqrt{\alpha\beta}$ is dominated by α, β . Moreover, interpreting $\mathcal{F}_+(V)$ as a partially ordered set, let $S \subset \mathcal{F}_+(V)$ be the subset of positive sesquilinear forms dominated by α, β , then $\sqrt{\alpha\beta} = \sup S$, i.e. $r \leq \sqrt{\alpha\beta}$ for all $r \in S$.*

Proof. The first statement is simply an application of the Cauchy-Schwarz inequality. For all $v, w \in V$ and for any compatible representations of α and β we have:

$$\begin{aligned} |\langle h(v), A^{1/2} B^{1/2} h(w) \rangle|^2 &= |\langle A^{1/2} h(v), B^{1/2} h(w) \rangle|^2 \\ &\leq \langle A^{1/2} h(v), A^{1/2} h(v) \rangle \langle B^{1/2} h(w), B^{1/2} h(w) \rangle \\ &= \langle h(v), Ah(v) \rangle \langle h(w), Bh(w) \rangle, \end{aligned} \quad (112)$$

having used the fact that A and B , and so their square roots, are Hermitian operators. Thus,

$$|\sqrt{\alpha\beta}(v, w)|^2 \leq \alpha(v, v)\beta(w, w), \quad \forall v, w \in V, \quad (113)$$

so $\sqrt{\alpha\beta}$ is dominated by α and β .

To prove the second statement, consider a form $r \in \mathcal{F}_+(V)$ dominated by α, β and let $\alpha \sim (\mathcal{H}, h, A)$ and $\beta \sim (\mathcal{H}, h, B)$ be compatible representations. By applying the constructive proof of Theorem 4.1 to r , we can find a positive operator $C \in \mathcal{B}_H(\mathcal{H})$, such that $r(v, w) = \langle h(v), Ch(w) \rangle$, however the representation $r \sim (\mathcal{H}, h, C)$ will not, in general, be compatible with those of α and β . Since r is dominated by α and β , we have

$$|\langle h(v), Ch(w) \rangle|^2 \leq \langle h(v), Ah(v) \rangle \langle h(w), Bh(w) \rangle, \quad \forall v, w \in V, \quad (114)$$

or, thanks to the fact that h is surjective on \mathcal{H} ,

$$|\langle x, Cy \rangle|^2 \leq \langle x, Ax \rangle \langle y, By \rangle, \quad \forall x, y \in \mathcal{H}. \quad (115)$$

Now, the second statement of the theorem, i.e. $r \leq \sqrt{\alpha\beta}$, means that $(\sqrt{\alpha\beta} - r)(v, v) \geq 0$ for all $v \in V$, and all $r \in S$, which is equivalent to $\langle u, (A^{1/2} B^{1/2} - C)u \rangle \geq 0$ for all $u \in \mathcal{H}$ and all C satisfying inequality (115).

It follows that the second statement of the theorem will be proven if we manage to show that $C \leq A^{1/2}B^{1/2}$. In order to obtain this result, a regularization procedure applied on the operators A, B is helpful: since $A \geq 0$ and $B \geq 0$, for all $\varepsilon > 0$, $A_\varepsilon := A + \varepsilon id_{\mathcal{H}}$, $B_\varepsilon := B + \varepsilon id_{\mathcal{H}}$ and their square roots $A_\varepsilon^{1/2}$, $B_\varepsilon^{1/2}$ are Hermitian, positive and invertible. Moreover, it is clear that $A \leq A_\varepsilon$, $B \leq B_\varepsilon$, hence

$$A_\varepsilon^{-1/2}AA_\varepsilon^{-1/2} \leq id_{\mathcal{H}}, \quad B_\varepsilon^{-1/2}BB_\varepsilon^{-1/2} \leq id_{\mathcal{H}}. \quad (116)$$

Finally, A_ε and B_ε are monotonically decreasing in ε w.r.t. the Löwner ordering and $A_\varepsilon \rightarrow A$ and $B_\varepsilon \rightarrow B$, as $\varepsilon \rightarrow 0$. For all $x, y \in \mathcal{H}$, there exist unique vectors $u, v \in \mathcal{H}$ such that

$$x := A_\varepsilon^{-1/2}u, \quad y := B_\varepsilon^{-1/2}v, \quad (117)$$

then, inequality (115) becomes

$$|\langle A_\varepsilon^{-1/2}u, CB_\varepsilon^{-1/2}v \rangle|^2 \leq \langle A_\varepsilon^{-1/2}u, AA_\varepsilon^{-1/2}u \rangle \langle B_\varepsilon^{-1/2}v, BB_\varepsilon^{-1/2}v \rangle, \quad (118)$$

i.e.

$$|\langle u, A_\varepsilon^{-1/2}CB_\varepsilon^{-1/2}v \rangle|^2 \leq \langle u, A_\varepsilon^{-1/2}AA_\varepsilon^{-1/2}u \rangle \langle v, B_\varepsilon^{-1/2}BB_\varepsilon^{-1/2}v \rangle \leq \langle u, u \rangle \langle v, v \rangle, \quad (119)$$

having used the inequalities written in (116).

By considering $u = v$, and taking into account that $A_\varepsilon^{-1/2}CB_\varepsilon^{-1/2}$ is a positive operator, we can write

$$\langle u, A_\varepsilon^{-1/2}CB_\varepsilon^{-1/2}u \rangle \leq \langle u, u \rangle, \quad \forall u \in \mathcal{H}, \quad (120)$$

which implies that $A_\varepsilon^{-1/2}CB_\varepsilon^{-1/2} \leq id_{\mathcal{H}}$, thus $C \leq A_\varepsilon^{1/2}B_\varepsilon^{1/2}$. By taking the limit $\varepsilon \rightarrow 0$, we get $C \leq A^{1/2}B^{1/2}$ and so also the second statement of the theorem is proven. \square

The property of the geometric mean just proven allows us to extend the ordering relation between two positive sesquilinear forms to their interpolations, in the sense specified by the following theorem.

Theorem 4.4 *Let V be a vector space and let $\alpha, \alpha', \beta, \beta' \in \mathcal{F}_+(V)$ such that $\alpha' \leq \alpha$ and $\beta' \leq \beta$, then*

$$\gamma_{\alpha' \rightarrow \beta'}^t \leq \gamma_{\alpha \rightarrow \beta}^t, \quad \forall t \in [0, 1]. \quad (121)$$

Proof. The statement is clearly satisfied for $t = 0, 1$. Setting $t = 1/2$, thanks to the first part of Theorem 4.3, we get

$$|\gamma_{\alpha' \rightarrow \beta'}^{1/2}(v, w)|^2 \leq \alpha'(v, v)\beta'(w, w) \leq \alpha(v, v)\beta(w, w), \quad \forall v, w \in V, \quad (122)$$

where the second inequality follows from the hypotheses of this theorem. This means that $\gamma_{\alpha' \rightarrow \beta'}^{1/2}$ is dominated by α, β and so the extremality of the geometric mean implies:

$$\gamma_{\alpha' \rightarrow \beta'}^{1/2} \leq \gamma_{\alpha \rightarrow \beta}^{1/2}. \quad (123)$$

If we now use eq. (108) with $t = t_2 = 1/2$ and $t_1 = 0$, we get

$$\gamma_{\alpha \rightarrow \beta}^{1/4} = \gamma_{\gamma_{\alpha \rightarrow \beta}^0 \rightarrow \gamma_{\alpha \rightarrow \beta}^{1/2}}^{1/2}, \quad \gamma_{\alpha' \rightarrow \beta'}^{1/4} = \gamma_{\gamma_{\alpha' \rightarrow \beta'}^0 \rightarrow \gamma_{\alpha' \rightarrow \beta'}^{1/2}}^{1/2}. \quad (124)$$

By repeating the previous argument, this time using eq. (123), we show that $\gamma_{\alpha' \rightarrow \beta'}^{1/4} \leq \gamma_{\alpha \rightarrow \beta}^{1/4}$.

By iterating this procedure, we can prove the statement of the theorem for any $t \in [0, 1]$ of the type $t_{k,n} = k/2^n$, with $n, k \in \mathbb{N}$, $k \leq 2^n$, which is a dense subset of $[0, 1]$.

Finally, the functions $t_{k,n} \rightarrow \gamma_{\alpha \rightarrow \beta}^{t_{k,n}}(v, w)$ and $t_{k,n} \rightarrow \gamma_{\alpha' \rightarrow \beta'}^{t_{k,n}}(v, w)$ are continuous for every fixed $v, w \in V$, hence the theorem holds for all $t \in [0, 1]$. \square

In a similar way, we can prove another important result. Let $\psi : U \rightarrow V$ be a linear map between vector spaces and let $\alpha, \beta \in \mathcal{F}_+(V)$. Then, ψ allows us to *pull-back* these sesquilinear forms on U as follows

$$\begin{aligned} \psi^* \alpha : U \times U &\longrightarrow \mathbb{F} & \psi^* \beta : U \times U &\longrightarrow \mathbb{F} \\ (v, w) &\longmapsto \alpha(\psi(v), \psi(w)) & (v, w) &\longmapsto \beta(\psi(v), \psi(w)) \end{aligned} \quad (125)$$

The following theorem shows that the pull-back of an interpolation of forms in $\mathcal{F}_+(V)$ is always ‘smaller’ than the interpolation of their pull-backs, with respect to the partial ordering of $\mathcal{F}_+(U)$.

Theorem 4.5 *Let U, V be vector spaces, $\psi : U \rightarrow V$ be a linear map and $\alpha, \beta \in \mathcal{F}_+(V)$, then*

$$\psi^* \gamma_{\alpha \rightarrow \beta}^t \leq \gamma_{\psi^* \alpha \rightarrow \psi^* \beta}^t, \quad \forall t \in [0, 1]. \quad (126)$$

Proof. The argument that we use is quite similar to the one appearing in the previous proof. The statement is true for $t = 0$ and $t = 1$ because in these cases we have $\psi^* \alpha \leq \psi^* \alpha$ and $\psi^* \beta \leq \psi^* \beta$, respectively.

Let us now consider $t = 1/2$, which gives rise to the geometric mean, so

$$|\psi^* \sqrt{\alpha \beta}(v, w)|^2 = |\sqrt{\alpha \beta}(\psi(v), \psi(w))|^2, \quad \forall v, w \in U. \quad (127)$$

Since $\sqrt{\alpha \beta}$ is dominated by α, β , we can write

$$|\psi^* \sqrt{\alpha \beta}(v, w)|^2 \leq \alpha(\psi(v), \psi(v)) \beta(\psi(w), \psi(w)) = \psi^* \alpha(v, v) \psi^* \beta(w, w), \quad (128)$$

which shows that $\psi^* \sqrt{\alpha \beta}$ is dominated by $\psi^* \alpha, \psi^* \beta$. Thus, by the extremal property of the geometric mean, the statement of the theorem holds for $t = 1/2$.

By iterating this reasoning as done in the proof of the previous theorem, the validity of inequality (126) can be generalized to all $t \in [0, 1]$. \square

4.2 Definition of the relative entropy in terms of interpolations of forms

In this subsection, we apply the results previously established to reformulate the relative entropy in a manner that will facilitate the proof of its monotonicity under partial trace, to be presented in the next subsection.

Adopting analogous notations as those introduced at the beginning of Section 3, we identify the vector space V , on which the forms of interest for us will be defined, with $\mathcal{B}(\mathcal{H}_{ab})$. As we know, this is a Hilbert space w.r.t. the Hilbert-Schmidt inner product $\langle A, B \rangle_{ab} = \text{Tr}(A^\dagger B)$, $A, B \in \mathcal{B}(\mathcal{H}_{ab})$, which is a positive-definite sesquilinear form.

Given two density operators $\rho, \sigma \in \mathcal{B}_H(\mathcal{H}_{ab})$, we can define the following positive sesquilinear forms $\rho_L, \sigma_R : \mathcal{B}(\mathcal{H}_{ab}) \times \mathcal{B}(\mathcal{H}_{ab}) \rightarrow \mathbb{F}$,

$$\rho_L(A, B) := \text{Tr}(\rho B A^\dagger) = \text{Tr}(A^\dagger \rho B) = \langle A, L_\rho B \rangle, \quad (129)$$

$$\sigma_R(A, B) := \text{Tr}(\sigma A^\dagger B) = \text{Tr}(A^\dagger B \sigma) = \langle A, R_\sigma B \rangle, \quad (130)$$

where the operators L_ρ and R_σ are defined as in eq. (46). We immediately recognize the two representations

$$\rho_L \sim (\mathcal{B}(\mathcal{H}_{ab}), id_{ab}, L_\rho), \quad \sigma_R \sim (\mathcal{B}(\mathcal{H}_{ab}), id_{ab}, R_\sigma), \quad (131)$$

where id_{ab} is the identity map on $\mathcal{B}(\mathcal{H}_{ab})$. These representations are compatible because, thanks to eq. (47), $[L_\rho, R_\sigma] = 0$.

We can now define the *relative entropy positive sesquilinear form* between the two density operators ρ, σ , indicated with $S_{\rho||\sigma} : \mathcal{B}(\mathcal{H}_{ab}) \times \mathcal{B}(\mathcal{H}_{ab}) \rightarrow \mathbb{F}$, as the rate of change of the interpolation $\gamma_{\rho_L \rightarrow \sigma_R}^t$ with respect to ρ_L

$$S_{\rho||\sigma}(A, B) = - \liminf_{t \rightarrow 0^+} \frac{\gamma_{\rho_L \rightarrow \sigma_R}^t(A, B) - \rho_L(A, B)}{t}, \quad A, B \in \mathcal{B}(\mathcal{H}_{ab}). \quad (132)$$

Remark. The use of \liminf instead of an ordinary limit is motivated by the following two arguments:

1. When ρ and σ are not invertible, the interpolation function $t \mapsto \gamma_{\rho_L \rightarrow \sigma_R}^t(A, B)$ may not be differentiable at $t = 0^+$. The ordinary limit of the differential quotient at $t = 0^+$ might fail to exist due to oscillations or lack of smoothness in the interpolation path. In such situations, the ordinary limit does not exist. However, the \liminf always exists (possibly infinite), thereby ensuring that the entropy form $S_{\rho\|\sigma}(A, B)$ is always well-defined.
2. The function $t \mapsto \gamma_{\rho_L \rightarrow \sigma_R}^t(A, B)$ is convex in t , as it arises from the interpolation $f_t(x, y) = x^{1-t}y^t$, which is jointly operator convex for $x, y > 0$. For convex functions, the left and right derivatives at an endpoint may differ, and the correct notion of derivative from the right at $t = 0$ is the lower right Dini derivative, i.e. the \liminf of the difference quotient. Thus, the use of \liminf aligns with standard practice in convex analysis.

The relative entropy between states ρ, σ can be recovered as follows.

Theorem 4.6 *For all density operators $\rho, \sigma \in \mathcal{B}_H(\mathcal{H}_{ab})$ we have*

$$S(\rho\|\sigma) = S_{\rho\|\sigma}(id_{ab}, id_{ab}). \quad (133)$$

Proof. Let us first consider the case of invertible density operators $\rho > 0$ and $\sigma > 0$. Then,

$$\begin{aligned}
S_{\rho\|\sigma}(id_{ab}, id_{ab}) &= -\liminf_{t \rightarrow 0^+} \frac{\gamma_{\rho_L \rightarrow \sigma_R}^t(id_{ab}, id_{ab}) - \rho_L(id_{ab}, id_{ab})}{t} \\
&= -\liminf_{t \rightarrow 0^+} \frac{\langle id_{ab}, L_\rho^{1-t} R_\sigma^t id_{ab} \rangle - \langle id_{ab}, L_\rho id_{ab} \rangle}{t} \\
&= -\liminf_{t \rightarrow 0^+} \frac{\text{Tr}(\rho^{1-t} \sigma^t) - \text{Tr}(\rho)}{t} = -\left. \frac{d}{dt} \right|_{t=0} \text{Tr}(\rho^{1-t} \sigma^t) \\
&= -\left. \frac{d}{dt} \right|_{t=0} \text{Tr}(\exp((1-t)\log \rho) \exp(t\log \sigma)) \\
&= -\text{Tr} \left[-\exp((1-t)\log \rho) \log \rho \exp(t\log \sigma) \right. \\
&\quad \left. + \exp((1-t)\log \rho) \exp(t\log \sigma) \log \sigma \right]_{t=0} \\
&= -\text{Tr}(-\rho \log \rho + \rho \log \sigma) = \text{Tr}(\rho \log \rho - \rho \log \sigma) \\
&= S(\rho\|\sigma).
\end{aligned} \quad (134)$$

In the case of σ and ρ that are not invertible, we can use the regularized version of relative entropy. We have seen an equivalent definition of the relative entropy in (40) and then it is natural to consider

$$\lim_{\epsilon \rightarrow 0^+} S_{\rho_\epsilon\|\sigma_\epsilon}(id_{ab}, id_{ab}). \quad (135)$$

Now we have that both $\rho_\epsilon, \sigma_\epsilon$ are positive definite and, thus, $\log \rho_\epsilon, \log \sigma_\epsilon$ are well defined. By repeating the steps in (134), we get that

$$\lim_{\epsilon \rightarrow 0^+} S_{\rho_\epsilon\|\sigma_\epsilon}(id_{ab}, id_{ab}) = \lim_{\epsilon \rightarrow 0^+} \text{Tr}(\rho_\epsilon \log \rho_\epsilon - \rho_\epsilon \log \sigma_\epsilon). \quad (136)$$

Obtaining the same definition of relative entropy as in (27). \square

4.3 Proof of the monotonicity of the relative entropy under partial trace

Thanks to formula (133), the monotonicity of the relative entropy under the partial trace $\text{Tr}_b : \mathcal{B}(\mathcal{H}_{ab}) \rightarrow \mathcal{B}(\mathcal{H}_a)$ will be proven if we show that

$$S_{\text{Tr}_b(\rho)\|\text{Tr}_b(\sigma)}(id_a, id_a) \leq S_{\rho\|\sigma}(id_{ab}, id_{ab}). \quad (137)$$

As in Petz's proof, the adjoint of the partial trace plays a central, though conceptually distinct, role in establishing the monotonicity of relative entropy.

Specifically, in Uhlmann's approach, Tr_b^\dagger acts as a pull-back map: we define $\psi := \text{Tr}_b^\dagger : \mathcal{B}(\mathcal{H}_a) \rightarrow \mathcal{B}(\mathcal{H}_{ab})$ and use it to pull back the positive sesquilinear forms ρ_L and σ_R introduced in Eqs. (129) and (130), respectively. Thanks to Theorem 4.5 we have

$$\psi^* \gamma_{\rho_L \rightarrow \sigma_R}^t \leq \gamma_{\psi^* \rho_L \rightarrow \psi^* \sigma_R}^t, \quad (138)$$

i.e.

$$\gamma_{\rho_L \rightarrow \sigma_R}^t(\text{Tr}_b^\dagger(X), \text{Tr}_b^\dagger(X)) = \psi^* \gamma_{\rho_L \rightarrow \sigma_R}^t(X, X) \leq \gamma_{\psi^* \rho_L \rightarrow \psi^* \sigma_R}^t(X, X), \quad (139)$$

for all $X \in \mathcal{B}(\mathcal{H}_a)$. Now we can use the fact that Tr_b preserves density operators and that Tr_b^\dagger is a Schwarz map to write

$$\begin{aligned} \psi^* \rho_L(X, X) &= \rho_L(\text{Tr}_b^\dagger(X), \text{Tr}_b^\dagger(X)) = \text{Tr}(\text{Tr}_b^\dagger(X)^\dagger \rho \text{Tr}_b^\dagger(X)) = \text{Tr}(\rho \text{Tr}_b^\dagger(X) \text{Tr}_b^\dagger(X)^\dagger) \\ &\leq \text{Tr}(\rho \text{Tr}_b^\dagger(X X^\dagger)) = \langle \rho, \text{Tr}_b^\dagger(X X^\dagger) \rangle = \langle \text{Tr}_b(\rho), X X^\dagger \rangle = \text{Tr}(\text{Tr}_b(\rho) X X^\dagger) \\ &= \text{Tr}_b(\rho)_L(X, X). \end{aligned} \quad (140)$$

Replacing ρ_L with σ_R , we find

$$\psi^* \sigma_R(X, X) \leq \text{Tr}_b(\sigma)_R(X, X). \quad (141)$$

Thus, we have proven that $\psi^* \rho_L \leq \text{Tr}_b(\rho)_L$ and $\psi^* \sigma_R \leq \text{Tr}_b(\sigma)_R$ and so Theorem 4.4 implies that, for all $t \in [0, 1]$, it holds that

$$\gamma_{\psi^* \rho_L \rightarrow \psi^* \sigma_R}^t \leq \gamma_{\text{Tr}_b(\rho)_L \rightarrow \text{Tr}_b(\sigma)_R}^t. \quad (142)$$

This result and inequality (139) imply

$$\gamma_{\rho_L \rightarrow \sigma_R}^t(\text{Tr}_b^\dagger(X), \text{Tr}_b^\dagger(X)) \leq \gamma_{\text{Tr}_b(\rho)_L \rightarrow \text{Tr}_b(\sigma)_R}^t(X, X), \quad (143)$$

for all $X \in \mathcal{B}(\mathcal{H}_a)$. By considering in particular $X = id_a$ and recalling that Tr_b^\dagger is a unital map, i.e. $\text{Tr}_b^\dagger(id_a) = id_{ab}$, we obtain

$$\gamma_{\rho_L \rightarrow \sigma_R}^t(id_{ab}, id_{ab}) \leq \gamma_{\text{Tr}_b(\rho)_L \rightarrow \text{Tr}_b(\sigma)_R}^t(id_a, id_a). \quad (144)$$

Now, since Tr_b is trace-preserving, we have

$$\begin{aligned} \text{Tr}_b(\rho)_L(id_a, id_a) &= \langle id_a, L_{\text{Tr}_b(\rho)} id_a \rangle_a = \text{Tr}(\text{Tr}_b(\rho)) = \text{Tr}(\rho) = \langle id_{ab}, L_\rho id_{ab} \rangle_{ab} \\ &= \rho_L(id_{ab}, id_{ab}), \end{aligned} \quad (145)$$

thus, we can rewrite inequality (144) as follows

$$\gamma_{\rho_L \rightarrow \sigma_R}^t(id_{ab}, id_{ab}) - \rho_L(id_{ab}, id_{ab}) \leq \gamma_{\text{Tr}_b(\rho)_L \rightarrow \text{Tr}_b(\sigma)_R}^t(id_a, id_a) - \text{Tr}_b(\rho)_L(id_a, id_a). \quad (146)$$

Thanks to eqs. (133) and (132), this last inequality implies the monotonicity of the relative entropy under partial trace.

5 Conclusions

We have revisited the monotonicity of relative entropy under the action of quantum channels by focusing on two important proofs: those by Petz and Uhlmann. While both approaches are foundational, their complexity has often hindered their pedagogical dissemination.

Our aim was to clarify and reconstruct these strategies within a finite-dimensional operator framework. In particular, we pointed out a subtle flaw in Petz's original argument, whose validity was nonetheless restored by Petz and Nielsen soon after, and showed how to rigorously extend this approach to incorporate non-invertible density operators.

It is also worth noting that our explicit construction of the isometric operator defined in equation (60) sheds new light on its structural role within the broader context of quantum information theory. In particular, this operator can be seen as an essential component of the *Petz recovery map*, originally introduced in [20] in the setting of von Neumann algebras. Given a quantum channel \mathcal{C} and a fixed full-rank state $\sigma \in \mathcal{B}_H(\mathcal{H}_{ab})$, the Petz recovery map associated to \mathcal{C} and σ is defined by

$$\mathcal{P}_{\sigma, \mathcal{C}}(\rho) := \sigma^{1/2} \mathcal{C}^\dagger \left(\mathcal{C}(\sigma)^{-1/2} \rho \mathcal{C}(\sigma)^{-1/2} \right) \sigma^{1/2}, \quad (147)$$

or equivalently, in terms of superoperators,

$$\mathcal{P}_{\sigma, \mathcal{C}} = L_{\sigma^{1/2}} \circ R_{\sigma^{1/2}} \circ \mathcal{C}^\dagger \circ R_{\mathcal{C}(\sigma)^{-1/2}}^a \circ L_{\mathcal{C}(\sigma)^{-1/2}}^a. \quad (148)$$

When $\mathcal{C} = \text{Tr}_b$, this reduces to

$$\begin{aligned} \mathcal{P}_{\sigma, \text{Tr}_b} &= L_{\sigma^{1/2}} \circ (R_{\sigma^{1/2}} \circ \text{Tr}_b^\dagger \circ R_{\text{Tr}_b(\sigma)^{-1/2}}^a) \circ L_{\text{Tr}_b(\sigma)^{-1/2}}^a \\ &= L_{\sigma^{1/2}} \circ V_\sigma \circ L_{\text{Tr}_b(\sigma)^{-1/2}}^a, \end{aligned} \quad (149)$$

or

$$V_\sigma = L_{\sigma^{-1/2}} \circ \mathcal{P}_{\sigma, \text{Tr}_b} \circ L_{\text{Tr}_b(\sigma)^{1/2}}^a. \quad (150)$$

So, just as the Petz recovery map characterizes the reversibility of quantum channels and identifies conditions for saturation of the monotonicity inequality, the operator V_σ captures explicitly the mechanism by which relative entropy is contracted under partial trace.

In recent developments, particularly in the work of Fawzi and Renner [5], the Petz recovery map plays a central role in quantitative refinements of the data processing inequality. Specifically, for states ρ, σ and a channel \mathcal{C} , the inequality

$$S(\rho \parallel \sigma) - S(\mathcal{C}(\rho) \parallel \mathcal{C}(\sigma)) \geq -2 \log F(\rho, (\mathcal{P}_{\sigma, \mathcal{C}} \circ \mathcal{C})(\rho)), \quad (151)$$

bounds the loss of distinguishability in terms of the fidelity F between the original state ρ and its recovered approximation via $\mathcal{P}_{\sigma, \mathcal{C}} \circ \mathcal{C}$.

The explicit identification of V_σ offers a concrete realization of this recovery mechanism, reinforcing its interpretive clarity and suggesting further applications in entropy inequalities and recoverability conditions.

Acknowledgments

The authors are grateful to Prof. Mark M. Wilde for his valuable insights on the literature concerning relative entropy and, in particular, for clarifying the correction of the flawed Petz argument through the proof that the operator V_ρ is an isometry.

References

- [1] Huzihiro Araki. Relative entropy of states of von Neumann algebras. *Publ. Res. Inst. Math. Sci.*, 11(3):809–833, 1976.
- [2] Rajendra Bhatia. *Matrix Analysis*. Springer, New York, NY, 1997.
- [3] Man-Duen Choi. A Schwarz inequality for positive linear maps on C^* -algebras. *Illinois Journal of Mathematics*, 18(4):565–574, 1974.
- [4] Nilanjana Datta and Mark M Wilde. Quantum markov chains, sufficiency of quantum channels, and rényi information measures. *Journal of Physics A: Mathematical and Theoretical*, 48(50):505301, 2015.

- [5] Omar Fawzi and Renato Renner. Quantum conditional mutual information and approximate markov chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015.
- [6] Frank Hansen and Gert K. Pedersen. Jensen’s operator inequality. *Bulletin of the London Mathematical Society*, 35(04):553–564, June 2003.
- [7] Teiko Heinosaari and Mário Ziman. *The mathematical language of quantum theory: from uncertainty to entanglement*. Cambridge University Press, 2011.
- [8] Marius Junge, Renato Renner, David Sutter, Mark M Wilde, and Andreas Winter. Universal recovery maps and approximate sufficiency of quantum relative entropy. In *Annales Henri Poincaré*, volume 19, pages 2955–2978. Springer, 2018.
- [9] Sumeet Khatri and Mark M. Wilde. Principles of quantum communication theory: A modern approach. arXiv preprint arXiv:2011.04672 [quant-ph], 2024. Preliminary version of the book.
- [10] O. E. Lanford and D. W. Robinson. Mean entropy of states in quantum statistical mechanics. *Journal of Mathematical Physics*, 9(7):1120–1125, 1968.
- [11] Elliott H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11(3):267–288, 1973.
- [12] Elliott H. Lieb and Mary Beth Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14(12):1938–1941, 1973.
- [13] Goran Lindblad. Completely positive maps and entropy inequalities. *Communications in Mathematical Physics*, 40(2):147–151, 1975.
- [14] V. Moretti. *Spectral theory and quantum mechanics*, volume 64. Springer, 2013.
- [15] Alexander G. Müller-Hermes and David Reeb. Monotonicity of the quantum relative entropy under positive maps. *Annales Henri Poincaré*, 18(5):1777–1788, 2017.
- [16] Michael A. Nielsen and Dénes Petz. A simple proof of the strong subadditivity inequality. *Quantum Information and Computation*, 5(6):480–486.
- [17] Juan Manuel Pérez-Pardo. On Uhlmann’s proof of the monotonicity of the relative entropy. In *Particles, Fields and Topology: Celebrating AP Balachandran*, pages 145–155. World Scientific, 2023.
- [18] D. Petz. Monotonicity of quantum relative entropy revisited. *Reviews in Mathematical Physics*, 15(01):79–91, 2003. arXiv:quant-ph/0209053 [quant-ph].
- [19] Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von neumann algebra. *Communications in Mathematical Physics*, 105(1):123–131, 1986.
- [20] Dénes Petz. Sufficiency of channels over von neumann algebras. *The Quarterly Journal of Mathematics*, 39(1):97–108, 1988.
- [21] Władysław Pusz and Stanisław L. Woronowicz. Functional calculus for sesquilinear forms and the purification map. *Reports on Mathematical Physics*, 8(2):159–170, 1975.
- [22] Naresh Sharma. More on a trace inequality in quantum information theory. *arXiv preprint arXiv:1512.00226*, 2015.
- [23] W Forrest Stinespring. Positive functions on C^* -algebras. *Proceedings of the American Mathematical Society*, 6(2):211–216, 1955.
- [24] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009. Published December 2009.

- [25] Armin Uhlmann. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21–32, 1977.
- [26] Hisaharu Umegaki. Conditional expectation in an operator algebra, IV (entropy and information). In *Kodai Mathematical Seminar Reports*, volume 14, pages 59–85. Department of Mathematics, Tokyo Institute of Technology, 1962.
- [27] Vlatko Vedral. The role of relative entropy in quantum information theory. *Reviews of Modern Physics*, 74(1):197, 2002.
- [28] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, Cambridge, UK, 2 edition, 2017.
- [29] Lin 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.