Quantum Data Processing

Rudolf Ahlswede and Peter Löber

Abstract—**We prove a data processing inequality for quantum communication channels, which states that processing a received quantum state may never increase the mutual information between input and output states.**

Index Terms—**Data processing, quantum information.**

I. Introduction

Our starting point was the question whether Holevo's upper bound could be viewed as a special case of a more general theorem, as a kind of data processing inequality, which has been known for classical channels for a long time. This inequality is about mutual information

$$I(P;W) = H(PW) - H(PW \,|\, P)$$

with W a channel (a stochastic matrix) and P an input process (a probability distribution on the channel's domain). Here, H denotes the (Shannon) entropy, respectively, conditional entropy (not to be confused with relative entropy).[1]

It is one of those important theorems from classical information theory that this purely algebraically defined mutual information gives the capacity of the memoryless channel, just by taking the supremum over all input processes [16]. One might interprete this number as the information that is shared by the input process P and the output process PW.

The concept of mutual information can easily be adapted for quantum channels, which gives the "number" that is known as the Holevo bound for classical communication over a quantum channel [8], and the bound can be achieved [7], [9], [15].

It is important to notice that the quantum version of mutual information does not use measurements at all (!). It is an algebraically defined number that depends only on the quantum channel and an input state (cf. Definition VI.2). If one interpretes this number again as the information that is shared by an input state A and the corresponding output state $W_*(A)$ of the quantum channel W_*, our main theorem—the quantum data processing inequality—may be formulated as follows.

Processing the channel's output state by any physically allowed transformation D_* (one may interprete this as another quantum channel, as well) can never increase the information that is shared by the input and the output state

$$I(A; D_* \circ W_*) \leq I(A; W_*).$$

We soon realized that a quantum data processing inequality is a corollary of Uhlmann's monotonicity theorem (see Section V) which is a generalization of a theorem of Lieb (see [11] or [1]). Here, as in the proof of the classical data processing inequality (which plays around with conditional mutual information[2]), convexity properties play a central role.

Root functions $(z \mapsto z^\mu; 0 < \mu \leq 1)$ are Pick functions (cf. Section II) and, therefore, operator monotone (cf. Section III). With the notation from Section IV we can use this fact to prove Uhlmann's monotonicity theorem (Section V). Finally (Section VI), we directly derive the quantum data processing inequality for quantum mutual information, and we present Holevo's bound as its (important) corollary.

For the special but essential class of quantum operations (physical maps) that are restrictions on physical subsystems Schumacher, Westmoreland, and Wootters established an inequality for a *Holevo bound functional* [14]. This leads to the quantum data processing inequality, too, because one gets quantum mutual information just by applying this functional to the respective output states of the channel. Later, Horodecki derived from this inequality for the Holevo bound functional[3] another bound for source coding (compression) of quantum information [10].

Indeed, we concentrated on the mutual information itself which is a very general concept of information theory with its roots in classical information theory. We refer the reader to *coherent information* which was discussed by Schumacher and Nielsen for error correction [13] as a very particular kind of quantum (!) information.

II. Pick Functions

This section introduces the notion of Pick functions. We will see in the next section that Pick functions are "operator monotone" (cf. Corollary III.4 for a rigorous formulation), a fact that is very useful because it is sometimes quite easy to decide whether a function is a Pick function or not. The theory we introduce here is developed in great detail (and with complete proofs) in [5] and [1].

Definition II.1: $H^\pm \triangleq \{x + iy \in \mathbb{C} \,|\, y \gtrless 0\}$ denote the two half spaces, respectively. $P \triangleq \{\varphi : H^+ \rightarrow H^+ \text{analytic}\}$ denotes the set of *Pick functions*.

Remark II.2: P is a convex cone, and if $f, g \in P$ then $g \circ f \in P$, too.

Example II.3: The function $\varphi(z) \triangleq z^\mu$ with $0 < \mu \leq 1$ is in P. A function

$$\psi(z) \triangleq \alpha z + \beta + \sum_{i=1}^{m} \frac{\gamma_i}{\delta_i - z}$$

with $\alpha, \gamma_i > 0$ and $\beta, \delta_i \in \mathbb{R}$ is in P, too.

The next theorem shows that the latter examples give essentially (i.e., up to closure) all Pick functions.

Theorem II.4: Every Pick function $\varphi \in P$ has a (unique) representation

$$\varphi(z) \triangleq \alpha z + \beta + \int \left(\frac{1}{x - z} - \frac{x}{x^2 + 1} \right) d\mu(x) \qquad (1)$$

with $\alpha \geq 0$, $\beta \in \mathbb{R}$, and μ a positive Borel measure on $\mathbb{R}$ for which $\int (x^2 + 1)^{-1} d\mu(x) < \infty$.

This theorem is from [5, p. 20 ff]. Its proof transforms the Pick function with an appropriate Möbius transformation (and its inverse) to a function with a positive real part that maps the unit disk into itself. Its real part is a positive harmonic function.

Manuscript received July 27, 1999; revised March 24, 2000. The material in this correspondence was presented in part at the Winter School on Coding and Information Theory, Denmark, December 13–16, 1998.

R. Ahlswede is with the Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany (e-mail: ahlswede@mathematik.uni-bielefeld.de).

P. Löber was with the Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany. He is now with Bull GmbH, GB Telco, 63225 Langen, Germany (e-mail: loeber@swh600.langen.bull.de).

Communicated by A. Barg, Associate Editor for Coding Theory.

Publisher Item Identifier S 0018-9448(01)00834-3.

[1] In Section VI we give concrete definitions for quantum channels, and they will be exact replicates of those for classical channels.

[2] cf. [3, p. 55] or [2, p. 32].

[3] He proved it with Uhlmann's monotonicity theorem, too!.

Definition II.5: For an open interval $(a,b) \subseteq \mathbb{R}$ let

$$\mathbb{C}(a,b) \triangleq H^+ \cup H^- \cup (a,b)$$

and

$$P(a,b) \triangleq \{\varphi : \mathbb{C}(a,b) \to \mathbb{C} : \varphi|_{H+} \in P \wedge \varphi(\bar{\cdot}) = \overline{\varphi(\cdot)}\}.$$

Remark II.6: $P(a,b)$ is a convex cone. Moreover, $\varphi \in P(a,b)$ implies that $\varphi|_{(a,b)}$ is a monotonically increasing real function. (As ϕ maps H^+ into H^+ and H^- into H^- it has to be real on the real axis. Let now $\varphi = u + iv$ and $z = x + iy$. By definition $(d/dy)v(x) \geq 0$, and the Cauchy–Riemann differential equations imply that $(d/dx)u(x) \geq 0$ as well.)

Example II.7: The function $\varphi(z) \triangleq z^\mu$ with $0 < \mu \leq 1$ is in $P(0,\infty)$. A function

$$\psi(z) \triangleq \alpha z + \beta + \sum_{i=1}^{m} \frac{\gamma_i}{\delta_i - z}$$

with $\alpha, \gamma_i > 0$, $\beta \in \mathbb{R}$, and $\delta_i \in \mathbb{R}\backslash(a,b)$ is also in $P(a,b)$.

Remark II.8: Let $\varphi \in P$ be a Pick function and μ its corresponding Borel measure (cf. (1)). Let $(a,b) \in \mathbb{R}$ be an interval. Then

$$\varphi \in P(a,b) \Leftrightarrow \mu((a,b)) = 0.$$

This remark is again from [5, p. 26], as is the next example [5, p. 27].

Example II.9:

$$\sqrt{z} = \frac{1}{\sqrt{2}} + \int_{-\infty}^{0} \left(\frac{1}{x-z} - \frac{x}{x^2+1}\right) \frac{\sqrt{x}}{\pi}\, dx.$$

III. Operator Monotonicity

The first part of this section introduces the notion of operator monotonicity. The results are taken from [5, pp. 67 ff], and they also appear in [1]. We show that Pick functions are operator-monotone. In the sequel, we consider operator convexity properties as was done in [6, pp. 230 ff]. It will be important for the next sections that root functions are operator-concave (cf. Example III.10). Here, all Hilbert spaces (usually denoted by $\mathcal{H}$, etc.) are supposed to be finite-dimensional.

Definition III.1: Let $\mathcal{H}$ be a finite-dimensional (complex) Hilbert space.

- $\mathcal{L}(\mathcal{H})$ denotes the algebra of linear operators on $\mathcal{H}$.
- $\mathcal{L}(\mathcal{H})^{\text{s.a.}}$ denotes the real vector space of self-adjoint operators on $\mathcal{H}$.
- An operator $A \in \mathcal{L}(\mathcal{H})$ is called *positive* if $\langle v \mid Av\rangle \geq 0$ for all $v \in \mathcal{H}$.
- $\mathcal{L}(\mathcal{H})^+$ denotes the convex cone of positive operators on $\mathcal{H}$. (Recall that $\mathcal{L}(\mathcal{H})^+ \subset \mathcal{L}(\mathcal{H})^{\text{s.a.}}$.)
- For operators $A, B \in \mathcal{L}(\mathcal{H})$ we write $A \leq B$ if $B - A$ is positive.

Definition III.2: Let $\mathcal{H}$ be a finite-dimensional (complex) Hilbert space, $I \subseteq \mathbb{R}$ an interval, $f : I \to \mathbb{R}$ a function, and $A \in \mathcal{L}(\mathcal{H})^{\text{s.a.}}$ with all its eigenvalues in I.

- For $A = \sum_{i=1}^n a_i |a_i\rangle\langle a_i|$, where $(|a_i\rangle)_{1\leq i\leq n}$ denotes an ON basis of $\mathcal{H}$, we define

  $$f(A) \triangleq \sum_{i=1}^{n} f(a_i)\,|\,a_i\rangle\langle a_i|.$$

  (Clearly, this is independent of the chosen basis.) In matrix notation we have

  $$A = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix} \overset{f}{\mapsto} f(A) = \begin{pmatrix} f(a_1) & & 0 \\ & \ddots & \\ 0 & & f(a_n) \end{pmatrix}$$

- f is called *operator monotone of order* n ("$f \in P_n(I)$") if

  $$B \leq C \Longrightarrow f(B) \leq f(C), \qquad \text{for all } B, C \in \mathcal{L}(\mathbb{C}^n)^{\text{s.a.}}.$$

- f is called *operator monotone* if $f \in P_n(I)$ for all $n \in \mathbb{N}$.

Lemma III.3:

- $P_n(I)$ is a convex cone.
- If $f \in P_n(I), g \in P_n(J)$ with $im(f) \subseteq J$ then $g \circ f \in P_n(I)$.
- $P_n(I)$ is closed (in the topology of pointwise convergence).
- $P_{n+1}(I) \subseteq P_n(I)$.
- For $\alpha > 0$, $\beta \in \mathbb{R}$, the function $x \mapsto \alpha x + \beta$ is operator monotone.
- $x \mapsto -(1/x)$ is operator monotone on $(0,\infty)$.

Proof: All assertions but the last are obvious. So, for $A \in \mathcal{L}(\mathcal{H})^+$ strictly positive (0 is no eigenvalue) and $v, w \in \mathcal{H}$ we have

$$\begin{aligned} |\langle v \mid w\rangle|^2 &= |\langle A^{-\frac{1}{2}}v \mid A^{\frac{1}{2}}w\rangle|^2 \\ &\leq \langle A^{-\frac{1}{2}}v \mid A^{-\frac{1}{2}}v\rangle\langle A^{\frac{1}{2}}w \mid A^{\frac{1}{2}}w\rangle \\ &= \langle v \mid A^{-1}v\rangle\langle w \mid Aw\rangle \end{aligned}$$

with equality, e.g., for $w = A^{-1}v$. (This is the Cauchy–Schwarz–Buniakowski inequality!) Therefore,

$$\langle v \mid A^{-1}v\rangle = \max_{w\neq 0} \frac{|\langle v \mid w\rangle|^2}{\langle w \mid Aw\rangle}$$

and this immediately implies that for $B, C \in \mathcal{L}(\mathcal{H})^{\text{s.a.}}$

$$B \leq C \Leftrightarrow B^{-1} \geq C^{-1} \Leftrightarrow -B^{-1} \leq -C^{-1}. \qquad \Box$$

Corollary III.4: Let $\varphi \in P(a,b)$. Then, $\varphi|_{(a,b)}$ is operator monotone.

Example III.5: $f(z) \triangleq z^\mu$ with $0 \leq \mu \leq 1$ is operator-monotone. (A direct proof of this fact appears also in [1, p. 115].) This is not (!) true if $\mu > 1$. For, e.g., $\mu = 2$

$$\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \leq \begin{pmatrix} 3.1 & 0 \\ 0 & 3.1 \end{pmatrix}$$

$$\text{but} \quad \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}^2 = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$$

$$\text{and} \quad \begin{pmatrix} 3.1 & 0 \\ 0 & 3.1 \end{pmatrix}^2 = \begin{pmatrix} 9.61 & 0 \\ 0 & 9.61 \end{pmatrix}.$$

Lemma III.6: Let $a \in \mathcal{L}(\mathcal{H}), b \in \mathcal{L}(\mathcal{H})^{\text{s.a.}}$, and $\varepsilon > 0$.

$$A_\lambda \triangleq \begin{pmatrix} \varepsilon\mathbf{1} & a \\ a^* & \lambda\mathbf{1} + b \end{pmatrix}$$

is a positive operator on $\mathcal{H} \oplus \mathcal{H}$ if $\lambda > 0$ is large enough.

Proof: We have to prove that $\langle\psi|A_\lambda|\psi\rangle \geq 0$ for all normed (!) $\psi = u \oplus v \in \mathcal{H} \oplus \mathcal{H}$.

For $v = 0$ (and $\lambda = 0$) we have $\langle\psi|A_0|\psi\rangle = \langle u|A_0|u\rangle = \varepsilon > 0$. Therefore, $\langle\psi|A_0|\psi\rangle > 0$ for a neighborhood N of the space with $v = 0$, and for all $\lambda > 0, \psi \in N : \langle\psi|A_\lambda|\psi\rangle > 0$.

So, we may assume that $\|v\|$ is larger than some positive constant γ. Let $\mu \in \mathbb{R}$ be the smallest eigenvalue of A_0. We have

$$\langle\psi|A_\lambda|\psi\rangle = \lambda \cdot \|v\|^2 + \langle\psi|A_0|\psi\rangle \geq \lambda\gamma^2 + \mu$$

and the proof is complete. $\Box$

Theorem III.7: Let $f \geq 0$ be a continuous and operator-monotone function on $[0,\infty)$, and let $x \in \mathcal{L}(\mathcal{H})^+$ and $a \in \mathcal{L}(\mathcal{H})$ with $\|a\|_{\text{op}} \leq 1$. Then

$$f(a^*xa) \geq a^*f(x)a. \tag{2}$$

Proof: Let $b \triangleq (\mathbf{1} - aa^*)^{1/2}$ and $c \triangleq (\mathbf{1} - a^*a)^{1/2}$, and define the operators

$$X \triangleq \begin{pmatrix} x & 0 \\ 0 & 0 \end{pmatrix} \qquad U \triangleq \begin{pmatrix} a & b \\ c & -a^* \end{pmatrix}$$

on $\mathcal{H} \oplus \mathcal{H}$. Let $\varepsilon > 0$, and let λ be large enough so that

$$Y \triangleq \begin{pmatrix} a^*xa + \varepsilon\mathbf{1} & 0 \\ 0 & \lambda\mathbf{1} \end{pmatrix} \geq \begin{pmatrix} a^*xa & a^*xb \\ bxa & bxb \end{pmatrix} = U^*XU.$$

$$\begin{aligned} \Rightarrow f(Y) &\geq f(U^*XU) = U^*f(X)U \\ &\geq U^* \begin{pmatrix} f(x) & 0 \\ 0 & 0 \end{pmatrix} U = \begin{pmatrix} a^*f(x)a & a^*f(x)b \\ bf(x)a & bf(x)b \end{pmatrix}. \end{aligned}$$

Therefore, $f(a^*xa + \varepsilon\mathbf{1}) \geq a^*f(x)a$, and this implies the claim.

If we assumed $\langle u | f(a^*xa + \varepsilon\mathbf{1}) - a^*f(x)a | u\rangle < 0$ for some $u \in \mathcal{H}$ and $\varepsilon = 0$, this would be also true for some $\varepsilon > 0$ (as $f(a^*xa + \varepsilon\mathbf{1})$ varies continuously with ε). □

Remark III.8: We call functions f that fulfil (2) *operator concave*. Indeed, a nonnegative continuous function on $[0, \infty)$ that is operator-concave is also operator-monotone (cf. [6, p. 232], or [1, p. 120]). Furthermore, operator-concave functions f fulfil *Jensen's inequality*. We have

$$f\Big(\sum_i a_i^* x_i a_i\Big) \geq \sum_i a_i^* f(x_i) a_i$$

for all $x_1, \ldots, x_n \in \mathcal{L}(\mathcal{H})^+$ and $a_1, \ldots, a_n \in \mathcal{L}(\mathcal{H})$ with $\sum_i a_i^* a_i \leq \mathbf{1}$.

Proof: Define

$$X \triangleq \begin{pmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_n \end{pmatrix}, \quad A \triangleq \begin{pmatrix} a_1 & \\ \vdots & 0 \\ a_n & \end{pmatrix}$$

$$\Rightarrow A^*XA = \begin{pmatrix} \sum_i a_i^* x_i a_i & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} f(\sum_i a_i^* x_i a_i) & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{pmatrix} = f(A^*XA)$$

$$\underset{\text{by (2)}}{\geq} A^*f(X)A = \begin{pmatrix} \sum_i a_i^* f(x_i) a_i & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & 0 & \\ 0 & & & \end{pmatrix}. \quad \Box$$

Corollary III.9: Let $f \geq 0$ be a continuous and operator monotone function on $[0, \infty)$, and let $x \in \mathcal{L}(\mathcal{H})^+$ and $a: \mathcal{H}' \rightarrow \mathcal{H}$ a (linear) contraction ($\|a\|_{\text{op}} \leq 1$). We have $f(a^*xa) \geq a^*f(x)a$.

Example III.10: For $0 \leq \mu \leq 1$ we have $(a^*xa)^\mu \geq a^*x^\mu a$ (with notation from Corollary III.9).

IV. Finite Quantum Systems and Physical Maps

In this section we introduce the notion of finite quantum systems and the maps of such systems that are considered to be in accordance with quantum physics' laws. Toward this goal we will consider both the Schrödinger and the Heisenberg pictures, starting with the latter and transforming it into the former.

Definition IV.1: A *finite quantum system* $\mathcal{A}$ is a finite-dimensional C^**-algebra*, i.e., a self-adjoint subalgebra (with identity) of some $\mathcal{L}(\mathcal{H})$, where $\dim(\mathcal{H}) < \infty$.

A *physical state* on a finite quantum system $\mathcal{A}$ is an element $A \in \mathcal{A}^+ \triangleq \mathcal{L}(\mathcal{H})^+ \cap \mathcal{A}$ for which $\mathrm{tr}(A) = 1$. If $A \in \mathcal{A}^+$ is a physical state we call $\mathrm{tr}(A\cdot) = \mathrm{tr}(A^*\cdot) \in \mathcal{A}^*$ a *physical state* as well.

Remark IV.2: Let $\mathcal{A} \subseteq \mathcal{L}(\mathcal{H})$ and A be a physical state on $\mathcal{A}$. Then (by abuse of notation)

$$A = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{pmatrix}$$

for an ON basis of A-eigenvectors, and A may be interpreted as a probability distribution on these basis vectors.

Definition IV.3: A $\mathbb{C}$-linear map $\alpha: \mathcal{B} \rightarrow \mathcal{A}$ is *positive* if $\alpha(\mathcal{B}^+) \subseteq \mathcal{A}^+$. It is called *completely positive* if all maps of the form

$$\begin{aligned} \alpha \otimes \mathbf{1}: \mathcal{B} \otimes \mathcal{C} &\rightarrow \mathcal{A} \otimes \mathcal{C} \\ \sum_i b_i \otimes c_i &\mapsto \alpha(b_i) \otimes c_i \end{aligned}$$

are positive.

The *Heisenberg picture* (HP) of a *physical map* of finite quantum systems $\mathcal{A}, \mathcal{B}$ is a completely positive and unity preserving $\mathbb{C}$-linear map $\alpha: \mathcal{B} \rightarrow \mathcal{A}$.

Remark IV.4: Physical maps are *Schwarz maps*: They fulfill, for all $x \in \mathcal{B}$, the inequality $\alpha(x^*x) \geq \alpha(x)^*\alpha(x)$. This can be deduced from Stinespring's theorem (see [4, p. 137] or [17] for a proof).

Theorem IV.5: Let $\alpha: \mathcal{A} \rightarrow \mathcal{L}(\mathcal{H})$ be a completely positive map of finite quantum systems. Then

$$\alpha(A) = V^*\rho(A)V \qquad (\text{for all } A \in \mathcal{A})$$

for some representation $\rho: \mathcal{A} \rightarrow \mathcal{L}(\mathcal{K})$ with $\mathcal{K}$ a finite-dimensional Hilbert space, and a linear map $V: \mathcal{H} \rightarrow \mathcal{K}$. ($\rho$ is an algebra homomorphism with $\rho(A^*) = \rho(A)^*$ for all $A \in \mathcal{A}$ and $\rho(\mathbf{1}) = \mathbf{1}$.)

Let $\alpha: \mathcal{B} \rightarrow \mathcal{A}$ be a $\mathbb{C}$-linear map of finite-dimensional C^*-algebras (HP). Its adjoint is a $\mathbb{C}$-linear map of the respective spaces of linear forms on the C^*-algebras

$$\begin{aligned} \alpha^*: \mathcal{A}^* &\rightarrow \mathcal{B}^* \\ \lambda &\mapsto \lambda \circ \alpha. \end{aligned} \tag{3}$$

We perform still another transformation, using the fact that

$$\langle A \mid A' \rangle_{\mathcal{A}} \triangleq \mathrm{tr}_{\mathcal{A}}(A^*A')$$

defines an inner product on the finite-dimensional C^*-algebra $\mathcal{A}$. $(\mathcal{A}, \langle \cdot | \cdot \rangle_{\mathcal{A}})$ is a Hilbert space. Therefore, there is a natural transformation $\mathcal{A} \rightarrow \mathcal{A}^*$ by $A \mapsto \mathrm{tr}(A^*\cdot)$, and we may rewrite (3) as a map $\alpha_* : \mathcal{A} \rightarrow \mathcal{B}$ that fulfils

$$\alpha^*(\mathrm{tr}_{\mathcal{A}}(A^*\cdot)) = \mathrm{tr}_{\mathcal{B}}(\alpha_*(A)^*\cdot). \tag{4}$$

We will refer to the map α_* as the *Schrödinger picture* (SP). The map α_* represents a physical map of finite quantum systems if it is completely positive and trace-preserving.

These definitions demonstrate the formal relationship between the Schrödinger and the Heisenberg picture, but, of course, there is still a more "practical" relationship between the two pictures. They lead to the same predictions concerning results of measurements (which are the relevant things in this context because they can be verified by an experiment).

Let $A \in \mathcal{A}^+$ be a physical state, $\alpha: \mathcal{B} \rightarrow \mathcal{A}$ a physical map (HP) corresponding to some prolongation of the physical system, and $(D_i)_{1 \leq i \leq N} \subset \mathcal{B}$ a positive operator measurement (POM: a tuple of positive operators that sum up to unity). It is the idea of the Schrödinger picture $\alpha_*\colon \mathcal{A} \to \mathcal{B}$ that $\alpha_*(A)$ represents the "outcome" (as a physical state in $\mathcal{B}$) of the prolongation, and so the probability that the result of the measurement will be $i = 1, \dots, N$ is $\langle \alpha_*(A) \mid D_i \rangle_{\mathcal{B}} = \mathrm{tr}_{\mathcal{B}}(\alpha_*(A) D_i)$.

Using the Heisenberg picture $\alpha\colon \mathcal{B} \to \mathcal{A}$ of the physical map, we can calculate these probabilities by applying α on the D_i's instead (!).

Proposition IV.6: Let $A \in \mathcal{A}^+$ be a physical state, $(D_i)_{1 \le i \le N} \subset \mathcal{B}$ a POM, $\alpha\colon \mathcal{B} \to \mathcal{A}$ the Heisenberg picture of a physical map, and $\alpha_*\colon \mathcal{A} \to \mathcal{B}$ its Schrödinger picture. Then

$$\langle \alpha_*(A) | D_i \rangle_{\mathcal{B}} = \langle A | \alpha(D_i) \rangle_{\mathcal{A}}, \qquad \text{for all } i = 1, \dots, N.$$

Proof:

$$\begin{aligned} \mathrm{tr}_{\mathcal{B}}(\alpha_*(A) D_i) &\underset{\mathrm{by}(3)}{=} \alpha^*(\mathrm{tr}_{\mathcal{A}}(A^* \cdot))(D_i) \\ &\underset{\mathrm{by}(4)}{=} \mathrm{tr}_{\mathcal{A}}(A^* \alpha(D_i)). \end{aligned} \qquad \square$$

V. Uhlmann's Monotonicity Theorem

Uhlmann's Monotonicity Theorem is our key tool to prove the quantum version of a data processing inequality. Its proof (the proof of the following lemma) uses the operator concavity of the root functions (see Example III.10). In this section we closely followed [12, p. 18 ff].

Lemma V.1: Let $\alpha\colon \mathcal{A}_2 \to \mathcal{A}_1$ be a physical map of finite quantum systems (HP), $S_1, T_1 \in \mathcal{A}_1^+$ and $S_2, T_2 \in \mathcal{A}_2^+$ with T_i invertible $(i = 1, 2)$. If for all $a \in \mathcal{A}_2^+$

$$\mathrm{tr}(S_1 \alpha(a)) \le \mathrm{tr}(S_2 a) \quad \text{and} \quad \mathrm{tr}(T_1 \alpha(a)) \le \mathrm{tr}(T_2 a)$$

then we have, for all $0 \le t \le 1$ and $x \in \mathcal{A}_2$

$$\mathrm{tr}\left(\alpha(x)^* S_1^t \alpha(x) T_1^{1-t}\right) \le \mathrm{tr}\left(x^* S_2^t x T_2^{1-t}\right).$$

Example V.2:

$$\mathrm{tr}\left(S_1^t T_1^{1-t}\right) \le \mathrm{tr}\left(S_2^t T_2^{1-t}\right).$$

Proof: For $a \in \mathcal{A}_2$ define a linear map $V\colon \mathcal{A}_2 \to \mathcal{A}_1$ by

$$a T_2^{1/2} \mapsto \alpha(a) T_1^{1/2}.$$

The map V is a contraction. In fact,

$$\begin{aligned} \left\| \alpha(a) T_1^{\frac{1}{2}} \right\|^2 &= \mathrm{tr}(T_1 \alpha(a)^* \alpha(a)) \le \mathrm{tr}(T_1 \alpha(a^* a)) \\ &\le \mathrm{tr}(T_2 a^* a) = \left\| a T_2^{\frac{1}{2}} \right\|^2. \end{aligned}$$

Define $\Delta_i \in \mathcal{L}(\mathcal{A}_i)^+$ by

$$\Delta_i(a T_i^{1/2}) \triangleq S_i a T_i^{-1/2}$$

for $i = 1, 2$ and $a \in \mathcal{A}_i$. (It is positive because $\langle a T_i^{1/2} | \Delta_i | a T_i^{1/2} \rangle = \langle a T_i^{1/2} \mid S_i a T_i^{-1/2} \rangle = \mathrm{tr}(T_i^{1/2} a^* S_i a T_i^{-1/2}) = \mathrm{tr}(a^* S_i a) \ge 0$.) We have

$$\Delta_i^t (a T_i^{1/2}) = S_i^t a T_i^{1/2 - t} \qquad (t \ge 0)$$

and $V^* \Delta_1 V \le \Delta_2$

$$\begin{aligned} \left\langle a T_2^{\frac{1}{2}} \middle| V^* \Delta_1 V \middle| a T_2^{\frac{1}{2}} \right\rangle &= \left\langle \alpha(a) T_1^{\frac{1}{2}} \middle| \Delta_1 \middle| \alpha(a) T_1^{\frac{1}{2}} \right\rangle \\ &= \left\langle \alpha(a) T_1^{\frac{1}{2}} \middle| S_1 \alpha(a) T_1^{-\frac{1}{2}} \right\rangle = \mathrm{tr}(\alpha(a)^* S_1 \alpha(a)) \\ &\le \mathrm{tr}(S_1 \alpha(a a^*)) \le \mathrm{tr}(S_2 a a^*) = \left\langle a T_2^{\frac{1}{2}} \middle| \Delta_2 \middle| a T_2^{\frac{1}{2}} \right\rangle. \end{aligned}$$

So,[4] $V^* \Delta_1^t V \le (V^* \Delta_1 V)^t \le \Delta_2^t$, and

$$\begin{aligned} \mathrm{tr}\left(T_1^{\frac{1}{2}} \alpha(x)^* S_1^t \alpha(x) T_1^{\frac{1}{2} - t}\right) &= \left\langle x T_2^{\frac{1}{2}} \middle| V^* \Delta_1^t V \middle| x T_2^{\frac{1}{2}} \right\rangle \\ &\le \left\langle x T_2^{\frac{1}{2}} \middle| \Delta_2^t \middle| x T_2^{\frac{1}{2}} \right\rangle = \mathrm{tr}\left(T_2^{\frac{1}{2}} x^* S_2^t x T_2^{\frac{1}{2} - t}\right). \end{aligned} \qquad \square$$

Definition V.3: Let $A \in \mathcal{A}^+$ be a physical state. Its *support* $(\mathrm{supp} A)$ is the projector on the space spanned by the eigenvectors corresponding to nonzero eigenvalues.

Two physical states $A, B \in \mathcal{A}^+$ have *divergence*

$$D(A \,\|\, B) \triangleq \begin{cases} \mathrm{tr}(A(\log A - \log B)), & \text{if } \mathrm{supp} A \le \mathrm{supp} B \\ \infty, & \text{otherwise.} \end{cases}$$

Physical states $\omega = \mathrm{tr}(A \cdot), \varphi = \mathrm{tr}(B \cdot) \in \mathcal{A}^*$ have *divergence*

$$D(\omega \| \varphi) \triangleq D(A \| B).$$

Note that divergence is often called *relative entropy*, too.

We present Uhlmann's monotonicity theorem [18].

Theorem V.4: Let $\alpha\colon \mathcal{A}_2 \to \mathcal{A}_1$ be a physical map of finite quantum systems (HP) and $\omega, \varphi \in \mathcal{A}_1^*$ physical states. Then

$$D(\omega \circ \alpha \| \varphi \circ \alpha) \le D(\omega \| \varphi).$$

Proof: Let $\omega = \mathrm{tr}(S_1 \cdot)$, $\varphi = \mathrm{tr}(T_1 \cdot)$, $\omega \circ \alpha = \mathrm{tr}(S_2 \cdot)$, and $\varphi \circ \alpha = \mathrm{tr}(T_2 \cdot)$, where without loss of generality $\mathrm{supp} T_1 = \mathbf{1}$.

As

$$\mathrm{tr}(S_1 \alpha(\cdot)) = \omega \circ \alpha = \mathrm{tr}(S_2 \cdot)$$

and

$$\mathrm{tr}(T_1 \alpha(\cdot)) = \varphi \circ \alpha = \mathrm{tr}(T_2 \cdot)$$

Lemma V.1 implies that

$$\mathrm{tr}\left(S_1^\mu T_1^{1-\mu}\right) \le \mathrm{tr}\left(S_2^\mu T_2^{1-\mu}\right), \qquad \text{for all } 0 \le \mu \le 1.$$

Consequently,

$$(1 - \mathrm{tr}(S_1^\mu T_1^{1-\mu}))/(1-\mu) \ge (1 - \mathrm{tr}(S_2^\mu T_2^{1-\mu}))/(1-\mu)$$

and by the limit $\mu \to 1$

$$\begin{aligned} D(\omega \,\|\, \varphi) &= \mathrm{tr}\left(S_1^\mu T_1^{1-\mu}\right)'\Big|_{\mu \to 1} \\ &\ge \mathrm{tr}\left(S_2^\mu T_2^{1-\mu}\right)'\Big|_{\mu \to 1} = D(\omega \circ \alpha \,\|\, \varphi \circ \alpha). \end{aligned} \qquad \square$$

VI. The Quantum Data Processing Inequality

In this final section we derive the quantum data processing inequality from Uhlmann's monotonicity theorem. We start with the formulation of the latter in the Schrödinger picture.

Corollary VI.1: Let $\alpha_*\colon \mathcal{A} \to \mathcal{B}$ be a physical map of finite quantum systems (SP) and $A_1, A_2 \in \mathcal{A}^+$ physical states. Then

$$D(\alpha_*(A_1) \,\|\, \alpha_*(A_2)) \le D(A_1 \,\|\, A_2).$$

Proof:

$$\begin{aligned} D(A_1 \,\|\, A_2) &= D(\mathrm{tr}(A_1 \cdot) \,\|\, \mathrm{tr}(A_2 \cdot)), \\ D(\alpha_*(A_1) \,\|\, \alpha_*(A_2)) &= D(\mathrm{tr}(\alpha_*(A_1) \cdot) \,\|\, \mathrm{tr}(\alpha_*(A_2) \cdot)), \end{aligned}$$

we have $\mathrm{tr}(\alpha_*(A_i) \cdot) = \mathrm{tr}(A_i \alpha(\cdot))$ (cf. the proof of Proposition IV.6), and the claim reduces to Theorem V.4. □

Definition VI.2: Let $W_*\colon \mathcal{A} \to \mathcal{B}$ be a quantum channel, i.e., a physical map of finite quantum systems (SP). Let $A \in \mathcal{A}^+$ be a physical

[4] We use Example III.10 (as we promised above).

state with $A = \sum_{a=1}^{n} p_a |a\rangle\langle a|$ its spectral decomposition, and let $B \triangleq W_*(A)$ be the corresponding output (physical) state. Furthermore, let

$$(A, B) \triangleq \sum_{a=1}^{n} p_a |a\rangle\langle a| \otimes W_*(|a\rangle\langle a|)$$

be the "joint state." The *mutual information* is given by

$$I(A; W_*) \triangleq H(A) + H(B) - H(A, B)$$

where $H(X) \triangleq -\mathrm{tr}(X \log X)$ denotes the Shannon–von Neumann entropy.

Remark VI.3: Let

$$H(B|a) \triangleq H(W_*(|a\rangle\langle a|))$$

be the entropy of the output state for the input state $|a\rangle\langle a|$ and

$$H(B|A) \triangleq \sum_{a=1}^{n} p_a H(B \mid a)$$

its expectation. We have $H(B|A) = H(A, B) - H(A)$, and mutual information may be written in the form

$$I(A; W_*) \triangleq H(B) - H(B \mid A)$$

which is known as Holevo's bound.

Proof: Provided that the logarithm respects eigenspaces we get

$$\begin{aligned}
&H(A, B) - H(A)\\
&= -\mathrm{tr}\left(\sum_{a=1}^{n} p_a |a\rangle\langle a| \otimes W_*(|a\rangle\langle a|)\right.\\
&\qquad \left.\log\left(\sum_{a'=1}^{n} p'_a |a'\rangle\langle a'| \otimes W_*(|a'\rangle\langle a'|)\right)\right)\\
&\quad + \mathrm{tr}\left(\sum_{a=1}^{n} p_a |a\rangle\langle a| \log\left(\sum_{a'=1}^{n} p'_a |a'\rangle\langle a'|\right)\right)\\
&= -\sum_{a=1}^{n} p_a \mathrm{tr}_2(W_*(|a\rangle\langle a|) \log(p_a W_*(|a\rangle\langle a|)))\\
&\quad + \sum_{a=1}^{n} p_a \log p_a\\
&= -\sum_{a=1}^{n} p_a \mathrm{tr}_2(W_*(|a\rangle\langle a|) \log W_*(|a\rangle\langle a|))\\
&= H(B \mid A).
\end{aligned}$$

□

The quantum data processing inequality.

Corollary VI.4: Let $W_*: \mathcal{A}_1 \to \mathcal{A}_2$ and $D_*: \mathcal{A}_2 \to \mathcal{A}_3$ be physical maps of finite quantum systems (SP) and $A \in \mathcal{A}_1^+$ a physical state. Then

$$I(A; D_* \circ W_*) \leq I(A; W_*).$$

Proof: We have

$$\begin{aligned}
D((A, B) \| A \otimes B) &= \mathrm{tr}((A, B)(\log(A, B) - \log(A \otimes B)))\\
&= \mathrm{tr}((A, B) \log(A, B)) - \mathrm{tr}((A, B) \log(A \otimes B))\\
&= -H(A, B) - \mathrm{tr}((A, B) \log(A \otimes \mathbf{1}))\\
&\quad - \mathrm{tr}((A, B) \log(\mathbf{1} \otimes B))\\
&= -H(A, B) + H(A) + H(B) = I(A; W_*)
\end{aligned}$$

and the claim follows from

$$D((\mathbf{1} \otimes D_*)(A, B) \,\|\, (\mathbf{1} \otimes D_*)(A \otimes B)) \leq D((A, B) \,\|\, A \otimes B).$$ □

As a special case, we get *Holevo's upper bound* for classical communication over quantum channels [8]. We interprete a POM $(D_i)_{1 \leq i \leq N} \subset \mathcal{A}_2$ (see Section IV) as a physical map D_* into the (commutative C^*-algebra) diagonal of $\mathcal{L}(\mathbb{C}^N)$, with

$$D_*(X) \triangleq \mathrm{diag}(\mathrm{tr}(D_1 X), \ldots, \mathrm{tr}(D_N X)).$$

Corollary VI.5: Let $W_*: \mathcal{A}_1 \to \mathcal{A}_2$ be a quantum channel and D_* a measurement on its output space, i.e., a physical map from $\mathcal{A}_2$ into some commutative C^*-algebra $\mathcal{A}_3$. Then $I(A; D_* \circ W_*) \leq I(A; W_*)$.

We remind the reader that Holevo's result is from 1973 whereas Uhlmann's monotonicity theorem is from 1977. Both use analytical considerations for the proofs.

There is a proof of Holevo's upper bound by H. P. Yuen and M. Ozawa deriving it directly from Uhlmann's monotonicity theorem [20], and another one by A. Winter using only information-theoretical considerations [19].

Acknowledgment

The authors wish to thank Andreas Winter for fruitful discussions about quantum information theory.

References

[1] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.: Academic, 1981.

[4] E. B. Davis, *Quantum Theory of Open Systems*. London, U.K.: Academic, 1976.

[5] W. F. Donoghue Jr., *Monotone Matrix Functions and Analytic Continuation*. Berlin, Germany: Springer-Verlag, 1970.

[6] F. Hansen and G. K. Pederson, "Jensen's inequality and Löwner's theorem," *Math. Ann.*, vol. 258, pp. 229–241, 1982.

[7] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A*, vol. 54, no. 3, pp. 1869–1876, 1996.

[8] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Probl. Pered. Inform.*, vol. 9, no. 3, pp. 3–11, 1973. (English translation: *Probl. Inform. Transm.*, vol. 9, no. 3, pp. 177–183, 1973).

[9] ——, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inform. Theory*, vol. 44, pp. 269–273, Jan. 1998.

[10] M. Horodecki, "Limits for compression of quantum information carried by ensembles of mixed states," *Phys. Rev. A*, vol. 57, no. 5, pp. 3364–3369, 1998.

[11] E. H. Lieb, "Convex trace functions and the Wigner-Yanase-Dyson conjecture," *Adv. Math.*, vol. 11, pp. 267–288, 1973.

[12] M. Ohya and D. Petz, *Quantum Entropy and Its Use (Texts and Monographs in Physics)*. Berlin, Germany: Springer-Verlag, 1993.

[13] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Phys. Rev. A*, vol. 54, no. 4, pp. 2629–2635, 1996.

[14] B. Schumacher, M. D. Westmoreland, and W. K. Wootters, "Limitation on the amount of accessible information in a quantum channel," *Phys. Rev. Lett.*, vol. 76, no. 18, pp. 3452–3455, 1996.

[15] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, 1997.

[16] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 339–425 and pp. 623–656, 1948.

[17] W. F. Stinespring, "Positive functions on C^*-algebras," in *Proc. Amer. Math. Soc.*, vol. 6, 1955, pp. 211–216.

[18] A. Uhlmann, "Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory," *Commun. Math. Phys.*, vol. 54, pp. 21–32, 1977.

[19] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2481–2485, Nov. 1999.

[20] H. P. Yuen and M. Ozawa, "Ultimate information carrying limit of quantum systems," *Phys. Rev. Lett.*, vol. 70, no. 4, pp. 363–366, 1993.